

anses

alimentation, environnement, 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

Avis de l'Anses
Rapport d'expertise collective

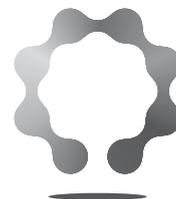
Janvier 2011

Édition scientifique



anses

alimentation, environnement, 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

Avis de l'Anses

Rapport d'expertise collective

Janvier 2011

Édition scientifique

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).

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.

**RAPPORT
d'expertise collective**

Groupe de travail « traitement statistique des données des études de toxicité sub-chronique de 90 jours chez le rat présentées dans les dossiers d'autorisation de mise sur le marché des OGM »

Comité d'experts spécialisé « Biotechnologie »

Janvier 2011

Il s'agit d'un rapport élaboré par le Groupe de travail « traitement statistique des données des études de toxicité sub-chronique de 90 jours chez le rat présentées dans les dossiers d'autorisation de mise sur le marché des OGM » et validé par le Comité d'experts spécialisé « BIOTECHNOLOGIE ».

Présentation des intervenants

COMPOSITION DU GROUPE DE TRAVAIL

Président

M Joël GUILLEMAIN – Expert Toxicologue, Sésame, Chambray les Tours.

Membres

Mme Michèle DE MONTE – Ingénieur de recherche en Biologie et en Biostatistique, INSERM U618, Faculté de médecine, Université François Rabelais, Tours.

M Marc LAVIELLE – Statisticien, Professeur à l'Université Paris-Descartes, actuellement en détachement à l'INRIA Saclay comme Directeur de Recherche, Saclay.

M Carlos LOPEZ – Chef du Service Biométrie, Institut de l'élevage, Paris.

M David MAKOWSKI – Directeur de recherche INRA en Agronomie et Biostatistique, UMR 211 Agronomie, INRA/AgroParisTech, Thiverval Grignon.

PARTICIPATION ANSES

Coordination scientifique

Mme Chantal ARAR – Chargée de projets scientifiques, Unité "Evaluation des risques biologiques dans les aliments", Département "Evaluation des risques liés aux aliments", Direction d'Evaluation des Risques, Anses, Maisons- Alfort.

Mme Sonia TENAILLEAU – Chef d'Unité "Evaluation des risques biologiques dans les aliments", Département "Evaluation des risques liés aux aliments", Direction d'Evaluation des Risques, Anses, Maisons- Alfort.

Appui scientifique et technique

Mme Diane CUZZUCOLI – Chargée de projets scientifiques, Unité "Appui et études en microbiologie et santé animale", Département "Appui méthodologique et études", Direction d'Evaluation des Risques, Anses, Maisons- Alfort.

Mme Françoise GAUCHARD – Chef d'Unité "Appui et études en microbiologie et santé animale", Département "Appui méthodologique et études", Direction d'Evaluation des Risques, Anses, Maisons- Alfort.

Mme Sonia POISSON – Chargée de projets scientifiques, Unité "Appui et études en microbiologie et santé animale", Département "Appui méthodologique et études", Direction d'Evaluation des Risques, Anses, Maisons- Alfort.

M Moez SANAA – chargé de mission "méthodologie et risques biologiques", Département "Appui méthodologique et études", Direction d'Evaluation des Risques, Anses, Maisons- Alfort.

Secrétariat administratif

Mme Fabienne M'BASSE

EXPERTS CONSULTES OU RELECTEURS

M Eric ANTIGNAC – Expert toxicologue Eurotox[®], Groupe l'Oréal, Paris.

M Marc DELOS – Ingénieur du Génie Rural, des Eaux et Forêts, Expert National Biovigilance et Grandes cultures auprès de la Direction Générale de l'alimentation, Service régionale de l'alimentation Midi-Pyrénées, Toulouse.

M Florian GUILLOU – Directeur de l'UMR INRA/CNRS/université de Tours/IFCE Physiologie de la Reproduction et des Comportements, UMR INRA /CNRS/Université Tours, Nouzilly.

Mme Martine KAMMERER – Professeur de pharmacie et toxicologie à L'Ecole Nationale Vétérinaire, Agroalimentaire et de l'Alimentation Nantes Atlantique, Oniris, Nantes.

M Michel LAURENTIE – Chargé de recherche en pharmacologie et physiologie animale de l'Ecole Nationale Polytechnique de Toulouse, Unité Pharmacocinétique-pharmacodynamie, Anses, Fougères.

SOMMAIRE

PRÉSENTATION DES INTERVENANTS	2
LISTE DES FIGURES, TABLEAUX ET ANNEXES	6
INTRODUCTION	7
CHAPITRE 1 : MÉTHODES D'ANALYSE STATISTIQUE DES DONNÉES ISSUES DES ÉTUDES DE TOXICITÉ SUB-CHRONIQUE DE 90 JOURS CHEZ LE RAT.	9
1. PLAN EXPÉRIMENTAL DES ÉTUDES DE TOXICITÉ SUB-CHRONIQUES DE 90 JOURS	10
1.1. Protocole expérimental 408 de l'OCDE	10
Toxicité orale à doses répétées-rongeurs: 90 jours	10
1.2. Protocole 408 tel qu'il est mis en œuvre dans les études OGM	10
1.2.1. Plan expérimental	11
1.2.2. Paramètres biologiques mesurés et leurs significations toxicologiques.	11
1.2.3. Méthodes statistiques utilisées	11
2. REVUE DES ANALYSES STATISTIQUES APPLICABLES AUX DONNÉES DU PROTOCOLE 408	12
2.1. Analyse descriptive des données	12
2.1.1. Détection de valeurs atypiques	12
2.1.2. Gestion des valeurs atypiques	13
2.1.3. Transformation des données	15
2.2. Analyse des paramètres biologiques hors évolution pondérale.	16
2.2.1. Test de différences	16
2.2.2. Multiplicité des tests et conséquence sur les risques liés aux tests	19
2.2.3. Puissance	20
2.2.4. Test de non différence ou d'équivalence	22
2.3. Analyse de l'évolution pondérale et de la consommation hebdomadaire.	26
3. PRÉSENTATION ET AIDE À L'INTERPRÉTATION	27

CHAPITRE 2 : APPLICATION DES MÉTHODES STATISTIQUES AUX DONNÉES D'UNE ÉTUDE DE TOXICITÉ SUB-CHRONIQUE CHEZ LE RAT PAR ADMINISTRATION RÉITÉRÉE VIA L'ALIMENTATION D'UNE PLANTE GM PENDANT 90 JOURS.	28
1. INTRODUCTION	29
2. MATÉRIELS ET MÉTHODES	29
2.1 Protocole expérimental de l'étude.	29
2.2 Gestion des données	29
2.3 Analyse statistique	30
3. RÉSULTATS	32
3.1 Données atypiques	32
3.2 Transformation des variables	32
3.3 Tests des différences entre les groupes OGM et témoin	34
3.4 Analyse de l'évolution pondérale.	39
3.5 Analyse de la puissance	42
3.6 Tests d'équivalence	42
4. INTERPRÉTATION STATISTIQUE	47
5. INTERPRÉTATION TOXICOLOGIQUE	48
CONCLUSION GÉNÉRALE	51
BIBLIOGRAPHIE	52
ANNEXES	55

Liste des figures, tableaux et annexes

Figures

Figure 1 : Valeurs des concentrations en ALAT (SGPTALT) pour les 10 animaux du groupe OGM 11%.	13
Figure 2 : Arbre décisionnel pour les données atypiques.	14
Figure 3 : Fonction de répartition du paramètre ABSNEUT chez les femelles.	15
Figure 4: Puissance du test de comparaison de deux moyennes en fonction de la taille d'effet.	21
Figure 5: Nombre d'animaux nécessaire en fonction de la taille d'effet pour une puissance de 80%.	22
Figure 6 : Principe de l'équivalence range, l'eRg.	24
Figure 7: Nombre d'animaux par groupe pour le test d'équivalence en fonction du seuil de tolérance (Δ) et de la vraie différence entre les deux groupes comparés (taille d'effet).	25
Figure 8: Représentation graphique des 38 données atypiques identifiées dans l'étude MON810.	33
Figure 9: Valeurs atypiques de la femelle 14 du groupe OGM 11% pendant la période 2.	34
Figure 10: p-value des tests de comparaison du nombre de monocytes absolus en $10^3/\mu\text{l}$	38
Figure 11: Taille de l'effet observé pour le paramètre nombre de monocytes absolus en $10^3/\mu\text{l}$	38
Figure 12: Taille de l'effet observé pour le paramètre nombre de monocytes absolus.	39
Figure 13: Courbes de croissance pondérale individuelle.	39
Figure 14: Courbes de croissance moyenne par sexe pour chaque groupe.	40
Figure 15: Graphes explorant la qualité de l'ajustement.	41
Figure 16: Valeurs observées en fonction des valeurs prédites avec un modèle.	41

Tableaux

Tableau 1: Répartition des p-values inférieures à 0.1 dans les différents groupes de l'expérience.	34
Tableau 2: Tests paramétriques de comparaison de moyennes statistiquement significatifs.	36
Tableau 3: Paramètres du modèle non linéaire mixte de la croissance pondérale.	40
Tableau 4: Puissance des tests pour les variables ayant au moins un test significatif (p-value < 10%).	43
Tableau 5: Liste des tests concernant les variables sans différence significative et avec une puissance jugée insuffisante comparativement aux tailles d'effet de référence.	45

Annexes

ANNEXE 1 : Tableau de synthèse de la revue bibliographique.	55
ANNEXE 2 : Tableau de synthèse de la revue bibliographique.	59
ANNEXE 3 : Les mesures d'influence.	64
ANNEXE 4 : Le False Discovery Rate (FDR).	65
ANNEXE 5 : Paramètres étudiés dans l'étude MON810.	69
ANNEXE 6 : Programme SAS.	71
ANNEXE 7 : Comparaison des tests paramétriques et non paramétriques pour les variables quantitatives.	83
ANNEXE 8 : p-Value pour le test de comparaison pour les différentes variables.	84
ANNEXE 9 : Taille d'effet observée pour les différentes variables.	85
ANNEXE 10 : Puissance.	86
ANNEXE 11 : Transformation des variables quantitatives.	90

Introduction

L'évaluation des risques liés à la consommation des denrées alimentaires et aliments contenant des plantes ou dérivés de l'OGM repose notamment sur l'équivalence substantielle¹ entre l'OGM et son équivalent². Selon les Lignes Directrices de l'EFSA (EFSA, 2006, 2008a) cette évaluation comporte une série d'études dont l'objectif est d'identifier et de caractériser d'éventuels effets néfastes liés à la consommation par l'homme ou l'animal d'une plante génétiquement modifiée ou de ses produits dérivés.

L'innocuité ou non de la plante génétiquement modifiée dépend de deux éléments principaux :

- ✓ l'une est inhérente à l'expression du ou des produits du transgène ; il s'agit de rechercher si la ou les protéines d'intérêt nouvellement produites par la plante sont potentiellement toxiques. Cette évaluation comprend des essais toxicologiques par administration unique ou chronique de la dite protéine chez l'animal.
- ✓ l'autre est liée à l'existence éventuelle de phénomènes inattendus (e.g., la synthèse de facteurs anti-nutritionnels ou de substances toxiques naturellement présentes ou non dans la plante) induit par l'insertion du transgène dans le génome. L'existence de tels phénomènes peut être testée en effectuant des essais toxicologiques chez l'animal par administration répétée via l'alimentation d'un produit représentatif de la plante génétiquement modifiée, obtenu dans des conditions habituelles de culture. Ces essais permettent notamment d'étudier l'activité des protéines d'intérêt sur les molécules *in planta* (e.g. interaction avec les substances chimiques herbicides et leurs métabolites).

Dès sa création, l'Afssa a mené une réflexion sur la pertinence et le contenu des études toxicologiques conduites sur animaux de laboratoire réflexion qui, dans le cadre de l'évaluation de la sécurité d'aliments composés ou issus d'OGM, a fait l'objet de deux avis de l'agence. Ils apportent un éclairage sur le choix de la durée des essais sur les animaux (AFSSA, 2002, 2007). Ce sujet a également été discuté et documenté au sein du panel OGM de l'EFSA (EFSA, 2008b). Cependant, les agences d'évaluation ne se sont jamais prononcées sur la pertinence des méthodes et des outils d'analyse statistiques des données de ces études toxicologiques et alimentaires.

L'étude de l'OGM ou de produits issus de l'OGM par administration répétée 90 jours chez les rongeurs, via l'alimentation, est la seule étude à visée toxicologique réalisée sur l'aliment en tant que tel. Cette étude évalue généralement le produit tel que consommé par l'homme ou l'animal, prenant en compte l'effet de la matrice plante. Cet essai apparaît comme une étude sentinelle pertinente pour détecter les effets non intentionnels qui n'auraient pas été mis en évidence par les autres résultats d'analyses présentés dans le dossier (e.g. moléculaire ou analyse comparative de composition chimique). En effet, l'évaluation des effets potentiellement délétères dus à l'insertion, même si elle s'efforce d'être exhaustive, ne se base que sur des éléments connus. Il en est de même pour l'étude de l'équivalence en substance réalisée sur des composés chimiques définis. L'Afssa recommande donc fortement la mise en œuvre de l'étude de toxicité sub-chronique de 90 jours pour les événements primaires de transformation génétique (AFSSA, 2007).

Pour réaliser cette étude, le pétitionnaire suit les recommandations de la norme OCDE 408 (OCDE, 1998; OECD, 1998), établie à l'origine pour l'évaluation de la toxicité des substances chimiques. Ce protocole est également mis en œuvre pour évaluer les « nouveaux aliments » dans le cadre de leur mise sur le marché selon le règlement 258/97/CE (CE, 1997). Cependant, étant donné que le produit à évaluer n'est pas de même nature (aliment, structure chimique complexe ...), le protocole OCDE nécessite une adaptation aux spécificités de l'OGM.

Afin de mieux cerner les spécificités et contraintes méthodologiques liées à l'administration d'un aliment entier et d'approfondir l'analyse toxicologique, l'Afssa a créé un premier groupe de travail dont l'objectif est d'adapter le protocole standard OCDE 408 à la problématique spécifique des OGM, tant en termes de méthodologie que pour l'analyse statistique et l'interprétation des résultats.

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.

² **Le témoin non-GM (comparator)** se définit comme la plante non génétiquement modifiée ayant le fonds génétique le plus proche possible de la PGM.

Au cours de ce travail, il nous est apparu important de mieux définir les principes de l'analyse statistique des données, les documents guides disponibles ne paraissant pas suffisamment explicites sur ce point. Le protocole OCDE 408, indique par exemple : « s'il y a lieu les résultats numériques devraient être évalués par une méthode statistique appropriée et largement reconnue » (OCDE, 1998).

Un avis antérieur de l'Afssa (annexe 1 du rapport « consommation-OGM ») (AFSSA, 2002) s'était déjà préoccupé de l'évaluation de la puissance des tests statistiques dans les essais d'alimentarité et de tolérance nutritionnelle réalisés sur les animaux cibles. L'EFSA a publié un document formulant des recommandations sur le plan expérimental et l'analyse statistique des données issues des essais en champs, visant à comparer la composition des plantes GM à celle des plantes contrôles (témoin non GM) (EFSA, 2009). Dans son rapport, l'EFSA s'est attaché à faire des propositions sur l'approche statistique à adopter de façon à comparer, sur des paramètres de composition chimique ou de caractères agronomiques, les différences observées entre les deux variétés testées (OGM et quasi-isogénique) avec la variation naturelle que l'on observe entre les génotypes (variétés non transgéniques). Cette variation naturelle peut provenir notamment des interactions plus ou moins importantes des génotypes avec les facteurs environnementaux. Pour cela l'EFSA propose de calculer, dans une approche de test d'équivalence, les bornes de l'équivalence à partir des données obtenues avec les variétés commerciales incluses dans l'essai. L'EFSA souligne que les propositions faites dans son rapport ne s'appliquent pas directement aux données des études réalisées sur des animaux. Une réflexion, au cas par cas, est nécessaire avant leur utilisation dans le contexte d'une étude de toxicité sub-chronique de 90 jours. Dans un rapport relatif à l'évaluation du risque environnemental de la culture des OGM, l'EFSA recommande également d'utiliser des tests d'équivalence pour évaluer l'impact potentiel sur les organismes non cibles (NTO non target organisms) (EFSA, 2010; Perry *et al.*, 2009).

L'Afssa a créé un deuxième groupe de travail centré sur l'analyse des pratiques actuelles en matière de traitement statistique des données de l'étude de toxicité sub-chronique de 90 jours, et leur évolution possible afin d'accroître la pertinence des conclusions. Le groupe, comportant des statisticiens, biostatisticiens et toxicologues, s'est fixé 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 subchronique de 90 jours chez le rat.
- 3) D'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.

La démarche proposée est présentée en détail dans ce rapport et est illustrée à partir d'un cas concret d'analyse des données issues d'une étude de toxicité sub-chronique de 90 jours, réalisée avec les maïs génétiquement modifiés GM portant l'évènement de transformation MON810.

Le rapport est constitué de deux chapitres :

Après un rappel du principe du protocole OCDE 408 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, ce premier chapitre théorique décrit et explique les méthodes statistiques disponibles, leurs intérêts et leurs limites.

Après cet état des lieux, le second chapitre est consacré à l'application de la méthodologie jugée la plus pertinente, aux cas des données issues d'une étude de toxicité sub-chronique réalisée par administration chez des rats de grains de maïs MON810 pendant 90 jours. Une liste de recommandations est finalement proposée.

Chapitre 1 : Méthodes d'analyse statistique des données issues des études de toxicité sub-chronique de 90 jours chez le rat.

1. Plan expérimental des études de toxicité sub-chroniques de 90 jours

1.1. Protocole expérimental 408 de l'OCDE

Toxicité orale à doses répétées-rongeurs: 90 jours

Le protocole OCDE 408 s'inscrit dans une série de protocoles standardisés destinés à évaluer la toxicité des produits chimiquement définis. Les études visant à rechercher des effets par administration répétée correspondent à plusieurs types de protocoles, dont la numérotation renvoie à l'espèce (rongeur ou non-rongeur), à la voie d'administration (orale, cutanée, inhalation) et à la durée du traitement (28 jours, 90 jours ou 2 ans pour l'évaluation de la cancérogénèse).

Dans le cas du protocole OCDE 408, l'espèce est le rongeur, la voie d'administration est orale et la durée de traitement est de 90 jours. Au moins 20 animaux (10 femelles et 10 mâles) doivent être employés pour chaque groupe d'essai. Trois concentrations et un groupe témoin, au moins, doivent être mis en œuvre.

Le produit testé est administré par gavage, via la nourriture ou la boisson. Un essai limite peut être pratiqué quand aucun effet pour une dose de 1000 mg/kg poids corporel/j n'est attendu.

Les effets recherchés sont de type local ou systémique, notamment vis-à-vis des principaux organes cibles comme les poumons et voies respiratoires supérieures, le foie, les reins, le cœur et les vaisseaux sanguins, le cerveau et le système nerveux central.

Les résultats de l'étude incluent :

- ✓ des observations deux fois par jour (morbidité et mortalité) et quotidiennes (examen clinique général),
- ✓ des observations et mesures hebdomadaires (examen clinique complet, prise de poids et consommation de nourriture et d'eau),
- ✓ des examens et analyses (hématologiques, biochimie sanguine, urinaires) en fin d'étude et éventuellement à un temps intermédiaire,
- ✓ un examen ophtalmologique, un examen clinique et une autopsie générale en fin d'étude,
- ✓ de l'histopathologie sur les tissus et organes des animaux témoins et ayant reçu la dose la plus élevée, et aux doses inférieures selon les résultats à forte dose.

Un essai de toxicité sub-chronique de 90 jours chez le rat correctement effectué doit permettre de fournir une évaluation satisfaisante d'un niveau de dose sans effet observé (NOEL³), avec des effets observés faibles ou non reliés à une toxicité (NOAEL ou LOAEL⁴).

La ligne directrice 408 originale a été adoptée en 1981 et a été révisée le 21 septembre 1998. La version révisée accorde davantage d'importance aux effets neurologiques et donne des indications relatives aux effets sur le système immunitaire et sur la reproduction. Elle insiste également sur la nécessité d'observer très attentivement les animaux sur le plan clinique, en vue d'obtenir le plus d'informations possible. Le but de cette révision était de mieux identifier les produits chimiques susceptibles d'avoir une action neurotoxique ou des effets sur le système immunitaire ou les organes reproducteurs, pouvant justifier des études plus approfondies. Concernant l'analyse statistique, le protocole OCDE 408 est non directif.

1.2. Protocole 408 tel qu'il est mis en œuvre dans les études OGM

Une revue des publications⁵ des études de toxicité sub-chronique de 90 jours destinées à démontrer la sécurité d'une plante GM a été réalisée. Cette synthèse permet d'identifier les points communs et les différences dans la méthode appliquée. Les trois aspects suivants sont analysés :

- ✓ Les plans expérimentaux
- ✓ Les paramètres biologiques mesurés
- ✓ La méthodologie statistique employée

³ NOEL no observed effect level ou dose sans effet observé

⁴ NOAEL no observed adverse effect level ou dose sans effet néfaste observé

⁵ Au total, 17 articles publiés dans la littérature scientifique (annexe 1 et 2).

1.2.1. Plan expérimental

Le tableau en annexe 1 résume certains aspects des protocoles utilisés dans 17 articles publiés dans la littérature.

Les différents éléments du protocole expérimental qui ont été examinés concernent notamment :

- le produit testé : espèces, événement de transformation, caractères apportés par l'événement de transformation, les protéines nouvellement exprimées dans la plante, plantes contrôles utilisées ;
- les animaux : nombre d'animaux par groupe, souche de rats utilisées, provenance des animaux, conditions d'hébergement, le centre investigateur ;
- le protocole expérimental : nombres de doses, nombre et période de mesures, nombre et type de groupe.

L'espèce est toujours le rat et la souche la plus souvent « Sprague-Dawley ». Les effectifs sont de 10 à 20 animaux de chaque sexe par groupe, à raison d'un à deux animaux par cage. La mise en œuvre de deux « doses » (11 et 33%) n'est observée que dans 50% des études. Les doses maximales incorporées dans les régimes dépendent de l'espèce végétale, objet du test. Elles sont de 76% pour des grains de maïs, de 60% pour de la farine de riz, de 20% pour des tourteaux de soja et de 10% pour des tourteaux de cotonniers. Les groupes de référence sont multiples et ils incluent toujours les rats nourris aux mêmes doses que l'OGM avec la plante témoin quasi-isogénique (même fonds génétique sans le transgène). Les rats nourris aux mêmes doses avec des variétés commerciales (de 3 à 6) constituent souvent les autres groupes de référence. Les prélèvements et mesures sur les paramètres hématologiques et biochimiques sont réalisés à un ou deux temps (6-7 semaines et 13 semaines). Seuls les résultats en fin d'expérience (13 semaines) sont présentés dans les publications alors que les résultats à mi-parcours de ces mêmes études sont présentés et analysés dans les dossiers déposés par les industriels.

Pour les plantes génétiquement modifiées tolérantes à un herbicide, il est rarement précisé si la plante étudiée a reçu un traitement herbicide au cours de sa culture. Nous avons relevé deux études contenant un groupe supplémentaire correspondant à la plante GM ayant été traitée par l'herbicide pour lequel elle est résistante (Appenzeller *et al.*, 2008, 2009b).

1.2.2. Paramètres biologiques mesurés et leurs significations toxicologiques.

Le terme paramètre désigne les différentes variables d'intérêt qui sont mesurées au cours de l'étude et qui sont l'objet d'analyses statistiques.

Les paramètres pris en compte permettent notamment d'évaluer un effet sur les grandes fonctions et organes cibles majeurs (fonction hématopoïétique, hépatique, rénale, immunitaire, système nerveux central). Classiquement, le nombre de paramètres est de l'ordre de 15 pour l'hématologie, 2 pour la coagulation, 15 pour la biochimie sanguine, 17 pour l'urine, 8 pour le poids des organes et 40 pour les examens histologiques (voir la liste en annexe 5).

1.2.3. Méthodes statistiques utilisées

La revue des méthodes statistiques (tableau en annexe 2) utilisées dans les publications pour analyser les données des études permet de constater que :

- La normalité des distributions et l'homoscédasticité (égalité des variances) sont toujours testées. Les intervalles de confiance sont établis pour chaque variable et chaque groupe. Les distributions des données font l'objet d'une analyse descriptive.
- Les données de poids et de consommation font l'objet d'une représentation graphique, puis d'un test ANOVA sur les poids finaux des animaux ; les modèles mixtes non linéaires adaptés aux données de type longitudinales ne sont jamais mis en œuvre.
- Selon le type de données (et mesures associées), des méthodes paramétriques ou non paramétriques sont utilisées.
- Les tests de différences sont effectués avec des ANOVA en se basant sur des contrastes permettant de comparer les groupes consommant un régime contenant la plante témoin et les groupes consommant un régime à base de plante GM.

- Des données issues des variétés commerciales sont parfois utilisées par le toxicologue en cas de résultat statistiquement significatif du test de différence entre les animaux ayant reçu l'OGM et ceux ayant reçu le témoin. Le toxicologue compare alors les différences entre groupes OGM et témoin aux données des groupes ayant consommé des variétés commerciales. Les données issues des variétés commerciales ne sont pas utilisées si le test de différence est non-significatif et elles n'entrent pas dans l'analyse statistique globale. La même démarche est parfois appliquée à des données historiques.
- Le niveau de risque α est toujours fixé à 5% et les tests sont toujours bilatéraux.
- La multiplicité des tests est quelquefois prise en compte, notamment par l'ajustement des p-values avec le test de Dunn (Dunn, 1964) mais l'ajustement au taux de faux positifs (FDR) n'est jamais effectué, excepté dans le cas de l'étude relative à l'évaluation de la toxicité des tourteaux de cotonniers GM (Dryzga *et al.*, 2007).

Cette revue bibliographique révèle que, les pétitionnaires utilisent des tests de différences paramétriques ou non paramétriques pour ce type d'étude. La puissance de ces tests n'est pas calculée et le test d'équivalence n'est jamais mis en œuvre.

2. Revue des analyses statistiques applicables aux données du protocole 408

Cette section présente les principales méthodes statistiques qui sont ou pourraient être utilisées pour analyser les données issues des études de toxicité sub-chronique de 90 jours chez le rat. Leurs intérêts et leurs limites sont analysés.

2.1. Analyse descriptive des données

Avant de procéder à toute analyse statistique, il est essentiel de visualiser les données mesurées pour évaluer leur qualité, leur variabilité ou leur distribution. Cette visualisation permet également d'identifier les méthodes statistiques les plus appropriées.

2.1.1. Détection de valeurs atypiques

Une *valeur atypique* peut se définir comme une observation qui s'écarte de façon marquée de l'ensemble des autres observations de l'échantillon dans lequel elle apparaît. Les données, surtout lorsqu'elles sont nombreuses, comportent souvent des données aberrantes ou atypiques⁶.

Les valeurs atypiques peuvent concerner un ou plusieurs animaux ou une ou plusieurs variables. Elles peuvent résulter de :

- ✓ La variabilité inhérente au critère biologique mesuré : caractéristique naturelle de la population, celle-ci est difficilement contrôlable
- ✓ L'erreur de mesure : elle est liée à la méthode de mesure et peut être aléatoire ou systématique. Habituellement, on réalise plusieurs analyses par individus et à partir d'un même prélèvement pour réduire l'incertitude.
- ✓ L'erreur d'exécution : mauvaise manipulation de l'animal, erreur de saisie, mauvaise conservation du prélèvement voire erreur d'identification de l'animal.

Lorsque les valeurs atypiques sont attribuables à des erreurs de mesure ou à une erreur d'exécution déterminée, les valeurs peuvent être corrigées ou éliminées.

Ces données atypiques peuvent être une source de biais en déformant l'information obtenue à partir des données brutes. Par exemple, la moyenne peut être fortement influencée par les valeurs atypiques.

Il est donc primordial d'identifier ces valeurs de manière systématique en utilisant un test statistique adapté. Le test de Grubbs (Grubbs, 1969) est un exemple de test qui peut être employé. Il est basé sur la comparaison des valeurs absolues des écarts réduits avec les valeurs théoriques maximums admissibles

⁶ Le terme « données aberrantes » est traduit de l'anglais « Outliers ». On utilisera dans ce document le terme de données atypiques plutôt qu'aberrantes jugé plus approprié au contexte.

pour un risque de première espèce donné (α). Cette méthode est conçue au départ pour la recherche d'une valeur atypique unique, mais elle peut s'adapter à la recherche de plusieurs données atypiques. Il suffit alors de l'appliquer de façon itérative ; après élimination de ou des observation(s) détectée(s) à l'étape précédente, le test est réappliqué.

Cependant, le test de Grubbs ne s'applique qu'au cas où la variable suit une loi normale car il est très sensible à une possible dissymétrie de la distribution parente. Il est possible alors, avant d'appliquer le test, d'utiliser certaines variantes ; ce qui consiste à tronquer symétriquement l'échantillon afin d'obtenir une estimation plus robuste de la moyenne et de l'écart-type.

La troncature n'est pas toujours suffisante pour se rapprocher de l'hypothèse de normalité. Dans le cas de non normalité, il est recommandé d'appliquer le test de Grubbs sur des échantillons tronqués et transformés.

Les données atypiques peuvent aussi être détectées grâce à d'autres tests statistiques se basant sur le principe du test de Dixon (Dixon and Pechmann, 2005). Ce dernier consiste à calculer le ratio défini par: la différence entre la valeur suspecte avec la plus proche valeur observée, divisée par l'étendue de l'ensemble des valeurs observées. Ce ratio est comparé à des valeurs critiques qui dépendent de la taille de l'échantillon. Comme pour le test de Grubbs, le test de Dixon suppose la normalité de la variable étudiée.

Une fois détectées, les données "atypiques" doivent être systématiquement identifiées et présentées graphiquement.

A titre d'exemple, la figure 1 montre les données d'un paramètre sanguin aux deux périodes pour dix animaux mâles. Chaque point du graphe correspond à un animal. Lorsqu'une valeur est identifiée comme atypique, celle-ci est représentée par un point en rouge alors que celles correspondant aux autres animaux sont représentées par des points bleus. Le sexe, le groupe et le numéro du rat concerné sont indiqués au milieu du graphe. La ligne du haut correspond aux mesures à la période 1 (P1) et la ligne du bas à la période 2 (P2). La valeur atypique du mâle 2 du groupe 1 en période 1 a été identifiée grâce à un test de Dixon. Ainsi représentée, la valeur anormalement élevée du mâle 2 du groupe 1 en période 1 est facilement visualisable, et peut être comparée à la valeur mesurée en période 2 qui dans ce cas ne s'écarte pas des autres mesures.

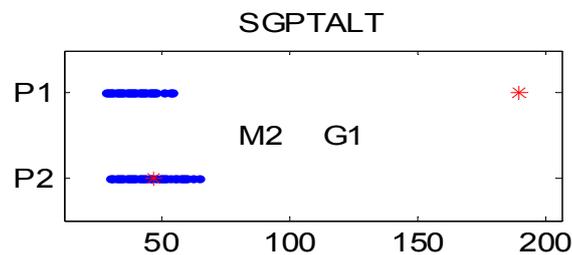


Figure 1 : Valeurs des concentrations en ALAT (SGPTALT) pour les 10 animaux du groupe OGM 11%. Elles sont représentées pour les deux dates de mesure P1 et P2. Les valeurs pour le mâle 2 sont représentées en rouge. Cette valeur est atypique en période 1.

2.1.2. Gestion des valeurs atypiques

Comme indiqué dans le paragraphe précédent, les données atypiques peuvent influencer fortement certaines variables numériques (comme la moyenne et la variance résiduelle d'un groupe) sur lesquelles se base l'analyse statistique. Il est donc important de porter une attention particulière à ces données. Une des solutions consiste parfois à les éliminer du jeu de données. D'une façon générale, la décision de retirer ou non certaines observations du jeu de données ne doit pas découler d'une application aveugle des méthodes statistiques, mais doit rester de la compétence de la ou des personnes qui sont responsables de l'étude (directeur de l'étude). Les données atypiques représentées graphiquement doivent faire l'objet d'une interprétation ou caractérisation d'un point de vue toxicologique afin de différencier les aléas expérimentaux d'une variabilité extrême et rare de réponse au traitement. L'arbre décisionnel de la figure 2 est proposé afin de faciliter cette prise de décision.

Lorsque les données de l'expérience sont nombreuses, un faible nombre de valeurs atypiques n'a que peu d'influence sur les résultats. Elles peuvent donc en principe être éliminées pour se rapprocher des conditions

d'application des tests paramétriques. Ces valeurs atypiques peuvent faire l'objet d'une représentation graphique afin d'en faciliter l'identification et l'analyse.

Lorsqu'un animal présente plusieurs données atypiques, il est recommandé de regarder de près le comportement de l'animal au long de l'étude (consommation, prise de poids, présence/absence de signe clinique) et également le lien entre les variables pour lesquelles l'animal a présenté des valeurs atypiques. A l'issue de cette analyse qualitative, le toxicologue peut conclure que ces observations ne sont pas attribuables au traitement reçu par l'animal concerné. Dans ce cas, les données atypiques peuvent être éliminées.

Dans les cas où il est difficile d'exclure un lien possible avec le traitement leur élimination n'est pas recommandée (figure 2). Des méthodes statistiques spécifiques doivent alors être utilisées : méthodes non paramétriques, transformations des variables ou méthodes d'imputation. La dernière consiste à remplacer les valeurs atypiques ou manquantes par d'autres calculées comme par exemple remplacer les valeurs atypiques par celles prises par certains quantiles.

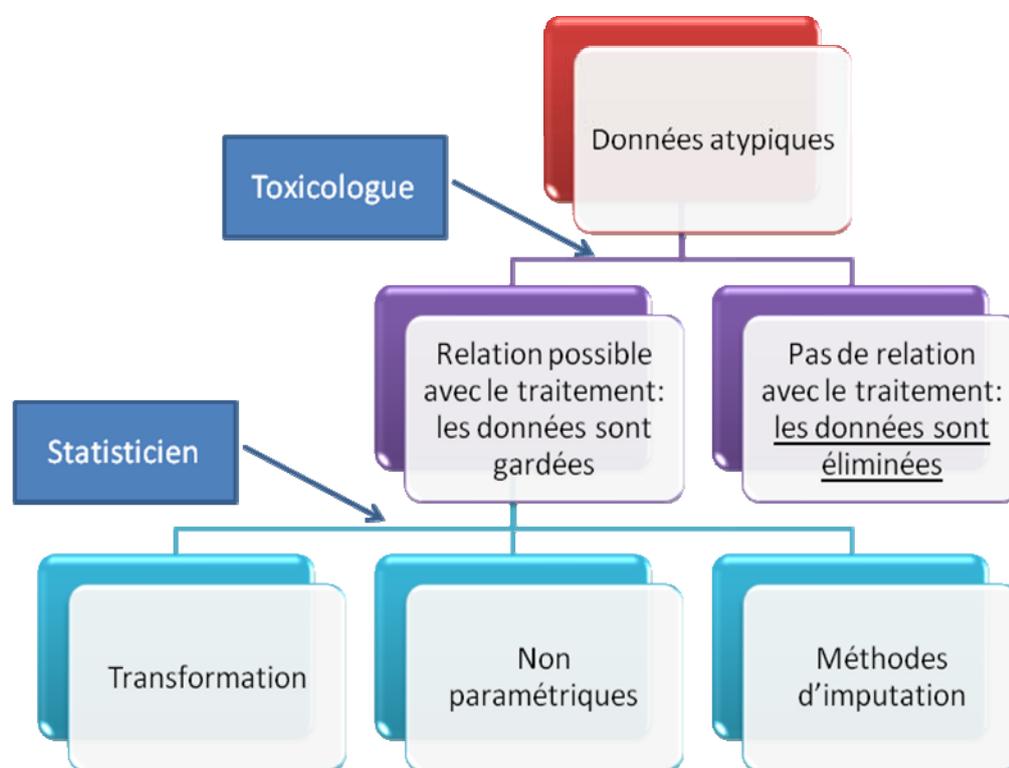


Figure 2 : Arbre décisionnel pour les données atypiques.

Les données dites « atypiques » peuvent avoir une influence importante sur certains paramètres statistiques, comme la moyenne et la variance, et ainsi empêcher l'utilisation des tests paramétriques. Ces données doivent être identifiées et présentées graphiquement pour juger facilement de leur répartition dans les différents groupes de l'expérience (animaux, traitement, période, sexe). La pertinence de leur élimination ou de leur conservation doit être analysée par des experts toxicologues en tenant compte du phénomène à l'origine des valeurs extrêmes. Lorsqu'elles sont conservées, des méthodes statistiques adaptées doivent être utilisées.

2.1.3. Transformation des données

Les tests de comparaisons de moyennes et les calculs d'intervalles de confiance de la moyenne sont basés sur l'hypothèse selon laquelle les distributions des variables étudiées sont normales. Il est à noter que l'hypothèse de normalité n'est pas indispensable lorsque les échantillons sont suffisamment grands pour pouvoir appliquer le Théorème Centrale Limite. Mais, il est d'autant plus important de vérifier cette hypothèse que les tailles d'échantillons sont réduites (comme dans de telles études). Il apparaît souvent nécessaire, de recourir à une transformation des données dans le cas de non normalité, avant de calculer les intervalles de confiance des moyennes.

Une transformation appropriée permet en général d'obtenir des intervalles centrés autour de la moyenne et d'amplitude plus réduite. De nombreuses transformations sont possibles et peuvent être recommandées.

A titre d'exemple :

- une asymétrie positive peut être corrigée avec une transformation de type log ; $y = \log(x - \alpha)$
- une asymétrie négative peut être corrigée avec une transformation de type puissance ; $y = (1 + x/\alpha)^\alpha$

Pour chaque paramètre, la vérification de la normalité et le choix de la transformation sont faites sur l'ensemble des données des groupes expérimentaux.

Le paramètre α peut être obtenu en minimisant la distance de Kolmogorov entre la distribution empirique des données transformées et la distribution gaussienne théorique (i.e. en maximisant la p-value d'un test de Kolmogorov).

Ci-dessous (figure 3), on présente à titre d'exemple la transformation appliquée aux données observées de la variable neutrophile absolue (ABSNEUT) de l'étude de cas MON810 (cf. deuxième partie) selon l'équation : $y = \log(x - 0.22)$

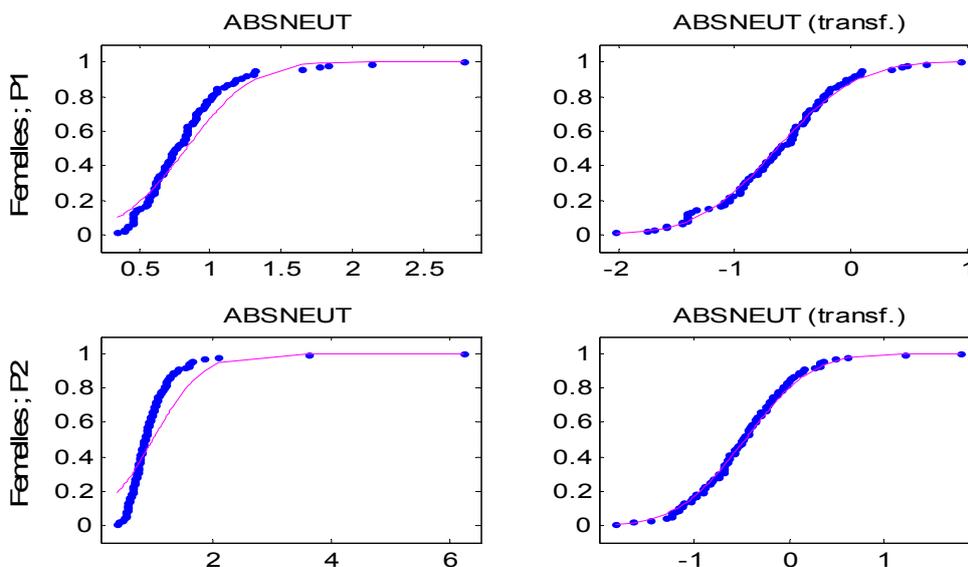


Figure 3 : Fonction de répartition du paramètre ABSNEUT chez les femelles.

P1= période 1 graphes supérieurs et P2 = période 2 graphes inférieurs (en pointillés).

La fonction de répartition observée est représentée en pointillés et la fonction de répartition de la loi normale ajustée sur les données est représentée en trait plein.

L'utilisation des tests paramétriques suppose la normalité de la distribution. Dans ces conditions, il est parfois nécessaire de procéder à une transformation des données pour se rapprocher d'une distribution Normale.

2.2. Analyse des paramètres biologiques hors évolution pondérale.

2.2.1. Test de différences

Les hypothèses testées

Un test statistique peut être considéré comme un outil d'aide à la décision permettant de rejeter ou de valider une hypothèse au vu des observations obtenues sur un échantillon. Le test consiste à opposer deux hypothèses: l'hypothèse nulle (notée H_0) et l'hypothèse alternative (notée H_1). Classiquement, les tests statistiques fonctionnent sur le rejet ou l'acceptation de l'hypothèse nulle qui représente une situation de *statu quo*. Dans le cas des études de toxicité sub-chronique de 90 jours, l'hypothèse H_0 correspond à l'absence d'effet du régime à base d'OGM. En d'autres termes, cette hypothèse suppose que le régime à base de plante GM n'entraîne pas d'effet néfaste sur les rats. Le rejet de l'hypothèse nulle revient donc à rejeter l'absence d'effet du régime à base d'OGM donc à admettre qu'il existe une différence. C'est pour cette raison que ces tests sont appelés « tests de différence ». Mais, le non rejet ou l'acceptation de l'hypothèse nulle ne signifie pas nécessairement l'absence d'effet du régime à base d'OGM ; le non rejet de H_0 peut être la conséquence de l'insuffisance de puissance du test.

Si l'hypothèse alternative H_1 suppose que les deux modalités ont des effets différents sans préjuger du sens de la différence, on parle de tests bilatéraux. *A contrario*, si la modalité expérimentale a un effet supposé négatif ou positif par rapport à la modalité témoin, on parle de tests unilatéraux. En pratique, en l'absence de connaissance *a priori* sur le sens de l'effet, on utilise des tests bilatéraux.

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 un effet de la modalité expérimentale alors que les modalités expérimentales et témoin sont rigoureusement identiques vis-à-vis de leurs effets sur les paramètres biologiques étudiés. C'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 quasi-isogénique.

Le «risque de seconde espèce» (ou risque β) est le risque de ne pas détecter un effet qui existe vraiment. C'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 quasi-isogénique. Le risque de 1^{er} espèce est contrôlé : on le fixe généralement à une valeur faible (souvent 5%). Le risque de 2^{nde} espèce est par contre généralement subi car il dépend de l'importance de l'effet testé.

Le risque β est lié à la puissance P du test. On définit la puissance d'un test par sa capacité à rejeter H_0 si H_1 est vraie, c'est la probabilité de conclure à raison à un effet. La puissance est le complément du risque β ; $P = 1 - \beta$.

Le choix d'un risque α à 5% est fréquent mais arbitraire. Une augmentation de ce risque, par exemple à 10%, permet d'augmenter la puissance des tests (cf paragraphe puissance). Cela augmente les chances de détecter des différences entre les animaux ayant reçu le régime contenant la plante GM et ceux ayant reçu la plante témoin quasi-isogénique. Autrement dit, cela permet d'augmenter la sensibilité de l'expérimentation à détecter des signaux de toxicité potentielle du régime étudié. Cette augmentation de la sensibilité renforce le rôle de l'expertise toxicologique dans l'interprétation biologique des résultats. Le toxicologue aura donc un plus grand nombre de tests significatifs à analyser tout en sachant que parmi eux se trouveront un plus grand nombre de faux positifs. L'analyse du toxicologue est multifactorielle et basée sur la convergence des effets sur plusieurs paramètres biologiques liés ou sur les mêmes paramètres mais dans différentes conditions expérimentales liées (par exemple le même effet sur deux périodes ou aux deux doses). L'augmentation du risque α conduit à une perte de spécificité qui sera compensée par la pertinence de l'analyse toxicologique.

Les tests paramétriques

Les tests paramétriques font l'hypothèse que les observations suivent les distributions de probabilité particulières. Les conclusions des tests paramétriques peuvent être invalides lorsque ces hypothèses faites sur les distributions ne sont pas vérifiées. A titre d'exemple, la validité d'un test paramétrique de comparaison de deux moyennes nécessite la normalité de la distribution de la variable étudiée et l'égalité des variances. L'égalité des variances est une conséquence de l'hypothèse selon laquelle l'effet d'un régime à base d'OGM sur un critère biologique donné se traduit uniquement sur une différence des moyennes observées dans les deux groupes et selon laquelle la dispersion des valeurs individuelles autour de leur moyenne est identique dans les deux groupes.

Nous présentons ci-dessous une procédure pour réaliser des tests de différences lorsque les observations suivent une loi normale. Cette procédure est basée sur un modèle d'analyse de la variance à k facteurs correspondant aux effets fixes de l'essai.

Par exemple :

Le « Régime » : facteur étudié⁷ à deux modalités (le régime transgénique et régime quasi-isogénique)

La « Dose » facteur contrôlé à deux modalités (dose 1 et dose 2)

La « Période » facteur contrôlé à deux modalités, période 1 (5^{ème} semaine), période 2 (14^{ème} semaine)

En général, les analyses sont réalisées pour chacun des deux sexes séparément.

Le modèle posé par sexe sera un modèle linéaire mixte saturé vis-à-vis des effets fixes, avec le facteur « Animal » considéré comme aléatoire⁸ :

$$Y_i = X_i\beta + Z_i\gamma + R_i$$

Y_i représente les valeurs observées chez l'animal i , X_i est la matrice permettant d'assigner l'appartenance aux différents groupes définis par le régime, la dose, et la période.

β représente le vecteur des paramètres décrivant les effets marginaux « Régime », « Dose », « Période » et l'ensemble des effets des interactions :

- ✓ d'ordre 1 « Régime/Dose », « Régime/Période » « Dose/Période ».
- ✓ d'ordre 2 « Régime/Dose/Période ».

$Z_i\gamma$ représente la partie aléatoire du modèle composée de l'effet « Animal » intra régime et dose (noté par la suite « Animal (Régime/Dose) »). L'effet animal est supposé suivre une loi normale centrée de variance σ_a^2 .

R_i représente les écarts au modèle (appelé par la suite l'erreur). Ces écarts sont supposés suivre indépendamment une loi normale de variance σ_e^2 , qui représente la composante de variance de l'interaction « Période*Animal(Régime*Dose) ».

Les contrastes d'intérêt sont les écarts entre régimes conditionnels à l'ensemble des facteurs de l'étude.

A titre d'exemple, dans un dispositif avec deux doses et deux périodes comme celui de l'étude MON810 considérée dans la partie II, nous distinguons les contrastes suivants :

Contraste 1 : "OGM versus ISO dose 1 & période 1"

Contraste 2 : "OGM versus ISO dose 2 & période 1"

Contraste 3 : "OGM versus ISO dose 1 & période 2"

Contraste 4 : "OGM versus ISO dose 2 & période 2"

Ici OGM désigne le groupe de rat ayant consommé le régime contenant la plante GM et ISO désigne le groupe de rat ayant consommé le régime à base de la plante quasi-ISOGénique.

Prise en compte de l'hétéroscédasticité (non égalité des variances).

Le modèle suppose que la variabilité des variables mesurées conditionnellement aux effets fixes est constante (hypothèse « d'homoscédasticité »). Cette hypothèse n'est pas toujours vraie notamment lorsque des mesures sont effectuées pour chaque animal à différentes périodes au cours de l'essai. L'hypothèse d'homoscédasticité peut être testée en modélisant la variance animale intra régime dose qui dépend de la période (test du rapport de vraisemblance). Si le test indique une hétéroscédasticité significative et que les variances diffèrent de façon importante, il est conseillé de poser un modèle hétéroscédastique.

⁷ On appelle facteur « étudié » le facteur qui est la raison d'être de l'essai. Les facteurs « contrôlés » sont des facteurs dont on veut estimer les variabilités afin de diminuer la variance résiduelle et au final accroître la puissance expérimentale de l'essai.

⁸ On appelle facteur « aléatoire » un facteur dont toutes les modalités (i-e les valeurs prises par le facteur) ne figurent pas dans l'essai ; on ne dispose que d'un échantillon représentatif des modalités. C'est le cas ici du facteur « Animal ».

Validation du modèle

Le traitement statistique des données devra inclure la validation du modèle posé. Celle-ci inclut :

- ✓ La vérification du caractère gaussien de la distribution des résidus.
- ✓ L'identification de données influentes⁹ (cf. annexe 3).
- ✓ La réalisation de tests supplémentaires sans ces données influentes.

L'exclusion de ces données devra dans tous les cas être justifiée.

Présentation des résultats des tests paramétriques.

Les résultats fournis pour chacun des tests réalisés et pour chacun des paramètres biologiques analysés comprendront :

- ✓ les écarts entre les moyennes des deux modalités comparées (contrastes entre régimes)
- ✓ les écarts-types des écarts entre moyennes
- ✓ les degrés de signification ou p-value
- ✓ les Carrés Moyens de Référence¹⁰ des tests statistiques

L'expression des écarts observés (taille de l'effet) entre valeurs moyennes des groupes en absolu (dans leurs unités d'origine) et en relatif (par rapport à la valeur moyenne du groupe témoin) peut permettre de nuancer le résultat des tests statistiques et aider l'interprétation en écartant des effets significatifs jugés peu importants mais également en identifiant des effets déclarés non significatifs par les tests mais jugés importants par les toxicologues.

Pour les variables transformées, il est utile d'exprimer les résultats dans leurs unités initiales. L'application directe de la fonction inverse de la transformation n'est pas conseillée. Il est recommandé d'avoir recours aux méthodes de simulation de Monte-Carlo.

Les tests non-paramétriques

Un test statistique non paramétrique est un test qui ne nécessite aucune hypothèse sur la distribution des observations. Ils permettent d'inclure les données atypiques dans l'analyse sans pour autant perturber les résultats. Le recours aux méthodes non paramétriques (type test de Wilcoxon, (Wilcoxon, 1945)) est recommandé lorsque les données «atypiques» n'ont pas été exclues et qu'elles empêchent la normalisation des distributions.

On peut aussi utiliser ces tests pour certaines variables qualitatives dont la nature discrète nécessite de sortir du modèle gaussien (par exemple le paramètre présence/absence de sang dans les urines).

Les tests non paramétriques sont basés sur des comparaisons de rangs ou de médianes, alors que les tests paramétriques comparent les moyennes. Les médianes devront être présentées dans les résultats pour faciliter l'interprétation.

Le test non paramétrique de différence réalisé dans le chapitre II du rapport porte sur l'effet régime conditionnellement au type de groupes générés par le protocole expérimental (selon sexe, période, dose).

Les tests de différence peuvent être paramétriques ou non paramétriques. Les tests paramétriques sont généralement plus puissants. Ils offrent la possibilité de conduire l'analyse de façon plus cohérente avec le plan expérimental de l'étude (par exemple, la prise en compte des données répétées sur un même animal et la possibilité de tester les éventuelles interactions entre les différents facteurs de l'étude). Le recours aux tests non paramétriques se justifie lorsque les conditions d'applications des tests paramétriques (normalité, constance de la variance) ne sont pas satisfaites, même après d'éventuelles transformations de 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. Les tests de différence non significatifs sont souvent incorrectement interprétés car ils ne peuvent pas apporter la preuve de l'absence de différence.

⁹ On appelle donnée influente, une donnée qui a une influence disproportionnée sur un ou plusieurs aspects de l'estimateur d'un paramètre, en particulier les coefficients de régression. Une valeur atypique est souvent une observation influente. (*The Cambridge Dictionary of Statistics*. 2nd Edition ed, ed. C.U. Press. Vol. 121. 2002, Cambridge. 410.)

¹⁰ Les Carrés Moyens de Référence sont les combinaisons linéaires de composantes de variances qui servent de variabilités de référence pour les tests statistiques.

2.2.2. Multiplicité des tests et conséquence sur les risques liés aux tests

La réalisation d'un très grand nombre de tests statistiques conduit *de facto* à augmenter le risque d'erreur de 1^{ère} espèce. Le contrôle de ce risque apparaît donc comme un problème majeur (Hochberg and Tamhane, 1987).

D'importants développements théoriques ont été réalisés depuis une vingtaine d'années pour tenter de le résoudre. De nouveaux critères et de nouvelles procédures ont été définis et proposés dont le False Discovery Rate ou FDR proposé par Y. Benjamini et Y. Hochberg en 1995 (Benjamini and Hochberg, 1995).

Deux critères globaux sont habituellement définis :

Le « **Family Wise Error Rate** » (FWER) est la probabilité de conclure au moins une fois à tort à un effet significatif du facteur étudié.

$$\text{FWER} = \Pr(V \geq 1)$$

Où **V** désigne le nombre de tests significatifs à tort.

Le **False Discovery Rate** (FDR) qui est défini comme l'espérance du nombre de tests significatifs à tort (faux positifs) :

$$\text{FDR} = E \left[\frac{V}{R} \mid R > 0 \right] \cdot \Pr(R > 0)$$

Où **R** désigne le nombre de tests significatifs au niveau de risque α (Benjamini and Hochberg, 1995).

Sous l'hypothèse nulle complète, le FDR est égal au FWER. Si au moins une hypothèse nulle est fautive (« hypothèse nulle partielle »), le FDR est plus faible que le FWER.

Le FDR et le FWER ne contrôlent pas les mêmes risques. Choisir le FDR suppose implicitement que l'on préfère contrôler la proportion de faux positifs plutôt que la probabilité d'avoir au moins un faux positif.

Le FDR offre une solution acceptable au contrôle du risque de faux positifs lorsque le nombre de tests réalisés est très élevé. Dans ce cas, le FWER tend vers 1 et son contrôle entraîne une diminution importante de la puissance des tests (accroissement des faux négatifs). Le FDR est plus facilement contrôlable sans pénaliser de façon aussi excessive la puissance expérimentale.

Ces deux critères représentent des risques globaux qui sont supérieurs aux risques individuels. Le contrôle de ces deux risques globaux est obtenu en corrigeant les degrés de signification des tests individuels (« p-values ajustées ») (Westfall and Young, 1993).

Les caractéristiques de ces deux risques globaux sont décrits en annexe 4.

Une variante du FDR, le « **Positive FDR** » (pFDR) (Storey, 2002a; Storey, 2002b) est définie comme l'Espérance conditionnelle de faux positifs.

$$\text{pFDR} = E \left[\frac{V}{R} \mid R > 0 \right]$$

On définit alors la qValue comme la mesure analogue au degré de signification ajusté appliquée au pFDR (FDR¹¹).

Etant donné un test donné X_k , de degré de signification p_k , la qvalue est la plus petite valeur du FDR estimé pour laquelle le test peut être déclaré significatif à un niveau de risque global α .

$$\text{qvalue}(p_k) = \min_{p_k \leq \alpha} \{ \hat{\text{FDR}}(\alpha) \}$$

Cette statistique proposée par Storey (Storey, 2003) permet, par une inversion du raisonnement par rapport à la démarche classique, de fixer *a priori* le risque de première espèce au niveau des tests individuels et d'en déduire le FDR minimum pour ce niveau de risque.

Intérêt du contrôle du pourcentage de conclusions erronées (FDR).

Le False Discovery Rate (FDR) trie les tests statistiques qui restent significatifs lorsque l'on contrôle le pourcentage d'hypothèses nulles rejetées à tort à un niveau de risque (au sens du FDR) fixé *a priori*.

¹¹On considère le nombre de tests suffisamment grand pour qu'il y ait au moins une hypothèse nulle rejetée. Le pFDR est alors équivalent au FDR.

Les degrés de signification corrigés selon la procédure de Benjamini (Benjamini and Hochberg, 1995) au FDR permettent de conclure à un effet significatif pour tous les tests dont le degré de signification ajusté est inférieur au FDR tout en assurant un pourcentage de tests significatifs à tort, au plus égal au FDR.

Cette démarche très décisionnelle s'accompagne d'une démarche plus exploratoire proposée par Storey *et al.* (Storey, 2002a, 2003). Les q-values calculées par cette méthode pour chaque degré de signification permettent pour un niveau du risque de première espèce individuel (i-e au niveau de chaque test individuel) fixé *a priori*, d'en déduire le pourcentage de tests rejetés à tort (espérance de faux positifs).

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, telles que décrites ci dessus, sont largement utilisées lorsque le nombre de test est particulièrement élevé. Le calcul de l'un de ces critères globaux (FDR, q-Value) pour les données des études 90 jours permet d'estimer le taux de faux positifs dans les résultats.

2.2.3. Puissance

Le test de différence est classiquement utilisé pour vérifier un effet bénéfique physiologique ou pharmacologique, il est aussi utilisé de façon systématique dans les études de sécurité des produits chimiques, médicaments, des nouveaux aliments ou de produits cosmétiques afin d'identifier des effets néfastes. Dans le premier cas, l'expérience est mise en œuvre pour apporter la démonstration d'un effet positif. Un test de différence non significative n'est pas forcément dû à une absence de différence réelle au niveau des populations cibles; il peut également être dû à un manque de puissance induit par une variabilité importante de la mesure du paramètre ou par des échantillons de petites tailles. Il est donc utile de compléter les résultats d'un test de différence non significatif par un calcul de puissance pour vérifier que le test non significatif ne résulte pas d'une puissance trop faible.

La puissance d'un test statistique peut se définir comme la probabilité de rejeter l'hypothèse nulle à raison, c'est à dire la probabilité de détecter une vraie différence entre deux groupes de rats. Le calcul de la puissance d'un test requiert les données suivantes :

- le risque de première espèce α consenti,
- la taille d'effet que l'on veut pouvoir mettre en évidence,
- la variabilité des observations exprimées par la variance ou l'écart-type,
- la taille des échantillons (nombre d'animaux par groupe).

Les paramètres qui peuvent accroître la puissance d'un dispositif expérimental sont donc :

- une augmentation du risque α (par exemple de 5% à 10%),
- une augmentation de la taille de l'échantillon
- un meilleur contrôle des sources de variabilité (sexe, âge, période de mesure, dose ou qualité des mesures).

On peut aussi améliorer la puissance des tests en modifiant le protocole expérimental, par exemple en ajoutant un groupe traité à une dose intermédiaire entre la dose faible et la dose la plus forte.

La notion de taille d'effet (*effect size*) est encore très peu utilisée dans les rapports des études d'évaluation des risques liés aux OGM. Or la détermination de la taille de l'effet observé met en avant l'importance concrète des effets mesurés, témoignant de leur éventuelle signification biologique. Celle-ci peut permettre de nuancer ou de renforcer le résultat du test statistique.

La taille d'effet est une mesure qui évalue de combien on s'écarte de l'hypothèse nulle, $\mu_1 - \mu_2$ dans le cas de la comparaison de deux moyennes. La taille de l'effet que l'on souhaite mettre en évidence peut être déterminée par le toxicologue en fonction de considérations biologiques. Le statisticien peut demander aux biologistes ou aux toxicologues à partir de quelle taille d'effet ou différence, celle-ci devient biologiquement significative. On peut utiliser des tailles d'effet standardisées (différences entre moyennes exprimées en pourcentage de l'écart type). Cette standardisation peut être intéressante lorsque l'unité de la variable n'a pas de signification particulière mais aussi lorsque les variables ont des unités très différentes. Ceci permet aussi de comparer ces effets lorsque la variabilité d'un même paramètre est différente d'un échantillon ou d'une expérience à l'autre. Dans le cadre de comparaison de deux moyennes, la différence entre les deux moyennes est divisée par l'écart-type commun (variabilité intra-groupe). La taille d'effet standardisée est parfois appelée *d* de Cohen (Cohen, 1992).

Cohen propose l'interprétation suivante de la taille de l'effet :

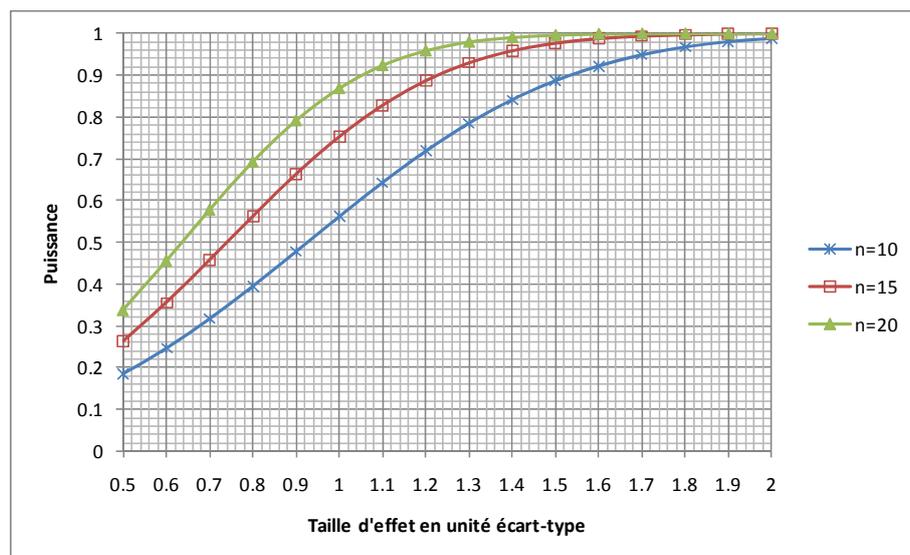
- d = 0,2 indique un léger effet
- d = 0,5 indique un effet modéré

$d = 0,8$ indique une grande taille d'effet

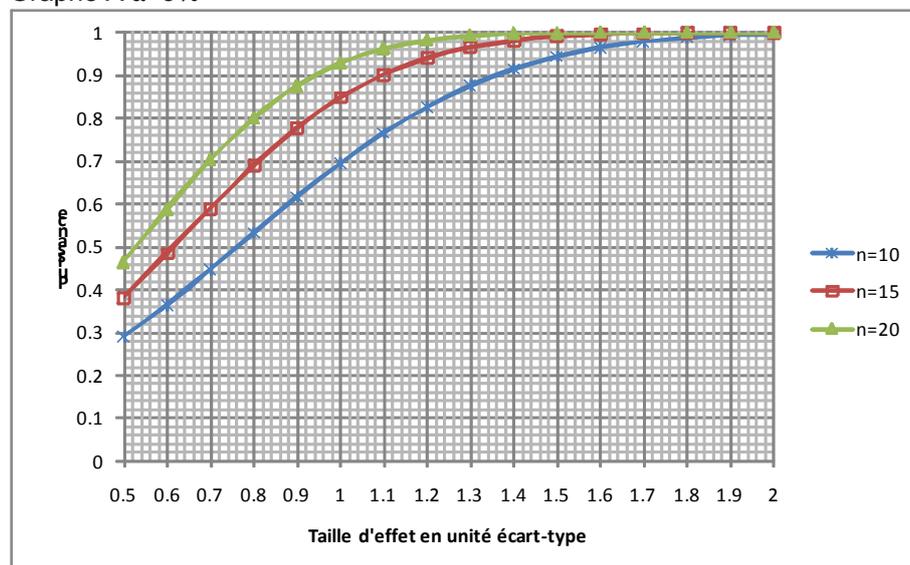
Sur la base des connaissances des experts en toxicologie, il sera utile de proposer ce type de catégorisation de la taille des effets.

Une analyse de puissance peut être faite de différentes manières selon les données disponibles :

- ✓ *a priori*, si le risque α , la taille d'échantillon et la taille d'effet sont fixés, on peut alors calculer la puissance. Ceci permet de voir si le dispositif expérimental est bien adapté (figure 4).
- ✓ *a priori*, si la puissance souhaitée (souvent 80%), le risque α et la taille d'effet à détecter sont fixés ; on peut alors calculer le nombre d'animaux nécessaire pour atteindre cette puissance (figure 5).
- ✓ *a posteriori* si les résultats de l'expérimentation sont disponibles, on peut calculer la taille de l'effet qui peut être mise en évidence par le test avec une probabilité donnée (par exemple 80%) (figure 4).



Graph A $\alpha=5\%$



Graph B $\alpha=10\%$

Figure 4: Puissance du test de comparaison de deux moyennes en fonction de la taille d'effet. n est le nombre d'animaux par groupe.

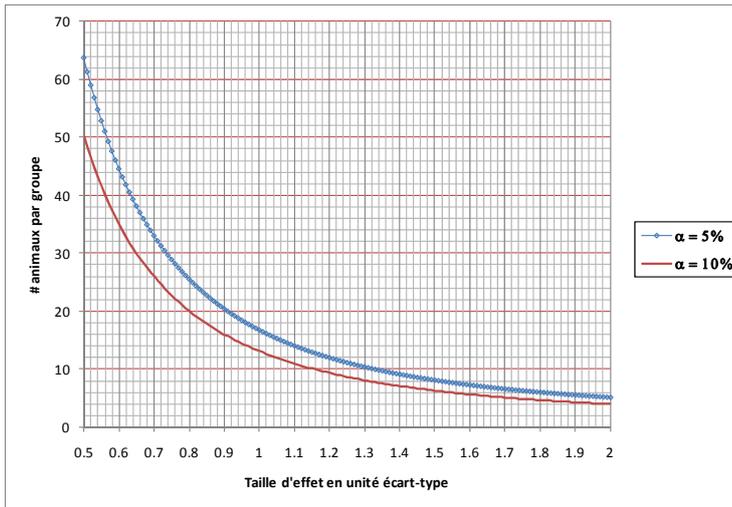


Figure 5: Nombre d'animaux nécessaire en fonction de la taille d'effet pour une puissance de 80%. α correspond à l'erreur de type I fixée ou p-value.

Ainsi selon la figure 4, graphe A ($\alpha=5\%$), avec un effectif de 10 animaux, il est possible de détecter un écart entre les moyennes comparées de 1,3 écart type pour une puissance de 80%.

La figure 5 montre que cet écart est de 1,16 lorsque le risque α est fixé à 10%.

Il est 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 atteindre une puissance jugée suffisante. Cependant, si la taille de l'échantillon est un facteur limitant, l'analyse de la puissance doit être présentée *a posteriori* telle que décrite ci dessus.

En particulier, on présentera la taille de l'effet qu'il est possible de détecter pour une puissance fixée à 80% afin de vérifier que le protocole mis en œuvre est capable de détecter un effet biologiquement significatif.

2.2.4. Test de non différence ou d'équivalence

Principe général

Dans une étude de toxicité sub-chronique de 90 jours par administration répétée de plante GM, l'expérience est mise en œuvre pour démontrer la non toxicité de l'alimentation contenant des OGM ; cette conclusion peut être prononcée au vu d'un résultat négatif (absence de différence). Mais cette approche n'est pas satisfaisante car un test de différence non significatif n'est pas la preuve de l'absence de différence. Les tests d'équivalence sont plus pertinents lorsque l'objectif de l'étude est d'apporter les preuves de non toxicité.

Les tests d'équivalence ont été initialement décrits par Westlake en introduisant la notion d'intervalle de confiance symétrique pour les études de bioéquivalence (Anderson and Hauck, 1983; Westlake, 1976). Schuirmann proposera plus tard d'utiliser le double test d'équivalence « TOST procedure » (Schuirmann, 1987). Les tests de bioéquivalence sont notamment mis en œuvre pour comparer un médicament dit princeps avec un générique que l'on souhaite substituer au premier. Dans ce cas, il est impératif de démontrer une équivalence biologique entre les deux médicaments. Depuis 1997 pour la FDA et 1998 pour l'EMA, des lignes directrices ont été publiées (EMA, 1998, 2008): l'équivalence est jugée sur un nombre limité de 3 critères cinétiques et l'intervalle de confiance (IC) 90% est construit avec des bornes d'équivalence entre 80% et 125%. L'élargissement de l'IC aux bornes 0,75 – 1,33 est admis s'il existe une forte variabilité intra-individuelle.

D'autres disciplines, comme l'écologie (Dixon and Pechmann, 2005) et la pathologie végétale (Garrett, 1997) font appel aux tests d'équivalence. Ainsi, Chen *et al.* (Chen *et al.*, 2010) s'appuient sur l'équivalence pour comparer différents procédés de nettoyage lors de la fabrication de produits biopharmaceutiques. Là encore, il s'agit de s'assurer de la substitution d'un procédé par un autre, avec les mêmes garanties d'efficacité.

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. Ces plantes sont 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 ces applications, le test de différence est mis en œuvre pour vérifier si la plante GM diffère de son équivalent, auquel cas cette différence pourrait éventuellement constituer un danger. Le test d'équivalence est proposé pour déterminer si les caractéristiques agronomiques, phénotypiques et de composition s'inscrivent dans la fourchette de variation naturelle.

Le test d'équivalence se base sur une hypothèse nulle contraire à celle du test de différence. L'hypothèse nulle H_0 est : « il existe une différence entre les traitements, ils ne sont pas équivalents ». Comme le test de différence, le test d'équivalence peut être à l'origine d'erreur de décision, mais les définitions de ces erreurs sont différentes car l'hypothèse H_0 de départ est inversée.

L'absence d'équivalence (hypothèse nulle) signifie dans ce contexte que le régime OGM a un effet sur le paramètre biologique étudié. Cet effet est traduit par l'existence d'une différence entre la moyenne du groupe ayant reçu l'OGM et celle du groupe témoin, supérieure à une valeur préétablie Δ . Le rejet de l'hypothèse nulle (absence d'effet supérieur à Δ) conduit à conclure à l'équivalence. Comme pour le test de différence, ne pas rejeter l'hypothèse nulle ne signifie pas que H_0 est vraie (le régime OGM a un effet sur le paramètre biologique étudié).

Le choix de la valeur seuil Δ est primordial dans les tests d'équivalence. Ce seuil peut être interprété comme étant la taille d'effet limite considérée comme non intéressante sur le plan biologique ou toxicologique « limits of concern ».

Le test d'équivalence se base sur la définition d'une ou de deux bornes Δ_{INF} et Δ_{SUP} pour chaque paramètre : Si le risque correspond à l'évolution du paramètre dans un sens uniquement (augmentation ou diminution dans les tests d'infériorité ou de supériorité), la définition d'une seule borne Δ suffit. La définition des bornes peut être réalisée de trois manières :

- par expertise sur la base des connaissances des toxicologues,
- en utilisant un niveau de tolérance arbitraire appliqué systématiquement à tous les paramètres (e.g, $\pm 10\%$ de la valeur estimée de la différence),
- en utilisant les mesures obtenues sur des rats nourris avec des régimes contenant différentes variétés commerciales de la même espèce que l'OGM,
- en utilisant les données historiques.

Concrètement, le test consiste à vérifier si les intervalles de confiance des différences entre les moyennes (e.g. entre la moyenne du groupe OGM à la dose 1 et la moyenne du groupe ISO à la dose 1) dépassent ou non les bornes Δ_{INF} ou Δ_{SUP} . Si l'intervalle de confiance de la différence entre les deux moyennes n'inclut pas les bornes Δ_{INF} ou Δ_{SUP} alors on rejette l'hypothèse de non équivalence et on conclut à l'équivalence pour le paramètre biologique étudié.

La définition des bornes Δ_{INF} et Δ_{SUP} par paramètre biologique est délicate. L'approche qui consiste à définir un niveau de tolérance systématique de ± 5 ou $\pm 10\%$ est arbitraire.

Lorsque l'expérience inclut des groupes de rats nourris avec différentes variétés commerciales, il est possible d'utiliser ces traitements supplémentaires pour définir les bornes Δ_{INF} et Δ_{SUP} . Une approche simple consiste à calculer toutes les différences entre, d'une part les moyennes obtenues avec la variété témoin quasi isogénique et, d'autre part, les moyennes obtenues avec les variétés commerciales, puis de fixer Δ_{INF} et Δ_{SUP} aux valeurs minimales et maximales de ces différences. L'avantage de cette approche est qu'elle ne fait aucune hypothèse sur la distribution des mesures, mais elle ne tient compte ni des erreurs de mesures, ni du fait que les variétés commerciales utilisées ne constituent qu'un échantillon de l'ensemble des variétés commerciales disponibles.

Une autre alternative consisterait à modéliser les variabilités intra et inter-variétés à l'aide d'un modèle probabiliste (e.g, un modèle mixte), il sera alors nécessaire de faire des hypothèses sur la distribution des mesures.

Enfin, les bornes peuvent être calculées à partir des données historiques. Celles-ci correspondent aux données obtenues pour des animaux de la même souche, élevés dans le même centre investigateur que

celui où a été réalisée l'étude. Ces données ne sont pas toujours fournies dans le dossier industriel mais elles existent toujours.

L'EFSA (EFSA, 2009) a proposé une méthode alternative pour calculer les seuils de tolérance. Ils sont calculés à partir d'une modélisation de la variabilité inter variétés basée sur un modèle mixte incluant un effet aléatoire "variété commerciale". Cette méthode a été proposée pour analyser des paramètres agronomiques de la culture et de compositions chimiques de la plante (e.g., teneur en protéines des grains) susceptibles de fluctuer fortement entre variétés. Dans ce contexte, l'application de la procédure EFSA conduit souvent à des intervalles d'équivalence larges (cf page 23 et 24 du rapport EFSA). Cette méthode a été conçue pour analyser des paramètres décrivant des caractéristiques variétales mais ne semble pas être adaptée à l'analyse des essais sur animaux de laboratoire. En effet, les études toxicologiques s'intéressent aux caractéristiques des animaux et non à la composition chimique des plantes. Même si la composition finale des régimes alimentaires peut varier du fait de l'introduction de variétés différentes dans les aliments (la composition varie d'une variété à une autre au sein d'une même espèce), les différences de composition n'induisent pas nécessairement des différences sur les paramètres biologiques mesurés sur les rats. On peut ainsi supposer que ce facteur aura peu d'influence sur le résultat des paramètres biologiques mesurés.

Calcul de l'eRg, l'Equivalence range

Une autre façon de conduire le test d'équivalence consiste à calculer l'eRg, ou amplitude (écart) d'équivalence (Tryon and Lewis, 2008). Il s'agit de l'écart maximum que l'on observe entre deux moyennes en tenant compte des intervalles de confiance autour des moyennes calculées (figure 6).

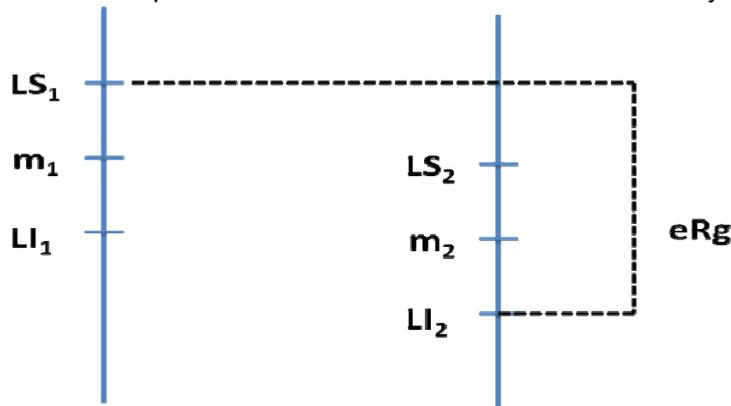


Figure 6 : Principe de l'équivalence range, l'eRg.

LI et LS représentent les limites inférieures et supérieures de l'intervalle de confiance calculé pour chacune des moyennes selon la formule corrigée de (Tryon and Lewis, 2008).

L'eRg représente le plus grand écart entre les extrêmes des intervalles de confiance corrigés (intervalle de confiance « hybride » entre intervalle de confiance descriptif et inférentiel) autour des moyennes des groupes comparés.

Les intervalles de confiances sont calculés au risque 2α . Si le test d'équivalence est fait avec un risque α de 10% alors les intervalles de confiances sont calculés avec un niveau de confiance de 80%.

On conclut à l'équivalence lorsque l'eRg est inférieur au seuil d'équivalence Δ .

Intérêts et limites du test d'équivalence

Le test d'équivalence a deux intérêts principaux :

- Le test d'équivalence définit explicitement la notion d'équivalence à partir des connaissances biologiques disponibles. Ceci permet d'apprécier la taille d'effet insignifiante sur le plan biologique.
- Le test d'équivalence est basé sur une hypothèse nulle de non-équivalence inverse au test de différence ce qui permet d'éviter toute conclusion erronée du type « un test de différence non significatif démontre l'absence d'effet ». Ce test est complémentaire du test de différence : lorsqu'il est significatif, il permet de conclure à l'équivalence ; les « faux négatifs » dans ce test correspondent à de « fausse » absence d'équivalence alors qu'ils correspondent à de « fausse » absence de différence en test de différence.

Toutefois, comme vu précédemment, la faisabilité d'un test d'équivalence dépend essentiellement de la définition du seuil de tolérance. Le choix de ce seuil reste délicat, il peut être basé sur les données historiques ou sur les données provenant d'un ou plusieurs régimes à base de variétés commerciales.

De plus, la démonstration de l'équivalence ne peut être obtenue que lorsque la puissance est suffisante ce qui nécessite d'augmenter les effectifs des groupes. Ainsi, la figure 7 présente le nombre de sujets nécessaire permettant de conclure à l'équivalence avec une puissance de 80% en fonction du seuil d'équivalence fixé (ici Δ est exprimé en unité écart-type intra-groupe et est fixé à 1 ou 1.5), du risque α et de la vraie différence entre les deux moyennes comparées. On observe que le nombre d'animaux par groupe augmente de façon exponentielle lorsque la vraie différence observable se rapproche du seuil de tolérance.

Pratiquement, si par exemple les différences observables sous l'hypothèse d'équivalence sont autour de 1 écart-type, le test d'équivalence nécessitera d'inclure plusieurs centaines d'animaux par groupe si le seuil d'équivalence (tolérance) a été fixé à un écart-type et environ 50 animaux par groupe dans le cas d'une tolérance fixée à 1.5 écart-type. Inversement, si le nombre de sujets par groupe est de 20, la tolérance devra être comprise entre 1 et 1.5 écart-type pour observer une vraie différence de 0.5 écart-type.

Dans l'exemple MON810 (cf. le chapitre 2), plus de 65% des différences observées sont inférieures à 0.5 écart-type et 80% sont inférieures à 0.75.

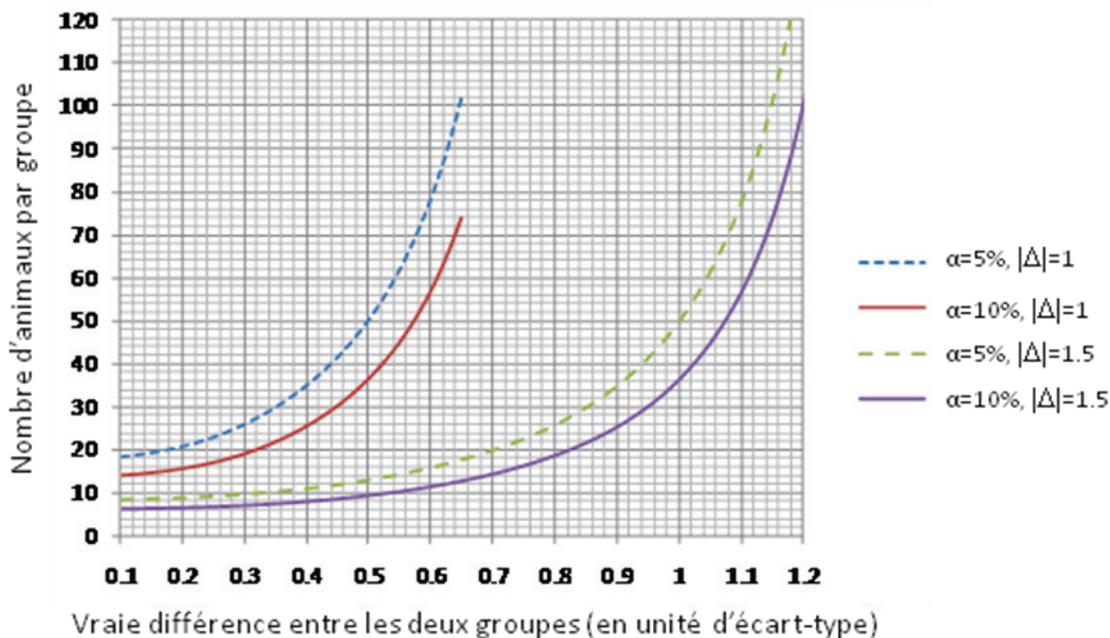


Figure 7: Nombre d'animaux par groupe pour le test d'équivalence en fonction du seuil de tolérance (Δ) et de la vraie différence entre les deux groupes comparés (taille d'effet).

La puissance = 80%

Comme pour les tests de différences, la multiplicité des tests d'équivalence pose également le problème de la gestion du risque global. Il n'existe pas actuellement d'équivalent du FDR pour les tests d'équivalence.

Le test d'équivalence permet de définir la notion d'équivalence à partir des connaissances biologiques disponibles. Il est complémentaire du test de différence et permet d'éviter toute conclusion erronée du type « un test de différence non significatif démontre l'absence d'effet ».

Une réflexion dédiée est nécessaire pour définir les seuils de tolérance pour chaque paramètre. Etant donnée la difficulté de détermination du seuil d'équivalence, il est proposé de calculer les eRg pour toutes les comparaisons prévues dans le protocole expérimental.

2.3. Analyse de l'évolution pondérale et de la consommation hebdomadaire.

Les données de poids et de consommation de tous les animaux sont mesurées chaque semaine: ce sont des données longitudinales.

Nous avons vu dans l'analyse bibliographique que souvent, les données de poids et de consommations sont traitées comme les autres données par des tests de différence (ANOVA) sans tenir compte des répétitions des observations sur les mêmes animaux.

Or, il est plus adapté d'analyser ces données selon les étapes suivantes:

- Tracer les courbes de croissances individuelles pour apprécier visuellement les formes de ces courbes, détecter des évolutions nettement différentes (comme par exemple, la différence entre sexe).
- Choisir un modèle mathématique pour décrire la courbe de croissance : Il s'agit le plus souvent de **modèle non linéaire** (exemple : le modèle de Mitscherlich, Gompertz...).
- Choisir le modèle statistique permettant d'intégrer les conditions expérimentales et de tester l'effet des traitements sur la croissance des animaux.
Par exemple dans le modèle non linéaire suivant :

$$f(t) = C + (A - C) e^{-Bt} \text{ (Modèle Mitscherlich)}$$

- le modèle d'ajustement statistique doit être clairement présenté en spécifiant les paramètres considérés aléatoires (entre les animaux) ou fixes, les lois de distribution des paramètres aléatoires, la prise en compte ou pas de corrélation entre les paramètres du modèle, et enfin préciser sur quels paramètres les effets du traitement sont testés (sur A, B ou C).
- Evaluer la qualité d'ajustement du modèle et analyser les résidus.

A titre d'exemple, considérons le modèle de Mitscherlich :

$$f(t) = C + (A - C) e^{-Bt}$$

où A et C représentent respectivement les poids initiaux et les poids limites, alors que B représente le taux de croissance.

Pour expliquer la variabilité des courbes de croissance observées, on considère que les paramètres du modèle sont des paramètres individuels. On modélise alors la série d'observations par :

$$Y_{ij} = C_i + (A_i - C_i) \cdot e^{-B_i t_{ij}} + e_{ij}$$

Avec Y_{ij} : poids de l'animal i au temps t_{ij}

A_i : Poids de l'animal i au début du traitement

C_i : Poids limite de l'animal i

B_i : Taux de croissance de l'animal i

e_{ij} : Erreur résiduelle supposée suivre une loi Normale, $e_{ij} \sim N(0, \sigma^2)$

Les courbes de croissance des rats dans les études de toxicité subchronique de 90 jours sont différentes suivant le sexe (poids et consommation chez les femelles plus faible que chez les mâles en général). L'analyse est donc faite par sexe.

Aussi bien pour les mâles que pour les femelles, le modèle suppose que les paramètres A, B et C sont variables d'un animal à un autre selon les caractéristiques suivantes :

$$\ln(A_i) = \ln(A) + \eta_{Ai}$$

$$\ln(B_i) = \ln(B) + \beta_B T_i + \gamma_B D_i + \eta_{Bi}$$

$$\ln(C_i) = \ln(C) + \beta_C T_i + \gamma_C D_i + \eta_{Ci}$$

Avec,

T_i : Traitement de l'animal i.

D_i : Dose reçue par l'animal i si prise en compte de deux doses dans l'essai.

Les effets aléatoires sont $(\eta_{Ai}, \eta_{Ci}, \eta_{Di})$ des vecteurs aléatoires gaussiens de moyenne nulle et de matrice de variance covariance Ω .

Les paramètres de population du modèle à estimer sont les effets fixes $(A, B, C, \beta_B, \beta_C, \gamma_B, \gamma_C)$ et la matrice de covariance Ω' .

Les tests à zéro des paramètres suivants : $\beta_B, \beta_C, \gamma_B, \gamma_C$, permettent de tester les effets du régime sur les paramètres de croissance.

Du fait de la répétition dans le temps des mesures pondérales, l'utilisation des modèles mixtes adaptés aux données répétées sur le même animal est nécessaire pour l'analyse des courbes de croissance. Ce type de modèles a fait l'objet de différents travaux théoriques et appliqués et divers outils informatiques sont disponibles pour faciliter la mise en œuvre de cette démarche. Compte tenu de leur complexité, le choix et l'exécution de ces modèles doivent être documentés.

3. Présentation et aide à l'interprétation

Les méthodes statistiques décrites ci-dessus constituent une aide à la décision pour le toxicologue. Elles permettent de réaliser un tri parmi les nombreuses données générées afin que le toxicologue identifie rapidement celles qui présentent un intérêt pour l'interprétation finale de l'expérience.

Il est donc primordial que les résultats de l'analyse statistique soient présentés avec un maximum de clarté et de visibilité pour le toxicologue. Les représentations graphiques sont des outils qui permettent de présenter les résultats ou un résumé pertinent de ces résultats.

La chapitre 2 de ce rapport donne des exemples des outils graphiques qui peuvent être utilisés.

Chapitre 2 : Application des méthodes statistiques aux données d'une étude de toxicité sub-chronique chez le rat par administration répétée via l'alimentation d'une plante GM pendant 90 jours.

1. Introduction

Cette partie du rapport vise à appliquer les méthodes statistiques présentées dans le chapitre précédent aux données de l'étude réalisée dans le cadre de la demande d'autorisation de mise sur le marché européen des maïs MON810.

Les données utilisées et les méthodes statistiques appliquées sont décrites dans la section 2. Les résultats sont présentés dans la section 3. La section 4 résume les principales conclusions déduites de l'analyse statistique. Enfin, la section 5 présente une interprétation toxicologique des résultats.

2. Matériels et méthodes

2.1 Protocole expérimental de l'étude.

L'étude¹² analysée dans cet exemple correspond à celle présentée dans le dossier industriel à l'appui de la demande de renouvellement d'autorisation de mise sur le marché des maïs portant l'événement de transformation MON810 au titre du règlement (CE) N°1829/2003 (application EFSA-GMO-RX-MON810).

L'étude de toxicité sub-chronique de 90 jours repose sur les lignes directrices de l'OCDE (protocole OCDE 408) avec une comparaison entre les paramètres observés sur des rats nourris avec un régime contenant du maïs MON810 (OGM) et ceux observés sur des rats nourris avec un régime contenant le maïs témoin quasi-isogénique (ISO). Le protocole inclut également 6 groupes de rats nourris avec un régime contenant une variété de maïs commerciale.

Quatre cent rats âgés de 6 semaines, 200 mâles et 200 femelles, ont été répartis aléatoirement en 10 groupes de 20 animaux :

- Groupe 1 : animaux nourris avec du maïs MON810 noté OGM, à hauteur de 11% du régime.
- Groupe 2 : animaux nourris avec du maïs témoin quasi-isogénique noté ISO, à hauteur de 11% du régime.
- Groupe 3 : animaux nourris avec du maïs OGM, à hauteur de 33% du régime.
- Groupe 4 : animaux nourris avec du maïs ISO, à hauteur de 33% du régime.
- Groupes 5 à 10 : animaux nourris avec une des 6 variétés commerciales du maïs correspondant notées de VC1 à VC6, à hauteur de 33% du régime.

La dose 33% correspond à la dose de grains de maïs maximale habituellement incorporée au régime des rongeurs sans créer de déséquilibre alimentaire. La composition chimique des grains et des régimes de maïs administrés aux animaux était non significativement différente entre les grains de maïs OGM et ISO. Les régimes ont été complétés et équilibrés en termes de contenu protéique et lipidique (régime standard PMI Certified Rodent Diet #5002).

Le comportement des animaux a été observé quotidiennement. Ils ont été pesés une fois par semaine. Sur 10 des 20 animaux, par groupe et par sexe, des prélèvements sanguins et urinaires ont été effectués à la fin de la 5^{ème} et de la 14^{ème} semaine pour des analyses biologiques. En fin d'expérience, les paramètres de coagulation sanguine, le poids des organes et leur analyse histologique ont été réalisés. Au total 109 paramètres ont été mesurés.

2.2 Gestion des données

Les données de l'étude MON810, disponibles sous format papier, ont été scannées et transférées dans des feuilles Microsoft Excel grâce à un logiciel de reconnaissance des caractères (Readiris®) puis les données ont été vérifiées (contrôle visuel, calcul de moyennes et d'écart types par groupe et par sexe puis comparaison aux valeurs du rapport de l'étude). L'ensemble des tableaux a été transféré dans deux tableaux SAS®, l'un sur les données pondérales et l'autre sur les données non pondérales.

¹² Lemen and Dudeck, 19 december 2001, Monsanto Company : 13 week feeding study in rats with grain from YieldGard (MON810) corn grain (DK551 Bt preceded by a 1-week baseline food consumption determination with PMI certified rodent diet # 5002).

2.3 Analyse statistique

Elle a portée sur 83 paramètres dont 67 sont présentés en annexe (annexe 5) ; les 16 autres sont des paramètres de croissance. Les paramètres histologiques n'ont pas été pris en compte car les méthodes statistiques proposées ne s'appliquent pas directement à ces données qualitatives.

- **Recherche des données atypiques**

Les valeurs centrales de chaque paramètre ont été analysées en calculant la moyenne, la médiane et le mode. La dispersion des valeurs de chaque paramètre a été décrite en calculant l'étendu, l'écart-type et les centiles. Des diagrammes de dispersion des valeurs par groupe et par sexe ont été construits afin de détecter d'éventuelles observations atypiques.

Une procédure automatique de recherche d'observations atypiques, basée sur le test de Dixon a aussi été appliquée. Elle consiste à calculer le ratio défini par la différence entre la valeur suspecte et la valeur observée la plus proche divisée par l'étendue de l'ensemble des données.

Les observations atypiques détectées ont été présentées aux toxicologues pour déterminer si ces données pouvaient être écartées de l'étude du fait de l'absence d'imputabilité des effets en relation avec le traitement et de contraintes statistiques conditionnant l'usage de certains tests.

Si les données atypiques doivent être conservées, ces valeurs peuvent invalider l'usage des tests paramétriques. Deux options sont alors considérées : i) transformer les variables ou ii) avoir recours aux tests non paramétriques.

- **Analyse de la normalité des données et transformation des paramètres si la distribution n'est pas gaussienne**

Avant d'appliquer les tests paramétriques, il convient de vérifier que la distribution des variables suit une loi normale. Cette vérification a été faite à l'aide de graphiques, diagramme des fréquences cumulatives, QQ-Plot, du calcul des moments d'ordre 3 et 4 (coefficient de symétrie et coefficient d'aplatissement) et du test de Kolmogorov. Les vérifications sont faites variable par variable sur l'ensemble des données après les avoir centrées et réduites par rapport à la moyenne et à l'écart type des séries de données des différents groupes définis par le sexe et la semaine du prélèvement. Lorsque les données ne sont pas distribuées selon une loi normale, deux options sont possibles : i) utiliser des tests non-paramétriques ou des tests d'imputation, ii) transformer¹³ les données pour que ces dernières suivent une distribution normale.

- **Tests de différences entre les groupes ISO et OGM**

Ces tests permettent de comparer les moyennes des paramètres biologiques observées dans les groupes OGM et ISO.

Le test repose sur la définition de deux hypothèses (hypothèse nulle H_0 et hypothèse alternative H_1) définies paramètre par paramètre, par sexe et par dose du maïs ISO et OGM incorporée dans le régime des animaux :

- H_0 : absence d'effet du régime sur le paramètre biologique qui se traduit par l'égalité des moyennes de ce paramètre dans les deux groupes de rats OGM et ISO comparés,
- H_1 : présence d'effet du régime qui se traduit par une augmentation ou une diminution de la moyenne du paramètre biologique dans le groupe de rats OGM comparativement au groupe de rats ISO.

Des tests de différence paramétriques et non-paramétriques (tests de Wilcoxon par sexe, dose et période) ont été réalisés au risque de première espèce α égal à 10% (p -value = 0,1). Les tests paramétriques ont été effectués à partir des estimations des différences des moyennes et variances obtenues avec un modèle ANOVA (analyse de la variance) adapté aux données. Le modèle utilisé pour analyser les paramètres biologiques mesurés sur deux périodes intègre les effets suivants :

- Effet régime : OGM vs ISO
- Effet dose : 11% vs 33%
- Effet période : période 1 vs période 2

¹³ Asymétrie à droite : transformation de type log : $y = \log(x - \alpha)$

Asymétrie à gauche : transformation de type puissance : $y = \left(1 + \frac{x}{\alpha}\right)^\alpha$

α est le paramètre de transformation maximisant la valeur P du test de Kolmogorov.

- Effet interactions d'ordre 1 et d'ordre 2 entre les facteurs précédents
- Effet animal (aléatoire)

Le modèle utilisé pour analyser les variables mesurées sur une seule période (période 2) ne tient compte que des effets suivants :

- Effet régime : OGM vs ISO
- Effet dose : 11% vs 33%
- Effet interaction entre régime et dose

Les conséquences de la multiplicité des tests de différence sur l'erreur de 1^{ère} espèce ont été étudiées en utilisant la démarche de Storey et la procédure de Benjamini-Hochberg.

Pour tester l'effet du régime OGM sur la croissance pondérale des rats, un modèle non linéaire mixte a été ajusté aux données pondérales pour analyser les caractéristiques des courbes de croissance.

L'ensemble des tests a été effectué au risque de première espèce égal à 10% ($p < 0,1\%$).

• Calcul de la puissance des tests de différence

La puissance des tests est généralement déterminée lors de la définition du protocole expérimental de manière à fixer des tailles d'échantillon suffisantes. Ce calcul n'ayant pas été réalisé par le pétitionnaire, nous avons calculé la taille d'effet détectable pour une puissance de 80%. Cette taille d'effet détectable correspond à la différence minimale entre traitements OGM et ISO permettant de rejeter l'hypothèse nulle (absence d'effet) avec une probabilité de 0.8. Elle est calculée pour chaque paramètre en utilisant les écart-types des mesures obtenues lors de l'expérimentation pour 2 tailles d'échantillons : 10 animaux ou 20 animaux.

Les valeurs calculées de taille d'effet détectable peuvent être utilisées pour interpréter les tests de différence non significatifs qui peuvent correspondre à deux situations différentes : i) une réelle absence d'effet ou ii) un manque de puissance qui ne permet pas de révéler une différence réelle ; l'analyse de la puissance et le calcul de la taille d'effet détectable peuvent être utilisés pour conclure à l'une de ces deux situations.

Les tailles d'effet détectables sont comparées à des seuils de tolérance ayant une signification toxicologique. Considérons un paramètre pour lequel la différence OGM vs. ISO n'est pas significative. Si pour ce paramètre, la taille d'effet détectable est inférieure au seuil de tolérance, il est possible de conclure que la puissance est suffisante et ce paramètre peut être écarté de l'analyse toxicologique. Inversement, si la taille d'effet détectable est supérieure au seuil de tolérance, il n'est pas exclu que la différence non significative soit due à une puissance insuffisante. Les résultats pour ce paramètre devront donc être pris en compte par le toxicologue.

La procédure décrite ci-dessus nécessite la définition d'un seuil de tolérance pour chaque paramètre. Un des moyens consisterait à définir ces seuils sur la base de connaissances biologiques des effets toxicologiques observés. De tels éléments sont cependant disponibles pour un nombre limité de paramètres (e.g. ALAT). Deux approches ont été retenues ici pour déterminer les seuils de tolérance : la première basée sur les données observées dans les six groupes nourris avec les variétés commerciales et la deuxième sur les données historiques provenant de la même souche de rat et publiées dans la littérature (Petterino and Argentino-Storino, 2006; Wolford *et al.*, 1986).

Dans le premier cas, le seuil de tolérance correspond à un écart-type intra-groupe. Cet écart-type est estimé grâce à un modèle d'analyse de la variance mixte dans le cas où les paramètres sont mesurés en deux dates ou à un modèle d'ANOVA simple dans le cas d'une mesure à une seule date. Lorsqu'un modèle mixte est utilisé, l'écart-type retenu est égal à la racine carrée de la somme des variances des deux termes aléatoires du modèle. Dans le cas d'une ANOVA simple, l'écart-type est celui de la résiduelle. Etant donné que les analyses sont faites sur des données transformées, une procédure de simulation de Monte-Carlo est utilisée pour exprimer ces écart-types dans leurs unités initiales respectives (annexe 6).

Dans le deuxième cas, les seuils de tolérance basés sur les données historiques publiées ont été fixés aux écart-types de ces données.

Cette procédure a permis de fournir au toxicologue pour chaque paramètre les éléments suivants:

- 1- Taille d'effet détectable pour une puissance de 80% (notée ensuite *delta limite*)

- 2- L'écart-type provenant des données historiques d'une étude de la littérature
- 3- L'écart-type provenant des données observées dans les groupes nourris avec les variétés commerciales
- 4- Puissance avec laquelle il est possible de détecter des différences au moins égales aux écart-types pour 2 tailles d'échantillons : 10 animaux et 20 animaux par groupe.

Si le delta limite est supérieur à un écart type des données commerciales ou des données historiques, la puissance est alors jugée insuffisante. Dans le cas contraire, la puissance est jugée suffisante. Nous rappelons qu'un test non significatif avec une puissance suffisante a plus de chance de correspondre à une situation d'absence d'effet qu'un test non significatif avec une puissance jugée non suffisante.

- **Tests d'équivalence**

Un test d'équivalence a été réalisé pour chaque paramètre avec les hypothèses suivantes :

- $H_0 : |\mu_1 - \mu_2| \geq \Delta$
- $H_1 : |\mu_1 - \mu_2| < \Delta$

où μ_1 et μ_2 sont les moyennes des paramètres biologiques respectivement dans le groupe OGM et ISO.

Comme pour l'interprétation des tailles d'effet détectables, la réalisation d'un test d'équivalence nécessite un seuil de tolérance noté ici Δ . Dans ce cas d'étude nous avons choisi le même seuil que celui déterminé pour les calculs de la puissance des tests de différence.

Le test d'équivalence à 10% a été réalisé pour chaque paramètre en calculant un écart entre les moyennes comparées deux à deux en tenant compte des imprécisions liées au dispositif expérimental. Au lieu de calculer, la différence entre les moyennes observées, on calcule une étendue dite d'équivalence (équivalence range : eRg), qui correspond à la différence entre les deux extrêmes des intervalles de confiances des moyennes observées (Tryon and Lewis, 2008). Le calcul de l'intervalle de confiance est fait sur les données transformées tout en respectant le principe du test d'équivalence puis les données sont retransformées pour être exprimées dans les unités initiales. Si l'eRg est inférieur au seuil de tolérance (Δ), le test d'équivalence est significatif ; on rejette dans ce cas l'hypothèse que la différence entre les deux moyennes dépasse Δ .

3. Résultats

3.1 Données atypiques

L'analyse a permis de détecter 38 valeurs atypiques (figures 8 et 9) sur 21 281 valeurs observées. Elles concernent 30 animaux et 28 paramètres biologiques et se répartissent uniformément dans tous les groupes d'animaux, à l'exception du groupe 1 (MON810 11%) et 9 (variété commerciale) pour lesquels une plus grande prévalence est observée.

La majorité des données atypiques (81.6 %) sont isolées (un paramètre atypique par animal lors d'une seule période). Sept données atypiques correspondent à la femelle 14 du groupe 1 (OGM 11%) en période 2 (figure 9).

Les toxicologues ont jugés ces données atypiques non imputables au traitement. Ces données peuvent donc être éliminées.

3.2 Transformation des variables

Sur les 60 variables quantitatives, 56 présentent une distribution asymétrique nécessitant une transformation logarithmique pour 35 variables et une transformation puissance pour 21 variables. Les paramètres de transformation sont présentés en annexe (Annexe 11).

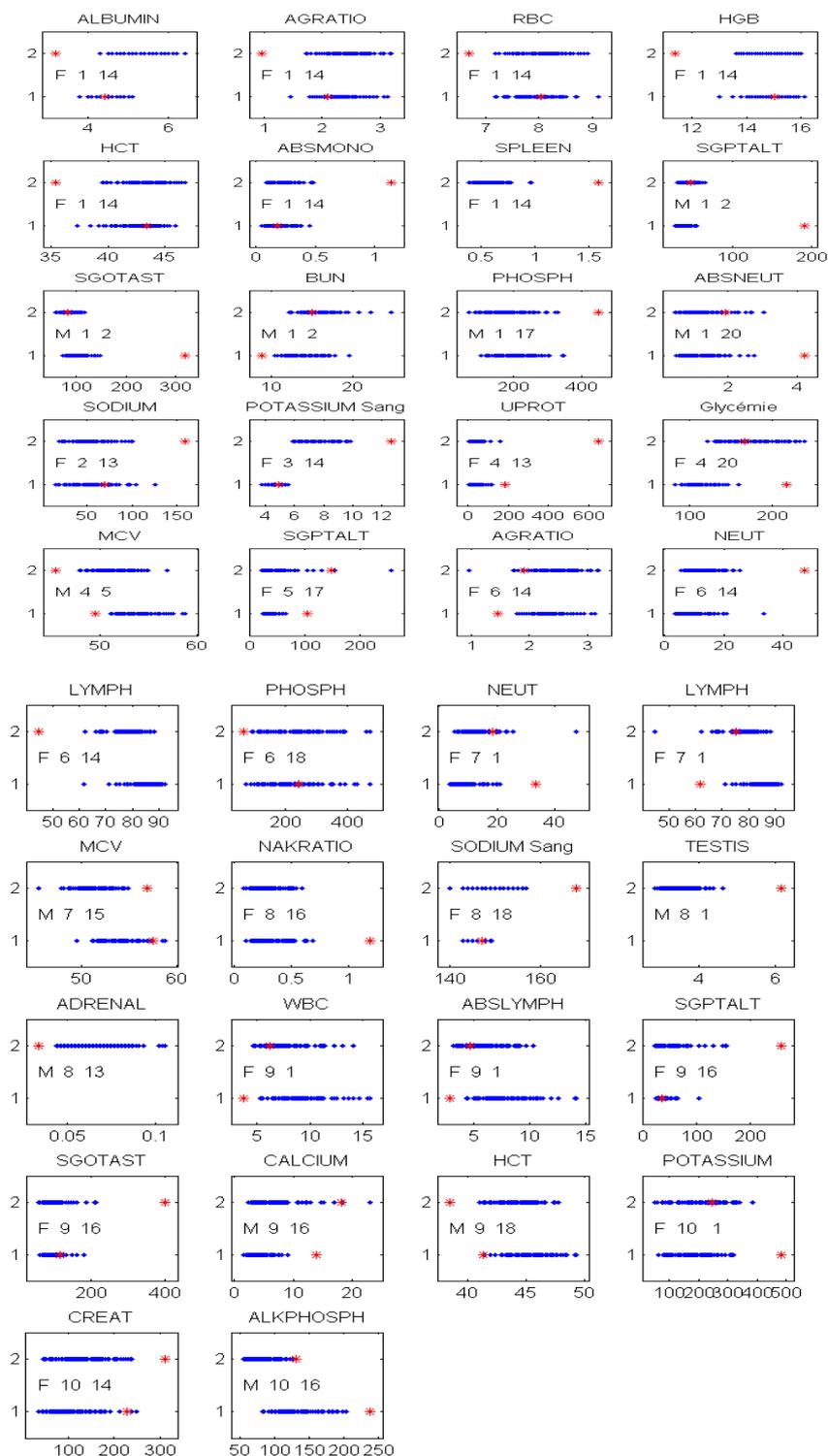


Figure 8: Représentation graphique des 38 données atypiques identifiées dans l'étude MON810.

Chaque graphe correspond aux résultats d'un paramètre pour un sexe. Chaque point des graphes correspond à un individu, les valeurs pour l'animal correspondant à une valeur atypique sont représentées en rouge pour les deux périodes P1 et P2. Le sexe, le groupe et le numéro du rat concerné sont indiqués au milieu du graphe.

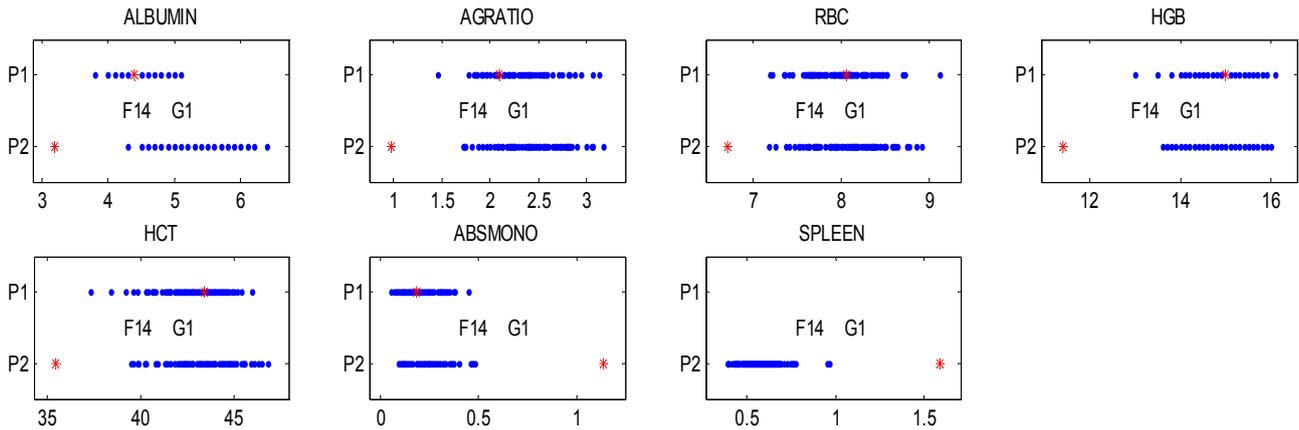


Figure 9: Valeurs atypiques de la femelle 14 du groupe OGM 11% pendant la période 2. Chaque graphe correspond à un paramètre biologique.

3.3 Tests des différences entre les groupes OGM et témoin

Analyses des 7 variables qualitatives

Sur les 56 tests non paramétriques conditionnellement au sexe, à la dose et à la période, deux sont significatifs au risque $\alpha=10\%$. Il s'agit du paramètre protéines dans l'urine (Var3) aux doses 11% et 33% pour la période 1.

Comparaisons des résultats des tests paramétriques et non paramétriques

Au total 432 tests paramétriques et non-paramétriques ont été réalisés pour comparer les groupes ISO et OGM. Les tests paramétriques mis en œuvre sans les données atypiques et les tests non paramétriques réalisés avec l'ensemble des données donnent respectivement 33 et 36 résultats significatifs sur 432 (annexe 7). Le nombre de discordances entre les deux types de test étant faible, le toxicologue n'analysera que les résultats des tests paramétriques.

Analyse des 60 variables quantitatives (données non pondérales)

Le tableau 2 présente les résultats des tests significatifs. Lorsqu'une variable est significative, les autres résultats des tests concernant cette même variable sont également présentés dans ce tableau. Au total, **33 tests** sur 432 sont significatifs au risque α de 10%. A titre indicatif, 22 tests le sont au risque α de 5%. Les différences significatives entre ISO et OGM sont observées soit pour la dose 11% soit pour la dose 33% et jamais pour les deux doses. Pour deux paramètres (AGRATIO et ALBUMIN), des différences significatives ont été obtenues pour les deux périodes considérées dans l'expérimentation chez les rats mâles à la dose 33%. Parmi les 33 tests significatifs ($p<0.10$), 21 (63%) le sont en deuxième période. Les différences significatives sont plus fréquentes en période 2 pour les femelles à la dose 11%, mais ce n'est pas le cas à la dose 33%. La répartition des tests significatifs entre sexes, périodes et doses est détaillée dans le tableau 1 ci-dessous.

Tableau 1: Répartition des p-values inférieures à 0.1 dans les différents groupes de l'expérience.

	Dose 11%		Dose 33%	
	mâles	femelles	mâles	femelles
Période 1	1	0	5	6
Période 2	3	12	4	2

Multiplicité des tests et conséquence sur les risques liés aux tests

Pour les 432 tests paramétriques, si l'on fixe à 10% la proportion de faux positifs (hypothèses nulles rejetées à tort), aucune hypothèse nulle ne peut être rejetée. Les valeurs de q (Storey) sont toutes supérieures à 10% quel que soit le degré de signification. L'espérance mathématique du nombre de faux positifs est égale au nombre de tests jugés significatifs (33). La démarche de Storey fournit le même résultat que la procédure de Benjamini-Hochberg.

A noter que sur les 488 tests non paramétriques réalisés (paramètres qualitatifs et quantitatifs), les résultats sont très similaires.

Tableau 2: Tests paramétriques de comparaison de moyennes statistiquement significatifs

Paramètre biologique	Unité	sexe	P	Dose 11%						Dose 33%						
				N	Moyenne		Différence	Différence/ISO	P	N	Moyenne		Différence	Différence/ISO	P	
					OGM	ISO					OGM	ISO				
Var09	SODIUM_Sang	mmol/L	F	2	20	147,12	146,38	0,74	0,005	0,46	20	145,51	147,55	-2,04	-0,014	0,04
Var09	SODIUM_Sang	mmol/L	M	2	20	149,01	148,38	0,63	0,004	0,55	20	144,76	148,66	-3,9	-0,026	0,01
Var10	AGRATIO		M	1	20	1,75	1,78	-0,03	-0,016	0,7	20	1,77	1,98	-0,21	-0,105	0,03
Var10	AGRATIO		M	2	20	1,68	1,66	0,02	0,013	0,87	20	1,66	1,85	-0,19	-0,104	0,03
Var11	ALBUMIN	g/dL	M	1	20	4,01	4	0,01	0,003	1	20	3,97	4,23	-0,26	-0,061	0,01
Var11	ALBUMIN	g/dL	M	2	20	4,31	4,29	0,02	0,004	0,91	20	4,15	4,45	-0,3	-0,067	0
Var13	BUN	mg/dL	F	1	20	15,57	14,78	0,79	0,054	0,54	20	17,26	15,03	2,22	0,148	0,04
Var14	CALCIUM_Sang	mg/dL	F	1	20	10,07	10,02	0,05	0,005	0,78	20	10,08	10,4	-0,32	-0,031	0,07
Var15	CHLORIDE_Sang	mmol/L	F	2	20	107,3	107,22	0,08	0,001	0,95	20	105,64	107,21	-1,57	-0,015	0,07
Var15	CHLORIDE_Sang	mmol/L	M	2	20	105,61	104,97	0,63	0,006	0,44	20	103,37	105,92	-2,55	-0,024	0
Var19	GLOBULINE	g/dL	F	2	20	2,35	2,11	0,24	0,115	0,05	20	2,18	2,2	-0,01	-0,006	0,85
Var23	TOTBILI	mg/dL	M	1	20	0,12	0,1	0,02	0,2	0,08	20	0,12	0,11	0,01	0,091	0,36
Var27	APTSEC	sec	F	2	20	14,94	13,55	1,4	0,103	0,04	20	14,08	13,69	0,39	0,028	0,53
Var30	PLT	.10 ³ /μL	F	2	20	1061,51	935,31	126,2	0,135	0,02	20	908,31	976,08	-67,78	-0,069	0,21
Var32	ABSLUC	.10 ³ /μL	F	2	20	0,21	0,13	0,08	0,635	0,02	20	0,14	0,17	-0,04	-0,209	0,24
Var33	ABSLYMPH	.10 ³ /μL	F	1	20	7,8	7,35	0,46	0,062	0,62	20	7,19	8,78	-1,59	-0,181	0,06
Var34	ABSMONO	.10 ³ /μL	F	1	20	0,21	0,16	0,05	0,299	0,11	20	0,17	0,23	-0,06	-0,247	0,09
Var35	ABSNEUT	.10 ³ /μL	F	2	20	1,29	0,94	0,34	0,364	0,08	20	0,82	0,96	-0,13	-0,138	0,48

Suite tableau 2

Paramètre biologique	Unité	sexe	P	Dose 11%						Dose 33%						
				N	Moyenne		Différence	Différence/ISO	P	N	Moyenne		Différence	Différence/ISO	P	
					OGM	ISO					OGM	ISO				
Var38	HCT	%	M	1	19	45,05	44,44	0,62	0,014	0,37	19	44,83	46,28	-1,46	-0,032	0,05
Var39	HGB	g/dL	M	1	19	15,48	15,15	0,32	0,021	0,25	19	15,55	16,04	-0,49	-0,031	0,05
Var40	LUC	%	F	2	20	2,32	1,73	0,59	0,341	0,03	20	1,84	2,19	-0,35	-0,16	0,17
Var43	MCHC	g/dL	F	2	20	33,71	34,5	-0,79	-0,023	0	20	34,21	34,39	-0,18	-0,005	0,44
Var43	MCHC	g/dL	M	2	20	34,12	33,48	0,64	0,019	0,02	20	33,86	33,79	0,07	0,002	0,73
Var45	RBC	.106/ μ L	M	1	19	8,4	8,17	0,24	0,029	0,24	19	8,25	8,66	-0,41	-0,047	0,05
Var48	WBC	.106/ μ L	F	1	20	9,13	8,69	0,44	0,051	0,59	20	8,1	10,02	-1,92	-0,191	0,05
Var49	CALCIUM	mg/dL	M	2	19	8,86	6,11	2,75	0,45	0,03	20	7,17	6,29	0,88	0,14	0,44
Var54	CREATCLEARANCE		F	1	18	0,51	0,61	-0,1	-0,157	0,25	19	0,44	0,57	-0,13	-0,234	0,08
Var57	UPROT	mg/dL	M	2	19	248,7	153,44	95,26	0,621	0,04	20	142,2	174,3	-32,1	-0,184	0,46
Var59	ADRENAL	g	F	2	40	0,08	0,07	0,01	0,114	0,02	40	0,07	0,07	0	-0,02	0,7
Var61	HEART	g	F	2	40	1,15	1,08	0,07	0,065	0,1	40	1,16	1,11	0,05	0,046	0,29
Var62	KIDNEY	g	F	2	40	2,37	2,24	0,13	0,057	0,05	40	2,24	2,23	0,01	0,004	0,93
Var63	LIVER	g	F	2	40	8,92	8,3	0,62	0,075	0,06	40	8,66	8,38	0,28	0,033	0,38
Var65	SPLEEN	g	F	2	39	0,6	0,54	0,06	0,108	0,06	40	0,59	0,54	0,05	0,093	0,1

Les différences significatives ($p < 0.1$) sont en rouge. P= période 1 : 5ème semaine et 2 : 14ème semaine. Différence : différence entre OGM et ISO, Différence/ISO : différence relative à la moyenne du groupe ISO. Les résultats des comparaisons entre groupe OGM 11% versus ISO 11% et groupe OGM 33% versus ISO 33% sont accolés. Ainsi, pour un même paramètre, période et sexe dont la valeur p est inférieure à 0.1 pour une des 2 comparaisons, les résultats pour la comparaison à l'autre dose sont visibles. Le signe de la différence indique s'il s'agit d'une baisse ou d'une hausse de la valeur observée dans le groupe OGM par rapport au groupe ISO. Les paramètres sont triés par type d'analyse: chimie du sang, hématologique, urinaire, poids des organes.

Représentations graphiques

Les valeurs de p-value (degré de signification) obtenues pour les différents régimes, sexes et périodes sont très nombreuses. La figure 10 présente un exemple de représentation graphique des résultats des tests de différence qui peut faciliter l'interprétation des résultats. Des diagrammes circulaires similaires ont été réalisés pour l'ensemble des paramètres (Annexe 8).

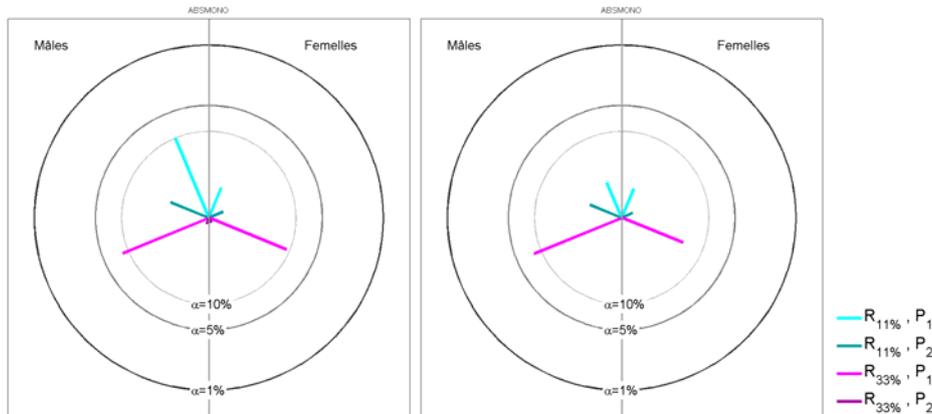


Figure 10: p-value des tests de comparaison du nombre de monocytes absolus en $10^3/\mu\text{l}$.

Comparaison des moyennes : diagramme de gauche

Comparaison des médianes (test de Wilcoxon) : diagramme de droite.

La longueur d'un segment est égale à $-\log_{10}(p\text{-value})$. Un long segment indique une petite p-value alors qu'une p-value proche de 1 sera représentée par un segment court. Les valeurs repères sont représentées (cercle à 10%, 5% et 1% pour le seuil de significativité). Ainsi, pour les 2 tests (non paramétriques et paramétriques), les p-values pour le régime 33% chez les mâles et chez les femelles pendant la période 1 sont proches de 0.10 alors qu'elles sont proches de 1 pour la période 2 et le même régime.

La p-value n'est pas le seul résultat à prendre en compte dans l'analyse des résultats. Il est aussi utile de représenter les tailles d'effet (différence des moyennes) dans l'unité initiale du paramètre, comme dans l'exemple ci dessous, correspondant au nombre absolu de monocytes en $10^3/\mu\text{l}$.

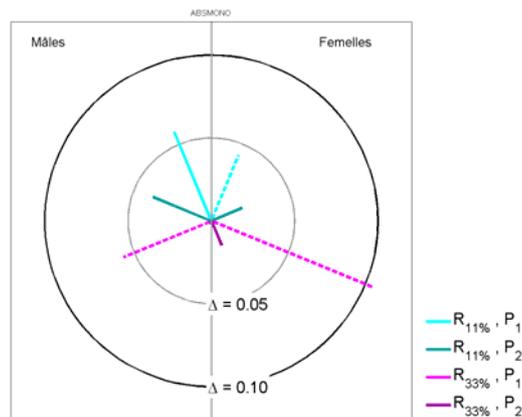


Figure 11: Taille de l'effet observé pour le paramètre nombre de monocytes absolus en $10^3/\mu\text{l}$.

Les différences entre les moyennes, exprimées dans l'unité initiale du paramètre ou exprimées en % relativement à la moyenne du groupe témoin sont également présentées sous forme de graphique.

Pour en faciliter l'interprétation, il est aussi intéressant de pouvoir comparer les différences observées sur plusieurs paramètres biologiques. Les différences peuvent être normalisées en les divisant par l'écart-type de la moyenne du groupe témoin (ISO) (exprimées en %) et représentées sous cette forme (figure 12).

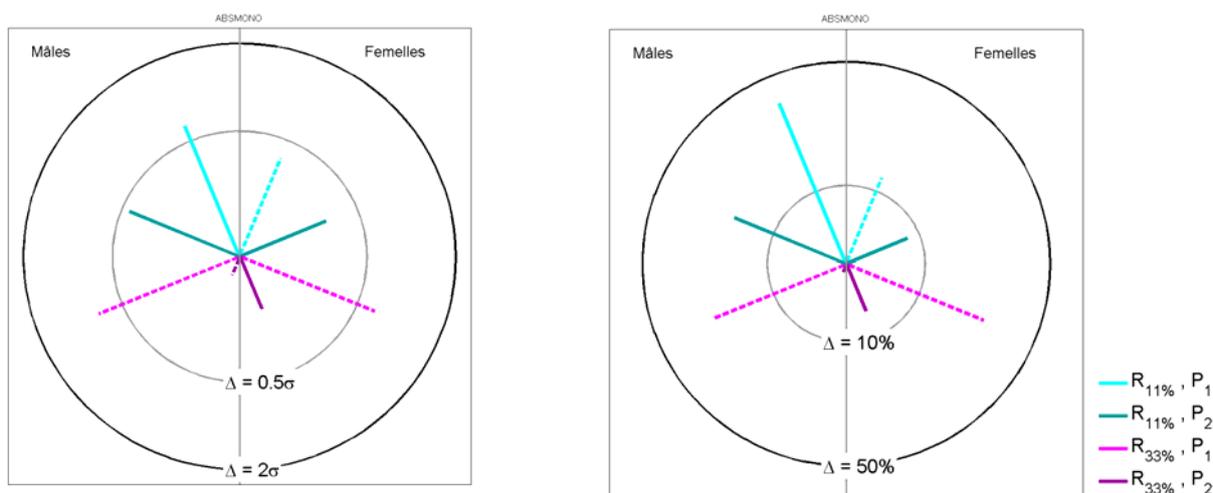


Figure 12: Taille de l'effet observé pour le paramètre nombre de monocytes absolus. Exprimés en écart-types (graphe de gauche) et en % relatif par rapport à la moyenne des groupes témoins (graphe de droite).

Les résultats pour les autres paramètres sont présentés en annexe (Annexe 9).

3.4 Analyse de l'évolution pondérale.

Ces données longitudinales (répétition des mesures au cours du temps) sont traitées telles que recommandées dans la première partie (cf 2.3 chapitre I).

La figure 13 présente le tracé des différentes courbes de croissance individuelles classées par sexe, dose et régime maïs ISO ou OGM.

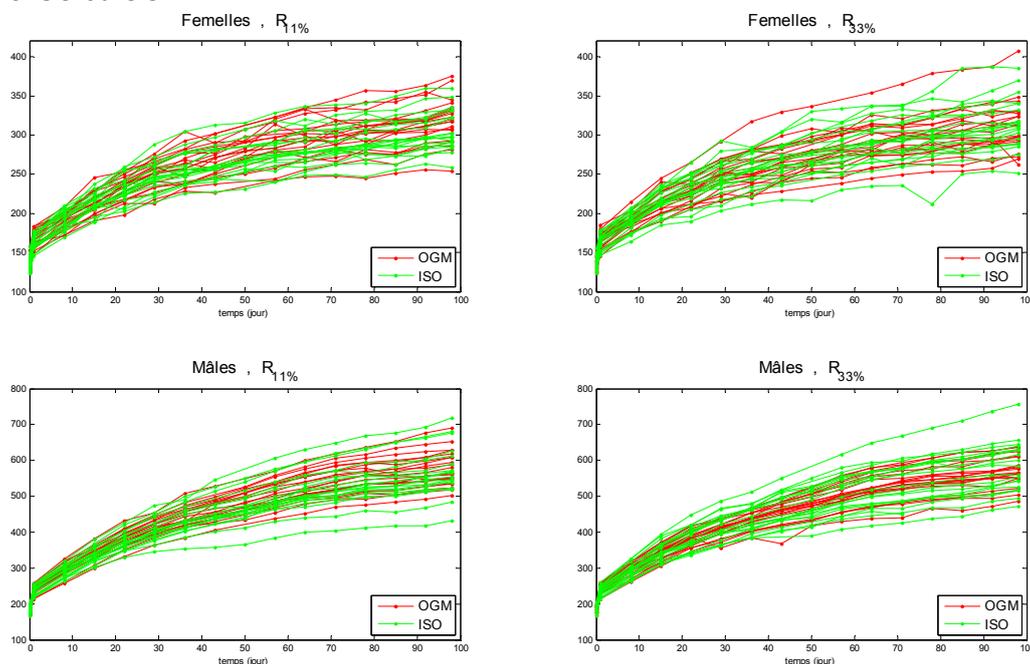


Figure 13: Courbes de croissance pondérale individuelle R11% dose maïs 11%, R33% dose maïs 33%

La figure 14 présente les courbes moyennes de croissance par groupe estimées à partir de modèle non linéaire (cf équation chapitre I § 2.3).

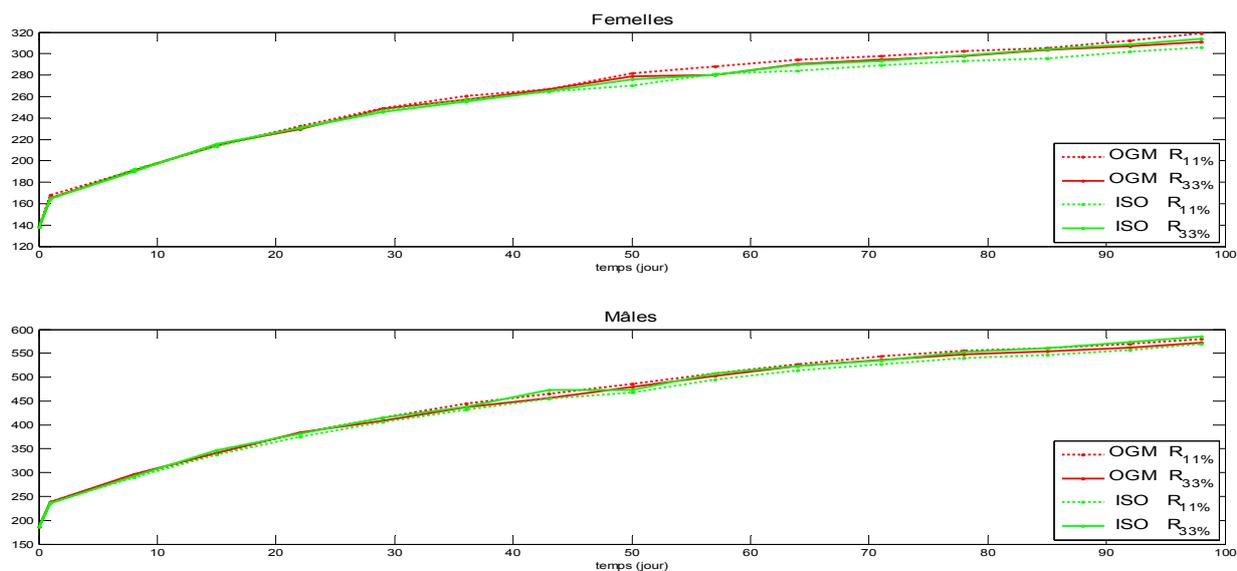


Figure 14: Courbes de croissance moyenne par sexe pour chaque groupe.
OGM 11%, OGM 33%, ISO 11% et ISO 33%

Les estimations des paramètres du modèle non linéaire mixte sont présentés dans le tableau 3.

Tableau 3: Paramètres du modèle non linéaire mixte de la croissance pondérale.

Paramètre modèle Mâles	Valeur paramètre	Ecart type	p-value
A (poids en g de l'animal à J0)	232	1.5	
B (taux de croissance de l'animal)	0.0207	0.00054	
Effet du régime OGM sur B	0.0114	0.03	0.7
Effet de la dose sur B	0.021	0.03	0.48
C (poids limite de l'animal)	618	14	
Effet du régime OGM sur C	0.00962	0.026	0.71
Effet de la dose sur B	-0.00307	0.026	0.91
Variance de A	0.0517	0.0048	
Variance de B	0.115	0.012	
Variance C	0.116	0.0094	
Variance résiduelle	6.49	0.15	
Paramètre Modèle femelles	Valeur paramètre	Ecart type	p-value
A (poids en g de l'animal à J0)	163	1	
B (taux de croissance de l'animal)	0.0255	0.001	
Effet du régime OGM sur B	-0.0386	0.047	0.41
Effet de la dose sur B	-0.0569	0.047	0.22
C (poids limite de l'animal)	318	6.1	
Effet du régime OGM sur C	0.0251	0.022	0.26
Effet de la dose sur B	0.0113	0.022	0.61
Variance de A	0.0495	0.005	
Variance de B	0.177	0.019	
Variance C	0.0967	0.0079	
Variance résiduelle	5.32	0.12	

Les figures 15 et 16, montrent la qualité de l'ajustement du modèle et l'intérêt du modèle mixte par rapport à un modèle non linéaire ignorant la variabilité inter-animaux.

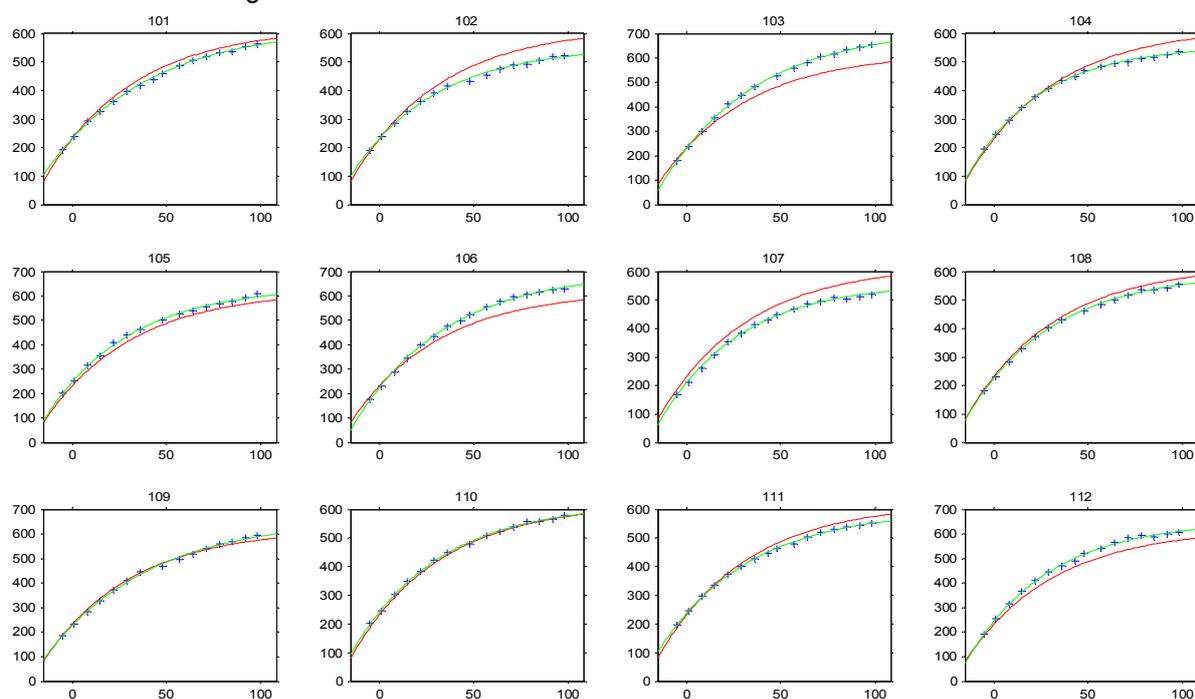


Figure 15: Graphes explorant la qualité de l'ajustement.

Les points expérimentaux (en bleus), la courbe utilisant le modèle mixte (verte), la courbe utilisant un modèle sans prise en compte des effets aléatoires (rouge).

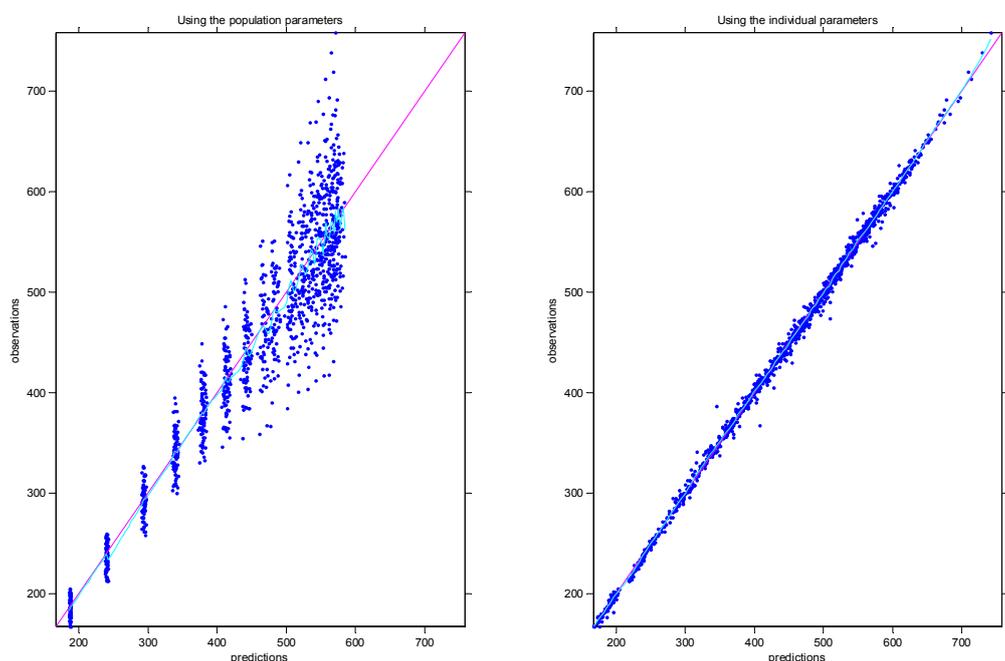


Figure 16: Valeurs observées en fonction des valeurs prédites avec un modèle.

Graphes de gauche : le modèle ne prend pas en compte la variabilité inter-animal

Graphes de droite : le modèle est un modèle mixte.

Les tests des effets du régime OGM par rapport au régime ISO sont tous non significatifs au risque 10% (tableau 3).

3.5 Analyse de la puissance

Le calcul des seuils de tolérance montre que les données historiques de la bibliographie¹⁴ ont dans la majorité des cas un écart-type plus grand que celui des données issues des groupes nourris avec les variétés commerciales. L'annexe 10 présente les écart-types historiques et les écart-types issus des données concernant les variétés commerciales et la taille d'effet détectable avec une puissance de 80%.

Sur les 216 tailles d'effet détectables¹⁵ calculées, seulement 49 (22%) sont inférieures à la valeur de 1 écart-type des données issues des variétés commerciales. La puissance des tests prenant comme taille d'effet l'écart-type des données issues des variétés commerciales sont supérieures à 0.8 dans 20% des 216 effets étudiés, entre 0.7 et 0.8 dans 29% des effets étudiés et inférieure à 0.7 dans les 51% restant des effets étudiés.

Comme les données historiques ne sont disponibles que pour certains paramètres, seules 132 comparaisons entre tailles d'effet détectables et écart-types bibliographiques ont pu être réalisées. Les résultats montrent que 68 tailles d'effet détectables sont inférieures aux écart-types historiques (51%).

Ainsi, pour une analyse paramètre par paramètre, par sexe et par période, en prenant une taille d'effet de un écart type et une p-value de 0.1, la puissance du test est insuffisante (inférieure à 80%) dans 50% des tests de différence entre groupe OGM et groupe ISO pour les références calculées par rapport aux données historiques. Elle est insuffisante dans 80% des tests avec les seuils de tolérance calculées sur les données issues des rats nourris avec les variétés commerciales.

Cent seize tests concernent les variables ne présentant aucune différence significative. Parmi ces 116 tests, 110 présentent un manque de puissance comparativement à l'une des tailles d'effet de référence (tableau 5).

L'analyse de la puissance, avec l'hypothèse de doublement des effectifs par groupe (20 animaux au lieu de 10) montrent une puissance supérieure à 0.80 pour 94% des tests et à 0.70 pour 99.1% des tests (214/216). Il est important de noter que ces calculs sont valables pour un seuil de tolérance fixé à un écart type des données issues des rats nourris avec les maïs commerciaux (Annexe 10) et que d'autres résultats pourraient être obtenus en utilisant d'autres seuils de tolérance. En effet, nous avons considéré ici un seuil égal à un écart-type, des seuils supérieurs à un écart-type permettraient aussi d'obtenir des puissances plus élevées.

Le tableau 4 présente les puissances et tailles d'effet détectables des paramètres pour lesquels un test de différence significatif ($p < 10\%$) a été obtenu pour au moins une combinaison sexe, période, dose.

3.6 Tests d'équivalence

En prenant comme seuil de tolérance un écart-type des données issues des variétés commerciales, l'hypothèse nulle de non équivalence n'est jamais rejetée (0 test significatif sur 432). Avec un seuil de tolérance fixé à un écart-type des données bibliographiques, l'hypothèse nulle de non équivalence est rejetée pour 2 tests sur 264. Ces résultats montrent que, dans la majorité des cas, les données expérimentales ne permettent pas de rejeter l'hypothèse « la différence entre OGM et ISO est supérieure à un seuil de tolérance ».

¹⁴ Petterino and Argentino-Storino, 2006, Exp. and Tox. Pathol. 57, 213-219.
Wolford *et al.*, 1986, J. Tox and Environ. Health, 18, 161-188.

¹⁵ Le nombre d'effet détectable calculé est deux fois moins important que le nombre de test car pour les groupes d'animaux nourris avec les variétés commerciales, une seule dose 33% est mise en œuvre donc les données ne sont disponibles qu'à la dose de (33%).

Tableau 4: Puissance des tests pour les variables ayant au moins un test significatif (p-value < 10%).

Paramètre	sexe	periode	pValue 1-3	pValue 2-4	Ecart-type des données historiques	Ecart-type des données commerciales		d80% : Delta détectable pour une puissance de 80%		Données commerciales		Données historiques	
						Puissance	Delta D2	Régime 11%	Régime 33%	Puissance 33%	Puissance 33%		
Var09	SODIUM_Sang	F	1	0,17	0,65	8,60	0,631	1,882	2,323	2,313	0	1	
Var09	SODIUM_Sang	F	2	0,46	0,04	7,00	0,631	1,942	2,439	2,473	0	1	
Var09	SODIUM_Sang	M	1	0,68	0,88	9,00	0,530	2,199	2,973	3,132	0	1	
Var09	SODIUM_Sang	M	2	0,55	0,01	9,00	0,530	2,486	3,609	3,575	0	1	
Var10	AGRATIO	F	1	0,73	0,64	0,10	0,696	0,311	0,357	0,341	0	0	
Var10	AGRATIO	F	2	0,35	0,89	0,10	0,696	0,32	0,394	0,368	0	0	
Var10	AGRATIO	M	1	0,70	0,03	0,10	0,608	0,175	0,228	0,206	0	0	
Var10	AGRATIO	M	2	0,87	0,03	0,10	0,608	0,178	0,222	0,230	0	0	
Var11	ALBUMIN	F	1	0,84	0,55	0,40	0,737	0,37	0,377	0,400	0	1	
Var11	ALBUMIN	F	2	0,60	0,51	0,40	0,737	0,326	0,403	0,396	0	1	
Var11	ALBUMIN	M	1	1,00	0,01	0,40	0,753	0,246	0,231	0,235	1	1	
Var11	ALBUMIN	M	2	0,91	0,00	0,30	0,753	0,225	0,250	0,244	0	1	
Var13	BUN	F	1	0,54	0,04	21,60	0,761	2,302	2,615	2,303	0	1	
Var13	BUN	F	2	0,75	0,22	22,20	0,761	2,6	2,787	2,702	0	1	
Var13	BUN	M	1	0,90	0,61	22,20	0,717	1,723	1,976	2,082	0	1	
Var13	BUN	M	2	0,78	1,00	26,40	0,717	2,179	2,501	2,479	0	1	
Var14	CALCIUM_Sang	F	1	0,78	0,07	0,40	0,686	0,386	0,433	0,435	0	0	
Var14	CALCIUM_Sang	F	2	0,62	0,17	0,40	0,686	0,703	0,815	0,825	0	0	
Var14	CALCIUM_Sang	M	1	0,43	0,47	0,40	0,608	0,349	0,415	0,500	0	0	
Var14	CALCIUM_Sang	M	2	0,45	0,35	0,40	0,608	0,689	0,834	0,882	0	0	
Var15	CHLORIDE_Sang	F	1	0,40	0,84	4,50	0,765	2,145	2,202	2,262	0	1	
Var15	CHLORIDE_Sang	F	2	0,95	0,07	4,90	0,765	1,685	2,049	2,002	0	1	
Var15	CHLORIDE_Sang	M	1	0,99	0,39	3,90	0,862	2,428	2,477	2,380	1	1	
Var15	CHLORIDE_Sang	M	2	0,44	0,00	4,90	0,862	2,387	2,244	2,069	1	1	
Var19	GLOBULINE	F	1	0,59	0,87	0,20	0,634	0,235	0,290	0,292	0	0	
Var19	GLOBULINE	F	2	0,05	0,85	0,10	0,634	0,276	0,306	0,309	0	0	
Var19	GLOBULINE	M	1	0,69	0,36	0,30	0,627	0,207	0,258	0,255	0	1	
Var19	GLOBULINE	M	2	0,92	0,35	0,30	0,627	0,212	0,275	0,268	0	1	
Var23	TOTBILI	F	1	0,22	0,52	0,00	0,689	0,034	0,037	0,039	0	0	
Var23	TOTBILI	F	2	0,44	1,00	0,00	0,689	0,031	0,038	0,039	0	0	
Var23	TOTBILI	M	1	0,08	0,36	0,00	0,907	0,033	0,027	0,028	1	1	
Var23	TOTBILI	M	2	0,78	0,85	0,00	0,907	0,032	0,028	0,027	1	1	
Var27	APTSEC	F	2	0,04	0,53	0,00	0,782	1,672	1,852	1,681	0	0	
Var27	APTSEC	M	2	0,57	0,79	0,00	0,911	1,722	1,489	1,566	1	1	
Var30	PLT	F	1	0,43	0,88	141,50	0,617	123,272	156,629	165,568	0	0	
Var30	PLT	F	2	0,02	0,21	140,90	0,617	108,401	128,135	140,152	0	1	
Var30	PLT	M	1	0,71	0,19	153,00	0,527	142,407	188,161	195,860	0	0	
Var30	PLT	M	2	0,67	0,66	136,20	0,548	128,378	187,135	201,063	0	0	
Var32	ABSLUC	F	1	0,18	0,38	0,00	0,558	0,053	0,074	0,078	0	0	
Var32	ABSLUC	F	2	0,02	0,24	0,00	0,558	0,059	0,081	0,081	0	0	
Var32	ABSLUC	M	1	0,60	0,70	0,00	0,523	0,065	0,095	0,099	0	0	
Var32	ABSLUC	M	2	0,85	0,92	0,00	0,543	0,062	0,101	0,093	0	0	
Var33	ABSLYMPH	F	1	0,62	0,06	0,00	0,802	2,776	2,404	2,632	1	1	
Var33	ABSLYMPH	F	2	0,26	0,99	0,00	0,802	1,79	1,837	1,723	1	1	
Var33	ABSLYMPH	M	1	0,85	0,47	0,00	0,645	3,775	4,432	4,877	0	0	
Var33	ABSLYMPH	M	2	0,45	0,56	0,00	0,667	2,475	3,086	3,010	0	0	
Var34	ABSMONO	F	1	0,11	0,09	0,00	0,703	0,087	0,090	0,097	0	0	
Var34	ABSMONO	F	2	0,32	0,84	0,00	0,703	0,084	0,098	0,102	0	0	
Var34	ABSMONO	M	1	0,44	0,12	0,00	0,592	0,131	0,180	0,197	0	0	
Var34	ABSMONO	M	2	0,67	0,87	0,00	0,614	0,11	0,144	0,148	0	0	

P-value 1-3 : valeur de p du test comparant OGM à ISO à la dose 11%, P-value 2-4 : valeur de p du test comparant OGM à ISO à la dose 33%, Ecart-type des données bibliographiques (Petterino et Argentino-Storino, 2006). Puissance : calculé pour une taille d'effet égale à un écart-type à l'échelle des données transformées. Delta2 : correspond à un écart-type sur les données des variétés commerciales en unité initiale. Les cases colorées en vert indiquent que l'écart-type des données commerciales ou bibliographiques selon le cas est plus élevé que le plus petit écart détectable avec une puissance de 80% (puissance suffisante). Dans la colonne puissance : les ronds verts, oranges et rouges indiquent respectivement, une puissance supérieure ou égale à 80%, entre 70 et 80% et strictement inférieure à 70%.

Suite Tableau 4

Paramètre		sexe	periode	pValue 1-3	pValue 2-4	Ecart-type des données historiques	Ecart-type des données commerciales		d80% : Delta détectable pour une puissance de 80%		Données commerciales	Données historiques
							Puissance	Delta D2	Régime 11%	Régime 33%	Puissance 33%	Puissance 33%
Var35	ABSNEUT	F	1	0,73	0,61		0,991	0,774	0,464	0,454	1	
Var35	ABSNEUT	F	2	0,08	0,48		0,991	0,805	0,453	0,485	1	
Var35	ABSNEUT	M	1	0,93	0,29		0,650	0,457	0,591	0,560	0	
Var35	ABSNEUT	M	2	0,65	0,91		0,672	0,467	0,593	0,538	0	
Var38	HCT	F	1	0,55	0,83	2,40	0,734	1,639	1,703	1,855	0	1
Var38	HCT	F	2	0,12	0,74	2,10	0,734	1,771	1,724	1,906	0	1
Var38	HCT	M	1	0,37	0,05	2,40	0,711	1,535	1,743	1,820	0	1
Var38	HCT	M	2	0,61	0,73	2,50	0,733	1,692	1,752	1,634	1	1
Var39	HGB	F	1	0,88	0,83	0,70	0,809	0,575	0,599	0,582	0	1
Var39	HGB	F	2	0,79	0,95	0,70	0,809	0,612	0,590	0,617	0	1
Var39	HGB	M	1	0,25	0,05	0,70	0,744	0,495	0,640	0,532	0	1
Var39	HGB	M	2	0,13	0,95	0,70	0,765	0,623	0,621	0,633	0	1
Var40	LUC	F	1	0,15	0,58	0,40	0,594	4,906	4,879	5,563	0	0
Var40	LUC	F	2	0,03	0,17	0,40	0,594	15,337	11,081	18,442	0	0
Var40	LUC	M	1	0,89	0,36	0,40	0,641	5,81	8,627	6,759	0	0
Var40	LUC	M	2	0,86	0,43	0,40	0,663	12,738	18,037	15,087	0	0
Var43	MCHC	F	1	0,19	0,98	1,40	0,809	0,515	0,528	0,516	0	1
Var43	MCHC	F	2	0,00	0,44	1,00	0,809	0,581	0,529	0,572	1	1
Var43	MCHC	M	1	0,62	0,81	1,10	0,630	0,442	0,560	0,557	0	1
Var43	MCHC	M	2	0,02	0,73	1,10	0,652	0,479	0,653	0,588	0	1
Var45	RBC	F	1	0,97	0,98	0,40	0,783	0,339	0,360	0,371	0	1
Var45	RBC	F	2	0,84	1,00	0,40	0,783	0,331	0,348	0,367	0	1
Var45	RBC	M	1	0,24	0,05	0,40	0,650	0,345	0,436	0,447	0	0
Var45	RBC	M	2	0,47	0,18	0,50	0,672	0,369	0,441	0,427	0	1
Var48	WBC	F	1	0,59	0,05	2,20	0,815	3,403	2,509	3,364	1	0
Var48	WBC	F	2	0,16	0,74	2,40	0,815	2,39	2,194	2,287	1	1
Var48	WBC	M	1	0,95	0,41	2,40	0,622	4,467	5,440	6,680	0	0
Var48	WBC	M	2	0,69	0,48	2,40	0,644	3,109	3,695	3,597	0	0
Var49	CALCIUM	F	1	0,57	0,53		0,514	7,749	12,068	10,286	0	
Var49	CALCIUM	F	2	0,13	0,83		0,514	7,188	11,523	11,610	0	
Var49	CALCIUM	M	1	0,36	0,42		0,831	3,131	3,030	2,862	1	
Var49	CALCIUM	M	2	0,03	0,44		0,831	2,973	3,200	2,917	1	
Var54	CREATCLEARANCE	F	1	0,25	0,08		0,702	0,208	0,280	0,246	0	
Var54	CREATCLEARANCE	F	2	0,56	0,79		0,702	0,129	0,147	0,143	0	
Var54	CREATCLEARANCE	M	1	0,26	0,82		0,477	0,153	0,293	0,281	0	
Var54	CREATCLEARANCE	M	2	0,85	0,91		0,477	0,089	0,154	0,152	0	
Var57	UPROT	F	1	0,97	0,68		0,845	24,657	24,743	23,072	1	
Var57	UPROT	F	2	0,85	0,86		0,845	24,993	24,017	22,691	1	
Var57	UPROT	M	1	0,45	0,59		0,474	69,309	109,187	103,927	0	
Var57	UPROT	M	2	0,04	0,46		0,474	59,773	110,650	109,010	0	
Var59	ADRENAL	F	2	0,02	0,70		0,956	0,013	0,007	0,009	1	
Var59	ADRENAL	M	2	0,15	0,33		0,983	0,012	0,008	0,007	1	
Var61	HEART	F	2	0,10	0,29		0,799	0,119	0,114	0,119	1	
Var61	HEART	M	2	0,46	0,57		0,972	0,222	0,119	0,151	1	
Var62	KIDNEY	F	2	0,05	0,93		0,872	0,201	0,188	0,172	1	
Var62	KIDNEY	M	2	0,72	0,74		0,892	0,438	0,383	0,362	1	
Var63	LIVER	F	2	0,06	0,38		0,968	1,367	0,818	1,033	1	
Var63	LIVER	M	2	0,63	0,96		0,807	2,185	2,309	2,265	0	
Var65	SPLEEN	F	2	0,06	0,10		0,784	0,073	0,089	0,081	0	
Var65	SPLEEN	M	2	0,26	0,12		0,925	0,136	0,113	0,123	1	

Tableau 5: Liste des tests concernant les variables sans différence significative et avec une puissance jugée insuffisante comparativement aux tailles d'effet de référence.

Paramètre		sexe	periode	pValue 1-3	pValue 2-4	Ecart-type des données historiques	Ecart-type des données commerciales		d80% : Delta détectable pour		Comparaison 33%	
							Puissance	Delta D2	Régime 11%	Régime 33%	Commerciale	Historique
Var02	GRAVITY	F	1	0,69	0,81	0,01	0,626	0,014	0,020	0,019	0	0
Var02	GRAVITY	F	2	0,61	0,73	0,01	0,626	0,015	0,027	0,020	0	0
Var02	GRAVITY	M	1	0,19	0,71	0,01	0,562	0,008	0,011	0,012	0	0
Var02	GRAVITY	M	2	0,24	0,89	0,01	0,562	0,01	0,015	0,015	0	0
Var06	pH	F	1	0,76	0,39	0,01	0,785	0,428	0,482	0,449	0	0
Var06	pH	F	2	0,41	0,57	0,01	0,785	0,431	0,471	0,426	1	1
Var06	pH	M	1	0,76	0,76	0,01	0,719	0,387	0,400	0,375	1	1
Var06	pH	M	2	0,45	0,76	0,01	0,719	0,381	0,429	0,410	0	0
Var12	ALPHOSPH	F	1	0,59	0,46	99,50	0,602	25,136	34,734	33,393	0	1
Var12	ALPHOSPH	F	2	0,30	0,31	53,00	0,602	15,735	21,068	18,174	0	1
Var12	ALPHOSPH	M	1	0,94	0,22	113,50	0,740	31,976	34,399	34,917	0	1
Var12	ALPHOSPH	M	2	0,72	0,11	73,60	0,740	18,691	18,986	17,579	1	1
Var16	Creat sang	F	1	0,53	0,22	0,10	0,677	0,075	0,086	0,084	0	1
Var16	Creat sang	F	2	0,76	0,53	0,10	0,677	0,071	0,088	0,074	0	1
Var16	Creat sang	M	1	0,74	1,00	0,10	0,586	0,058	0,078	0,076	0	1
Var16	Creat sang	M	2	0,32	0,74	0,10	0,586	0,064	0,077	0,078	0	1
Var17	DIRBILI	F	1	0,69	0,69	0,01	0,541	0,044	0,060	0,053	0	0
Var17	DIRBILI	F	2	1,00	1,00	0,01	0,541	0,04	0,059	0,058	0	0
Var17	DIRBILI	M	1	0,27	0,58	0,01	0,869	0,047	0,042	0,043	1	0
Var17	DIRBILI	M	2	0,23	0,23	0,02	0,869	0,045	0,040	0,041	1	0
Var20	Glycémie	F	1	0,28	0,63	21,60	0,689	15,319	20,267	18,718	0	1
Var20	Glycémie	F	2	0,54	0,61	16,20	0,711	31,369	34,860	36,061	0	0
Var20	Glycémie	M	1	0,82	0,31	21,60	0,862	15,507	11,558	12,979	1	1
Var20	Glycémie	M	2	0,99	0,63	18,00	0,862	37,963	33,113	33,147	1	0
Var21	SGOTAST	F	1	0,86	0,63	20,20	0,772	24,074	26,899	25,360	0	0
Var21	SGOTAST	F	2	0,66	0,52	16,80	0,772	20,81	24,658	21,742	0	0
Var21	SGOTAST	M	1	0,65	0,41	31,30	0,673	14,782	19,975	18,423	0	1
Var21	SGOTAST	M	2	0,15	0,20	25,30	0,673	11,833	13,127	14,273	0	1
Var22	SGPTALT	F	1	0,66	0,74	8,70	0,906	14,519	11,101	11,593	1	0
Var22	SGPTALT	F	2	0,31	0,59	7,90	0,906	18,489	14,468	15,722	1	0
Var22	SGPTALT	M	1	0,81	0,29	11,40	0,728	6,224	7,001	7,026	0	1
Var22	SGPTALT	M	2	0,24	0,19	18,00	0,728	7,982	7,193	8,468	0	1
Var24	PHOSPH_Sang	F	1	0,60	0,71	0,62	0,774	0,851	0,819	0,838	1	0
Var24	PHOSPH_Sang	F	2	0,35	0,70	0,62	0,774	1,216	1,414	1,332	0	0
Var24	PHOSPH_Sang	M	1	0,55	0,28	0,62	0,781	0,752	0,796	0,815	0	0
Var24	PHOSPH_Sang	M	2	0,79	0,89	0,62	0,781	1,214	1,314	1,206	1	1
Var25	POTASSIUM_Sang	F	1	0,98	0,84	0,40	0,689	0,385	0,447	0,457	0	0
Var25	POTASSIUM_Sang	F	2	0,49	0,80	0,50	0,689	1,028	1,245	1,179	0	0
Var25	POTASSIUM_Sang	M	1	0,92	0,64	0,50	0,763	0,48	0,566	0,500	0	1
Var25	POTASSIUM_Sang	M	2	0,49	0,27	0,50	0,763	1,114	1,024	1,091	1	0
Var26	TOTPROT	F	1	0,49	0,64	0,40	0,670	0,371	0,431	0,423	0	0
Var26	TOTPROT	F	2	0,59	0,51	0,40	0,670	0,416	0,569	0,577	0	0
Var26	TOTPROT	M	1	0,76	0,22	0,40	0,682	0,322	0,339	0,398	0	1
Var26	TOTPROT	M	2	1,00	0,22	0,30	0,682	0,398	0,502	0,456	0	0
Var28	PTSEC	F	2	0,38	0,66	1,80	0,807	0,582	0,558	0,578	1	1
Var28	PTSEC	M	2	0,86	0,88	1,20	0,719	0,541	0,691	0,691	0	1
Var29	ABSBASO	F	1	0,60	0,61	0,62	0,855	0,016	0,014	0,015	1	1
Var29	ABSBASO	F	2	0,72	0,19	0,62	0,855	0,027	0,024	0,024	1	1
Var29	ABSBASO	M	1	0,58	0,37	0,62	0,556	0,022	0,030	0,035	0	0
Var29	ABSBASO	M	2	0,95	0,86	0,62	0,577	0,028	0,038	0,037	0	0
Var31	ABSEOS	F	1	0,46	0,30	0,62	0,750	0,069	0,064	0,081	0	0
Var31	ABSEOS	F	2	0,68	0,50	0,62	0,750	0,058	0,054	0,070	0	0
Var31	ABSEOS	M	1	0,42	0,95	0,62	0,706	0,071	0,083	0,083	0	0
Var31	ABSEOS	M	2	0,99	0,21	0,62	0,728	0,063	0,072	0,073	0	0

Les cases colorées en vert indiquent que l'écart-type des données bibliographiques ou commerciales selon le cas est plus élevé que le plus petit écart détectable (dose 33%) avec une puissance de 80%. Dans la colonne puissance : les ronds verts, oranges et rouges indiquent respectivement, une puissance supérieure ou égale à 80%, entre 70 et 80% et strictement inférieure à 70%.

Suite tableau 5

Paramètre	sexe	periode	pValue 1-3	pValue 2-4	Ecart-type des données historiques	Ecart-type des données commerciales		d80% : Delta détectable pour		Comparaison 33%		
						Puissance	Delta D2	Régime 11%	Régime 33%	Commerciale	Historique	
Var36	BASO	F	1	0,15	0,71	0,10	0,418	0,103	0,203	0,186	0	0
Var36	BASO	F	2	0,77	0,43	0,10	0,418	0,32	0,525	0,574	0	0
Var36	BASO	M	1	0,33	0,62	0,10	0,637	0,107	0,137	0,132	0	0
Var36	BASO	M	2	0,26	0,52	0,10	0,659	0,165	0,228	0,234	0	0
Var37	EOS	F	1	0,60	0,83	0,60	0,677	0,668	0,629	0,761	0	0
Var37	EOS	F	2	0,70	0,47	0,80	0,677	0,512	0,700	0,683	0	1
Var37	EOS	M	1	0,32	0,50	0,70	0,634	0,396	0,609	0,542	0	1
Var37	EOS	M	2	0,73	0,36	0,90	0,656	0,501	0,587	0,615	0	1
Var41	LYMPH	F	1	0,51	0,88	4,60	0,711	3,054	3,579	3,423	0	1
Var41	LYMPH	F	2	0,33	0,43	4,40	0,711					
Var41	LYMPH	M	1	0,55	0,60	5,40	0,716	3,425	3,847	3,860	0	1
Var41	LYMPH	M	2	0,39	0,86	6,30	0,738		4,684			
Var42	MCH	F	1	0,70	0,85	0,60	0,792	0,548	0,541	0,561	0	1
Var42	MCH	F	2	0,58	0,74	0,60	0,792	0,583	0,556	0,522	1	1
Var42	MCH	M	1	0,61	0,34	0,50	0,594	0,539	0,736	0,633	0	0
Var42	MCH	M	2	0,37	0,18	0,80	0,616	0,569	0,738	0,712	0	1
Var44	MCV	F	1	0,45	0,87	1,80	0,720	1,657	1,946	1,957	0	0
Var44	MCV	F	2	0,16	0,62	2,10	0,720	1,832	1,961	2,144	0	0
Var44	MCV	M	1	0,56	0,83	1,80	0,658	1,912	2,297	2,330	0	0
Var44	MCV	M	2	0,75	0,49	3,00	0,680	1,531	1,795	1,897	0	1
Var46	MONO	F	1	0,16	0,41	0,80	0,683	0,806	0,856	1,048	0	0
Var46	MONO	F	2	0,13	0,94	0,80	0,683	1,055	1,171	1,241	0	0
Var46	MONO	M	1	0,24	0,14	0,80	0,827	0,952	0,899	0,962	0	0
Var46	MONO	M	2	0,26	0,51	0,90	0,846	1,093	0,896	0,903	1	0
Var47	NEUT	F	1	0,74	0,50	4,30	0,723	3,356	4,698	3,796	0	1
Var47	NEUT	F	2	0,50	0,61	4,00	0,723	6,049	6,629	6,611	0	0
Var47	NEUT	M	1	0,30	0,59	5,30	0,749	3,931	4,715	4,652	0	1
Var47	NEUT	M	2	0,47	1,00	5,90	0,770	5,472	5,472	5,670	0	1
Var50	CHLORIDE	F	1	0,50	0,56		0,581	29,529	42,788	47,170	0	
Var50	CHLORIDE	F	2	0,88	0,42	27,32	0,581	23,389	54,188	31,872	0	0
Var50	CHLORIDE	M	1	0,46	0,61		0,503	27,189	38,235	45,958	0	
Var50	CHLORIDE	M	2	0,60	0,79	20,41	0,503	27,316	45,713	44,299	0	0
Var51	PHOSPH	F	1	0,97	0,84		0,739	88,982	114,832	98,077	0	
Var51	PHOSPH	F	2	0,87	0,82		0,739	89,206	128,352	95,203	0	
Var51	PHOSPH	M	1	0,63	0,45		0,693	55,097	68,244	60,824	0	
Var51	PHOSPH	M	2	0,65	0,73		0,693	53,268	72,086	65,726	0	
Var52	CREAT	F	1	0,71	0,92		0,585	49,015	67,976	66,901	0	
Var52	CREAT	F	2	0,75	0,91		0,585	47,518	78,631	66,507	0	
Var52	CREAT	M	1	0,15	0,67		0,635	32,648	44,064	38,521	0	
Var52	CREAT	M	2	0,20	0,99		0,635	38,473	56,173	46,227	0	
Var53	POTASSIUM	F	1	0,66	0,80		0,672	65,551	84,588	75,720	0	
Var53	POTASSIUM	F	2	0,95	0,84	27,32	0,672	68,822	90,166	78,776	0	0
Var53	POTASSIUM	M	1	0,13	0,31		0,566	49,938	63,579	66,377	0	
Var53	POTASSIUM	M	2	0,39	0,99	40,82	0,566	53,31	70,769	67,749	0	0
Var55	NAKRATIO	F	1	0,35	0,16		0,643	0,119	0,176	0,159	0	
Var55	NAKRATIO	F	2	0,43	0,16		0,643	0,133	0,162	0,169	0	
Var55	NAKRATIO	M	1	0,24	0,71		0,761	0,123	0,107	0,138	0	
Var55	NAKRATIO	M	2	0,64	0,81		0,782	0,127	0,118	0,131	0	
Var56	SODIUM	F	1	0,25	0,18		0,609	18,307	23,115	24,333	0	
Var56	SODIUM	F	2	0,31	0,54	27,32	0,587	18,149	30,452	25,238	0	1
Var56	SODIUM	M	1	0,98	0,26		0,690	20,576	20,968	22,658	0	
Var56	SODIUM	M	2	0,96	0,76	20,41	0,690	19,436	26,136	25,708	0	0
Var58	VOLTOTAL	F	1	0,86	0,75	2,61	0,592	2,881	4,842	4,750	0	0
Var58	VOLTOTAL	F	2	0,96	0,92	5,00	0,592	3,355	4,200	4,320	0	1
Var58	VOLTOTAL	M	1	0,25	0,90	2,05	0,587	3,2	3,737	4,512	0	0
Var58	VOLTOTAL	M	2	0,66	0,94	6,15	0,587	2,878	3,325	3,773	0	1

4. Interprétation statistique

Sur le plan statistique, cette étude de cas montre l'importance de la recherche des données atypiques. Leur analyse par le toxicologue permettra de déterminer s'il est justifié ou non de les éliminer et de vérifier les conditions d'application des tests paramétriques.

L'analyse statistique menée sur les données de l'étude MON810 montre que moins de 10% (33 sur 432 tests) des tests de différences sont significatifs à 10% (p -value < 0.10). En d'autres termes, en raison de la multiplicité des tests, la probabilité d'obtenir un test faussement significatif est aussi d'environ 10%. Ces tests de différences statistiquement significatifs doivent donc être interprétés comme des signaux d'alerte que les toxicologues devront analyser avec attention, mais pas comme une preuve de l'existence d'une réelle différence.

L'approche basée sur des tests de différence unidimensionnels doit être complétée par un calcul de puissance de manière à identifier les paramètres biologiques pour lesquels un manque de puissance est avéré. Cependant, le calcul de puissance nécessite la définition de seuils de tolérance, c'est-à-dire du niveau d'effet à partir duquel la différence a un sens toxicologique. La définition de ces seuils est une tâche difficile et doit être réalisée à partir des connaissances toxicologiques disponibles, des données historiques, et des données obtenues avec les variétés commerciales.

Les seuils utilisés dans le cadre de cette étude ont été déterminés à partir des données bibliographiques et des groupes témoins (nourris avec des maïs commerciaux). La référence choisie est égale à un écart-type des valeurs historiques ou des valeurs observées chez les animaux nourris avec les variétés commerciales.

Dans la présente application, selon l'origine des données, bibliographiques ou variétés commerciales, la puissance est insuffisante dans respectivement 50 et 80% des tests. Les paramètres présentant une différence non significative mais une puissance insuffisante doivent aussi être examinés par le toxicologue. Les calculs de puissance montrent qu'une augmentation des effectifs à 20 animaux par groupe et par sexe au lieu de 10 conduirait à une puissance satisfaisante pour presque tous les paramètres ; cette augmentation des effectifs est de nature à fiabiliser et à faciliter l'analyse du toxicologue.

Des tests « d'équivalence » ont été réalisés afin de tester l'hypothèse nulle « différences OGM vs. ISO supérieures aux seuils de tolérance ». Les résultats montrent que, dans la quasi-totalité des cas, cette hypothèse ne peut pas être rejetée. L'absence de test d'équivalence significatif n'est pas nécessairement due à un effet régime OGM mais pourrait être liée à un effectif insuffisant et aux choix des seuils de tolérance.

La démarche décrite par l'EFSA a été appliquée et conduit à des seuils de tolérance encore inférieurs à ceux que nous proposons. Cette méthode a été conçue pour analyser des paramètres décrivant des caractéristiques variétales qui présentent par essence de grandes différences d'une variété à l'autre. Nous considérons ici un essai de toxicité sub-chronique de 90 jours dans lequel les paramètres biologiques correspondent à des mesures réalisées non pas sur les plantes mais sur des animaux de laboratoire. Nous avons observé une faible variabilité des paramètres biologiques mesurés sur les rats nourris avec des régimes contenant différentes variétés commerciales de maïs. La démarche décrite par l'EFSA ne semble donc pas adaptée à l'analyse de ces données.

Les tests d'équivalence nécessitent des effectifs élevés pour atteindre une puissance suffisante (80%). A titre d'exemple, il serait nécessaire de mettre en œuvre plusieurs centaines d'animaux par groupe et par sexe pour démontrer l'équivalence sur la base d'un delta de 1 écart type.

Il apparaît donc important d'initier un travail approfondi sur la détermination des seuils de tolérance pour l'ensemble des paramètres biologiques mesurés dans cette étude. La détermination des seuils de tolérance ou des tailles d'effet, permettra une interprétation plus précise des différences observées.

Enfin, les tests d'équivalence comme les tests de différences ne tiennent pas compte de l'aspect multidimensionnel de l'analyse.

5. Interprétation toxicologique

Rappel des principes de l'analyse toxicologique

Si le toxicologue examine plus particulièrement les données qui concluent à des différences significatives entre les animaux témoins et « traités », son analyse ne se limite pas pour autant à la comparaison des valeurs moyennes des différents groupes de l'étude. Les données individuelles sont prises en compte, notamment lorsque les écarts type apparaissent élevés, traduisant des comportements atypiques de quelques animaux au sein d'un groupe.

Bien que la mise en évidence d'effets significatifs dans les deux sexes soit un élément qui renforce l'avis du toxicologue, un effet répondant aux critères habituels de jugement (effet dose, effet temps, convergence de paramètres..) dans un seul sexe sera pris en compte.

De même, l'analyse ne repose pas uniquement sur les valeurs ou comparaisons qui ont une signification statistique. **Une augmentation dose-dépendante ou durée du traitement dépendante, même non statistiquement significative, "retiendra" l'attention du toxicologue et sera prise en compte dans ses conclusions, si la donnée est jugée pertinente.**

D'une façon générale, l'analyse des données d'une étude toxicologique repose sur 1) la signification intrinsèque de la variation d'un paramètre, 2) le sens de cette variation en relation avec l'altération d'un ensemble de paramètres potentiellement liés à un effet sur un organe ou un système et 3) la notion de convergence de paramètres traduisant un effet sur une cible.

Ainsi, la modification d'un paramètre isolé dans un sens ou dans un autre, n'a pas obligatoirement de valeur toxicologique, même si cette modification est statistiquement significative.

De plus, si l'analyse et la prise en compte des paramètres cliniques et physiologiques (état général, évolution pondérale, consommation alimentaire, données hématologiques, résultats biochimiques, sanguins et urinaires, poids des organes,...) sont nécessaires, elles n'en sont pas moins insuffisantes.

La convergence de paramètres constituant la signature d'un effet sur un organe cible ou un système est confortée par des altérations histologiques permettant de conclure, avec plus de pertinence, à un effet ayant une signification toxicologique. L'histologie demeure, en effet, un élément clé à partir duquel un effet toxique peut être affirmé.

Enfin, il convient de rappeler que l'analyse d'une étude de toxicologie n'est qu'une composante de l'évaluation du risque.

Données atypiques

L'analyse de la distribution des données fait apparaître 38 données atypiques sur 21281. Un examen de ces données a été réalisé par les toxicologues afin de déterminer si ces valeurs étaient potentiellement liées à un effet du traitement OGM. Ces valeurs sont réparties dans tous les groupes, avec une prévalence supérieure dans les groupes 1 (MON810 11%) et 9 (variété commerciale).

Sept valeurs atypiques concernent toutefois la seule femelle 14 du groupe 1 (OGM 11%) en période 2 (figure 9). Dans ce cas, il convenait de s'interroger sur la cohérence de ces valeurs en relation avec un effet intrinsèque (effet régime) ou extrinsèque (pathologie intercurrente, erreurs de mesures...). L'examen des données individuelles de la femelle 14 ne met pas en évidence d'altération de l'évolution pondérale et de la consommation alimentaire. De même, aucun signe clinique particulier n'apparaît. L'origine de ces données atypiques chez cet animal uniquement à la dose la plus faible n'a pu être identifiée. La nature des modifications et leur présence à la seule dose faible, conduit le toxicologue à les considérer comme n'ayant pas de signification toxicologique.

Globalement, étant donné leur faible nombre et leur dispersion, ces données atypiques ne sont pas des marqueurs d'alerte d'une quelconque toxicité et un lien avec le traitement ne peut être établi. Dans ces conditions, et compte tenu de contraintes statistiques, ces valeurs pourraient être écartées de l'analyse statistique paramétrique.

Tests de différence

La mise en œuvre des tests paramétriques conduit à identifier 33 différences significatives entre les groupes OGM et les groupes ISO. Elles sont globalement plus nombreuses en période 2 qu'en période 1 pour la dose la plus faible, alors que c'est l'inverse qui est observé pour la dose la plus forte. Cette absence de relation dose/effet en fonction de la durée d'exposition au traitement, conduit à pondérer toute variation significative en période 1, non confirmée en période 2.

Deux paramètres (ALBUMIN et AGRATIO) présentent des différences significatives entre les moyennes OGM versus ISO à la dose la plus forte aux 2 périodes, chez les mâles uniquement. Ces deux paramètres sont liés, car la modification du taux d'albumine se répercute sur le rapport albumine/globuline (AGRATIO).

Le fait que ces paramètres liés ne concernent qu'un sexe, et qu'il n'y ait pas de lien avec d'autres paramètres qui renforceraient la présomption de toxicité, notamment d'altérations histologiques de l'organe cible conduit à ne pas considérer ces variations comme ayant un sens toxicologique.

En conclusion, la lecture du toxicologue conduit à estimer qu'au risque α de 10%, aucune des 33 différences significatives observées sur les 432 comparaisons n'a suffisamment de pertinence pour conclure à un quelconque effet toxique de l'OGM.

Multiplicité des tests

La multiplicité des tests conduit à augmenter le risque de faux positifs, c'est à dire le fait de conclure à tort à une différence alors qu'elle n'existe pas.

Il convient donc de prendre en compte ce risque en appliquant des facteurs de corrections.

Dans ces conditions, le nombre de faux positifs potentiels est quasiment égal au nombre de tests significatifs pour un niveau de risque $\alpha=10\%$. Ainsi, après correction de la p-value par la méthode de Benjamini et Hochberg, aucune différence n'apparaît significative. La prise en compte des facteurs de correction est une démarche nécessaire de pondération des résultats, dont le toxicologue doit avoir connaissance dans le cadre de son évaluation de la sécurité.

Choix des bornes de tolérance pour le calcul de puissance

Comme indiqué précédemment, le calcul de puissance *a posteriori* ne peut avoir lieu sans une détermination de la taille d'effet. Cette taille d'effet correspond, pour une puissance donnée, à la différence entre les moyennes des deux groupes témoin et traité pouvant être mise en évidence lors de l'étude. Cette différence devra être comparée aux différences jugées anormales ou en dehors de la variabilité naturelle pour ce paramètre.

L'analyse toxicologique reposant sur un faisceau d'éléments qui permet de conclure à une toxicité, la prise en compte de chaque paramètre isolé pour définir une taille d'effet est en soit insuffisante. Idéalement, la taille de l'effet devrait prendre en compte l'ensemble des paramètres convergents. Toutefois, afin de répondre à la question posée par les statisticiens, le toxicologue a proposé de prendre comme taille d'effet minimum celle correspondant à la variabilité observée naturellement chez des animaux témoins pour chacun des paramètres mesurés. Cette valeur, correspondant à 1 déviation standard, est la plus petite valeur caractérisant une variation d'un paramètre, disponible *via* les données issues d'animaux témoins (données historiques).

C'est ainsi que deux références ont été proposées. La première « SD » correspond à l'écart type de données historiques de rat de la même souche décrite dans deux publications¹⁶. Cette référence peut être considérée comme une référence externe à l'expérience. La seconde est une référence interne et correspond à un écart type calculé à partir des données des groupes nourris avec les variétés commerciales « D2 » dans la même étude.

Il est à noter que la littérature fait plutôt état de la prise en compte de 1.95 fois l'écart-type estimé comme borne de tolérance qui correspond à une fourchette de valeurs pour lesquelles 95% des données sont incluses (EFSA, 2008a). Le seuil de 1 écart type qui a été proposé apparaît donc comme conservateur par rapport aux variations spontanées de certains paramètres.

¹⁶ Petterino and Argentino-Storino, 2006, Exp. and Tox. Pathol. 57, 213-219 et Wolford *et al.*, 1986, J. Tox and Environ. Health, 18, 161-188. ; A l'avenir ce sont les données historiques du centre investigateur qui devront être disponibles.

Analyse de la puissance sur la base des bornes de référence précédemment définies.

Au-delà de l'analyse des différences significatives qui sont systématiquement prises en compte par le toxicologue, il est nécessaire de considérer la puissance du test.

Le principe qui a été retenu est celui de critères d'évaluation choisis de façon à augmenter le nombre de différences significatives que le toxicologue aura à examiner, sans préjuger de leur signification toxicologique. C'est ce qui a conduit à fixer une p-value à 10%, versus 5% habituellement proposée dans ce type d'étude.

De même, l'approche retenue pour choisir les bornes ou la taille d'effet s'est voulue conservatrice. Ce choix est important puisqu'il conditionne en retour la puissance du test.

Dans ces conditions, il revient au toxicologue de juger *in fine* si une insuffisance de puissance découlant d'un choix de bornes est de nature à masquer un effet potentiellement toxique.

Un calcul théorique de puissance réalisé sur la base des données de l'étude MON810 conclut à une puissance suffisante pour l'ensemble des paramètres si les mesures sont réalisées sur 20 animaux au lieu des 10 pris en compte dans l'étude. Dans le protocole expérimental de l'étude MON810, la prise en compte des mesures sur 20 animaux a d'autant plus de sens que c'est l'effectif qui a été mis en œuvre pour chacun des groupes.

Considérant que les données de l'étude, en dehors des mesures de l'évolution pondérale et des poids d'organes réalisées chez tous les animaux (n=20), portaient sur 10 animaux, il convenait d'analyser 1) la puissance du test dans ces conditions et 2) la pertinence de la taille de l'effet en regard de sa signification toxicologique.

Cette approche a été appliquée à tous les paramètres de l'étude qui présentaient une insuffisance de puissance d'après le calcul statistique. Le principe de la démarche suivie par les toxicologues est illustré par différents exemples issus des données hématologiques, biochimiques sanguines et urinaires et du poids des organes (annexe n°12).

Bien que les tests statistiques ne soient pas applicables aux examens histologiques et qu'en conséquence, il n'y ait pas de calcul de puissance sur ces résultats microscopiques, la prise en compte de ces données est essentielle dans le jugement du toxicologue pour conclure *in fine* à une toxicité sur un système ou un organe.

Pour les paramètres dont la puissance est inférieure à 80%, l'examen des bornes D2 ou SD proposées pour juger de la taille de l'effet, associé à la prise en compte des variations "spontanées" mini-maxi de chacun des paramètres et de la cohérence des variations dans une approche multi-paramètres, permet d'estimer, qu'à l'exception des basophiles, aucune mesure ne présente une taille détectable susceptible d'induire un jugement erroné par rapport aux tests de différences mis en œuvre.

En conclusion, au regard de l'ensemble des données cliniques et biologiques et de l'absence d'altérations macroscopiques et microscopiques des organes examinés, les toxicologues concluent à l'absence de différence entre groupes traités OGM et groupes témoins, susceptible de traduire une toxicité du produit étudié dans les conditions expérimentales rapportées.

Conclusion générale

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 les rats.

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 1^{ère} 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 en reproductibilité.

Bibliographie

- AFSSA (2002) Evaluation des risques relatifs à la consommation de produits alimentaires composés ou issus d'organismes génétiquement modifiés (Risk assessment for the consumption of foods composed of or derived from GMOs). **annexe 1**.
<http://www.afssa.fr/Documents/BIOT-Ra-ConsoOGM.pdf>
- AFSSA (2007) Avis relatif aux études de toxicité réalisées dans le cadre des demandes de mises sur le marché d'OGM (Opinion on toxicity studies carried out in the context of applications for marketing authorisations for GMOs). Avis du 29 février 2008. *saisine 2007-SA-0396*.
<http://www.afssa.fr/Documents/BIOT2007sa0396.pdf>
- Anderson, S., Hauck, W.W. (1983) A new procedure for testing equivalence in comparative bioavailability and other clinical trials. *Communications in Statistics - Theory and Methods* **12**, 2663 - 2692.
<http://www.informaworld.com/10.1080/03610928308828634>
- Appenzeller, L.M., Malley, L., Mackenzie, S.A., Hoban, D., Delaney, B. (2009a) Subchronic feeding study with genetically modified stacked trait lepidopteran and coleopteran resistant (DAS-O1507-1xDAS-59122-7) maize grain in Sprague-Dawley rats. *Food Chem Toxicol* **47**, 1512-1520.
- Appenzeller, L.M., Munley, S.M., Hoban, D., Sykes, G.P., Malley, L.A., Delaney, B. (2008) Subchronic feeding study of herbicide-tolerant soybean DP-356O43-5 in Sprague-Dawley rats. *Food Chem Toxicol* **46**, 2201-2213.
- Appenzeller, L.M., Munley, S.M., Hoban, D., Sykes, G.P., Malley, L.A., Delaney, B. (2009b) Subchronic feeding study of grain from herbicide-tolerant maize DP-O9814O-6 in Sprague-Dawley rats. *Food Chem Toxicol* **47**, 2269-2280.
- 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**, 289-300.
- CE (1997) Règlement (CE) n° 258/97 du Parlement européen et du Conseil du 27 janvier 1997 relatif aux nouveaux aliments et aux nouveaux ingrédients alimentaires. *JOCE n° L 043 du 14/02/1997* 1-6.
- Chen, C., Rathore, N., Wenchang, J., Germansderfer, A. (2010) Statistical equivalence testing for assessing bench-scale cleanability. *Biopharm International* **23**, 40-45.
- Cohen, J. (1992) A power primer. *Psychological bulletin* **112**, 155-159.
- Dixon, P.M., Pechmann, J.H.K. (2005) A statistical test to show negligible trend. *Ecology* **86**, 1751-1756.
- Dryzga, M.D., Yano, B.L., Andrus, A.K., Mattsson, J.L. (2007) Evaluation of the safety and nutritional equivalence of a genetically modified cottonseed meal in a 90-day dietary toxicity study in rats. *Food Chem Toxicol* **45**, 1994-2004.
- Dunn, O.D. (1964) Multiple Comparisons Using Rank Sums. *Technometrics* **6**, 241-262.
- 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.
<http://www.efsa.europa.eu/en/scdocs/doc/99.pdf>
- EFSA (2008a) **DRAFT** Updated Guidance document for the risk assessment of genetically modified plants and derived food and feed *The EFSA Journal* **727**, 1-135.
<http://www.efsa.europa.eu/en/scdocs/doc/293rax1.pdf>
- EFSA (2008b) Safety and nutritional assessment of GM plants and derived food and feed: the role of animal feeding trials. *Food Chem Toxicol* **46 Suppl 1**, S2-70.
- EFSA (2009) Scientific opinion on statistical considerations for the safety evaluation of GMOs, on request of EFSA. *The EFSA Journal* **1250**, 1-62. <http://www.efsa.europa.eu/en/scdocs/doc/1250.pdf>
- EFSA (2010) Scientific opinion on the assessment of potential impacts of genetically modified plants on non-target organisms *The EFSA Journal* **8**, 1877.
<http://www.efsa.europa.eu/en/scdocs/doc/1877.pdf>

- EMEA, E.M.A. (1998) Non-clinical guideline on drug-induced hepatotoxicity, 24 January 2008. *EMEA/CHMP/SWP/150115/2006*.
<http://www.ema.europa.eu/pdfs/human/swp/15011506en.pdf>
- EMEA, E.M.A. (2008) Guideline on the investigations of bioequivalence, 24 July 2008. *CPMP/EWP/QWP/1401/98, 2008 rev1*.
<http://www.ema.europa.eu/pdfs/human/qwp/140198enrev1.pdf>
- Garrett, K.A. (1997) Use of statistical tests of equivalence (bioequivalence tests) in plant pathology. *Phytopathology* **87**, 372-374.
- Grubbs, F. (1969) Procedures for detecting outlying observations in samples. *Technometrics* **11**, 1-21.
- Hammond, B., Dudek, R., Lemen, J., Nemeth, M. (2004) Results of a 13 week safety assurance study with rats fed grain from glyphosate tolerant corn. *Food Chem Toxicol* **42**, 1003-1014.
- Hammond, B., Lemen, J., Dudek, R., Ward, D., Jiang, C., Nemeth, M., Burns, J. (2006a) Results of a 90-day safety assurance study with rats fed grain from corn rootworm-protected corn. *Food Chem Toxicol* **44**, 147-160.
- Hammond, B.G., Dudek, R., Lemen, J.K., Nemeth, M.A. (2006b) Results of a 90-day safety assurance study with rats fed grain from corn borer-protected corn. *Food Chem Toxicol* **44**, 1092-1099.
- Hammond, B.G., Lemen, J.K., Ahmed, G., Miller, K.D., Kirkpatrick, J., Fleeman, T. (2008) Safety assessment of SDA soybean oil: results of a 28-day gavage study and a 90-day/one generation reproduction feeding study in rats. *Regul Toxicol Pharmacol* **52**, 311-323.
- He, X.Y., Huang, K.L., Li, X., Qin, W., Delaney, B., Luo, Y.B. (2008) Comparison of grain from corn rootworm resistant transgenic DAS-59122-7 maize with non-transgenic maize grain in a 90-day feeding study in Sprague-Dawley rats. *Food Chem Toxicol* **46**, 1994-2002.
- He, X.Y., Tang, M.Z., Luo, Y.B., Li, X., Cao, S.S., Yu, J.J., Delaney, B., Huang, K.L. (2009) A 90-day toxicology study of transgenic lysine-rich maize grain (Y642) in Sprague-Dawley rats. *Food Chem Toxicol* **47**, 425-432.
- Healy, C., Hammond, B., Kirkpatrick, J. (2008) Results of a 13-week safety assurance study with rats fed grain from corn rootworm-protected, glyphosate-tolerant MON 88017 corn. *Food Chem Toxicol* **46**, 2517-2524.
- Hochberg, Y., Tamhane, A.C., (1987) Multiple comparison procedures. *New York: John Wiley & Sons*.
- MacKenzie, S.A., Lamb, I., Schmidt, J., Deege, L., Morrissey, M.J., Harper, M., Layton, R.J., Prochaska, L.M., Sanders, C., Locke, M., Mattsson, J.L., Fuentes, A., Delaney, B. (2007) Thirteen week feeding study with transgenic maize grain containing event DAS-O1507-1 in Sprague-Dawley rats. *Food Chem Toxicol* **45**, 551-562.
- Malley, L.A., Everds, N.E., Reynolds, J., Mann, P.C., Lamb, I., Rood, T., Schmidt, J., Layton, R.J., Prochaska, L.M., Hinds, M., Locke, M., Chui, C.F., Claussen, F., Mattsson, J.L., Delaney, B. (2007) Subchronic feeding study of DAS-59122-7 maize grain in Sprague-Dawley rats. *Food Chem Toxicol* **45**, 1277-1292.
- OCDE (1993) Safety evaluation of food derived by modern biotechnology : concept and principles. *Paris, France*. <http://www.oecd.org/dataoecd/37/18/41036698.pdf>
- OCDE (1998) Ligne directrice de l'OCDE pour les essais de produits chimiques Toxicité orale à doses répétées - rongeurs : 90 jours. *OCDE/OECD*.
<http://browse.oecdbookshop.org/oecd/pdfs/browseit/9740802E.PDF>
- OECD (1998) Guideline for testing of chemicals N°408. Repeated Dose 90-day Oral Toxicity Study in Rodents. *Paris, France*.
- Perry, J.N., Ter Braak, C.J., Dixon, P.M., Duan, J.J., Hails, R.S., Huesken, A., Lavielle, M., Marvier, M., Scardi, M., Schmidt, K., Tothmeresz, B., Schaarschmidt, F., van der Voet, H. (2009) Statistical aspects of environmental risk assessment of GM plants for effects on non-target organisms. *Environ Biosafety Res* **8**, 65-78.
- Petterino, C., Argentino-Storino, A. (2006) Clinical chemistry and haematology historical data in control Sprague-Dawley rats from pre-clinical toxicity studies. *Exp Toxicol Pathol* **57**, 213-219.

- Poulsen, M., Kroghsbo, S., Schroder, M., Wilcks, A., Jacobsen, H., Miller, A., Frenzel, T., Danier, J., Rychlik, M., Shu, Q., Emami, K., Sudhakar, D., Gatehouse, A., Engel, K.H., Knudsen, I. (2007a) A 90-day safety study in Wistar rats fed genetically modified rice expressing snowdrop lectin *Galanthus nivalis* (GNA). *Food Chem Toxicol* **45**, 350-363.
- Poulsen, M., Schroder, M., Wilcks, A., Kroghsbo, S., Lindecrona, R.H., Miller, A., Frenzel, T., Danier, J., Rychlik, M., Shu, Q., Emami, K., Taylor, M., Gatehouse, A., Engel, K.H., Knudsen, I. (2007b) Safety testing of GM-rice expressing PHA-E lectin using a new animal test design. *Food Chem Toxicol* **45**, 364-377.
- Schroder, M., Poulsen, M., Wilcks, A., Kroghsbo, S., Miller, A., Frenzel, T., Danier, J., Rychlik, M., Emami, K., Gatehouse, A., Shu, Q., Engel, K.H., Altosaar, I., Knudsen, I. (2007) A 90-day safety study of genetically modified rice expressing Cry1Ab protein (*Bacillus thuringiensis* toxin) in Wistar rats. *Food Chem Toxicol* **45**, 339-349.
- Schuirman, D.J. (1987) A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *J Pharmacokinet Biopharm* **15**, 657-680.
- Storey, J.D. (2002a) A direct approach to false discovery rates. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **64**, 479-498.
- Storey, J.D. (2002b) False discovery rates: Theory and applications to DNA microarrays. *Ph.D. dissertation Dept. Statistics, Stanford Univ.*
- Storey, J.D. (2003) The positive false discovery rate: A Bayesian interpretation and the q-value. *annals of statistics* **31**, 2013-2035.
- Tryon, W.W., Lewis, C. (2008) An inferential confidence interval method of establishing statistical equivalence that corrects Tryon's (2001) reduction factor. *Psychol Methods* **13**, 272-277.
- Wang, Z., Wang, Y., Cui, H., Xia, Y., Altosaar, I., Shu, Q. (2002) Toxicological evaluation of transgenic rice flour with a synthetic cry1Ab gene from *Bacillus thuringiensis*. *Journal of the Science of Food and Agriculture* **82**, 738-744.
- Westfall, P.H., Young, S.S., (1993) Resampling-based multiple testing : examples and methods for P-value adjustment, New York: John Wiley & Sons Edition.
- Westlake, W.J. (1976) Symmetrical confidence intervals for bioequivalence trials. *Biometrics* **32**, 741-744.
- 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*.
<ftp://ftp.fao.org/es/esn/food/gmreport.pdf>
- Wilcoxon, F. (1945) Individual comparisons by ranking methods. *Biometrics* **1**, 80-83.
- Wolford, S.T., Schroer, R.A., Gohs, F.X., Gallo, P.P., Brodeck, M., Falk, H.B., Ruhren, R. (1986) Reference range data base for serum chemistry and hematology values in laboratory animals. *J Toxicol Environ Health* **18**, 161-188.

ANNEXES

ANNEXE 1 : Tableau de synthèse de la revue bibliographique (Protocole expérimental)

Tableau de synthèse des principaux éléments méthodologiques des études de toxicité sub-chronique de 90 jours réalisées chez les rats à partir d'aliments issus ou consistant à des OGM publiées dans la littérature.

Références Articles	Espèce événement caractère	Protéines exprimées	Mat. végétal	Condition Rats	Groupes Traitements	centre investigateur	Paramètres toxicologiques observés	périodes et effectifs
(Appenzeller <i>et al.</i> , 2008) Subchronic feeding study of herbicide-tolerant soybean DP-356043-5 in Sprague-Dawley rats. Food and Chem. Tox. 46 2201-2213.	Soja 356043 tolérant à des herbicides.	GAT GM-HRA	Tourteau x toastés et coques de soja	12 rats/groupe/sexe Hébergement individuel rats CrI:CD (SD) IGS BR Charles River.	6 groupes (20% de tourteau et 1,5% de coques): OGM 356043 non traité, OGM 356043 traité glyphosate chlorimuron thifensulfuron, contrôle quasi isogénique 091, 3 variétés commerciales.	Dupont Haskell Global Centers for Health and environmental Sciences; BPL	Observations cliniques, ophtalmologie, évaluation comportement, pathologie clinique, hématologie et coagulation, chimie sanguine, analyse d'urine, croissance et pathologie anatomique	13 semaines 12 rats
(Appenzeller <i>et al.</i> , 2009a) Subchronic feeding study with GM stacked trait lepidopteran and coleopteran resistant (DAS-1507-1xDAS-59122-7)maize grain. Food and Chem. Tox. 47 1512-1520	Maïs 1507x59122 tolérant à des lépidoptères et herbicide.	Cry1F PAT Cry1Ab34 Cry1Ab35 PAT	Grain moulu	12 rats/groupe/sexe Hébergement individuel rats CrI:CD (SD) IGS BR Charles River.	6 groupes 34% d'incorporation : OGM 1507x59122; contrôle quasi isogénique 091; 3 variétés commerciales.	Dupont Haskell Global Centers for Health and environmental Sciences; BPL	Observations cliniques, ophtalmologie, évaluation comportement, pathologie clinique, hématologie et coagulation, chimie sanguine, analyse d'urine, croissance et pathologie anatomique	13 semaines 12 rats
(Appenzeller <i>et al.</i> , 2009b) Subchronic feeding study of grain from herbicide tolerant maize DP-98140-6 in Sprague Dawley rats. Food and Chem. Tox. 47 2269-2280.	Maïs 98140 tolérance glufosinate d'ammonium	GAT	Grain moulu	12 rats/groupe/sexe Hébergement individuel rats CrI:CD (SD)IGS BR Charles River	6 groupes 35-38% d'incorporation: OGM 98140 non traité; OGM 98140 traité Gly; contrôle quasi isogénique 091; 3 variétés commerciales	Dupont Haskell Global Centers for Health and environmental Sciences; BPL	Observations cliniques, ophtalmologie, évaluation comportement, pathologie clinique, hématologie et coagulation, chimie sanguine, analyse d'urine, croissance et pathologie anatomique.	13 semaines 12 rats
(Dryzga <i>et al.</i> , 2007) Evaluation of the safety and nutritionnal equivalence of a GM cottonseed meal in a 90 day dietary toxicity study in rats. Food and Chem. Tox. 45 1994-2004.	Cotonnier tolérance à des lépidoptères et glufosinate d'ammonium	Cry1F Cry1Ac PAT	Tourteau x	12 rats/groupe/sexe rats Sprague Dawley	5 groupes 10% d'incorporation OGM Widestrike; contrôle lignée quasi isogénique; 3 lignées commerciales	Dow chemical BPL	Consommation et croissance pondérale, Observations cliniques, pathologie clinique, hématologie et coagulation, chimie sanguine, analyse d'urine, histopathologie, poids des organes.	13 semaines 12 rats
(Hammond <i>et al.</i> , 2004) Results of a 13 week safety assurance study with rats fed grain from glyphosate tolerant corn. Food and Chem. Tox. 42 1003-1014.	Maïs NK603 tolérant au glyphosate	CP4 EPSPS	Grain	20 rats/groupe/sexe Hébergement individuel rats CrI:CD (SD) IGS BR Charles River	10 groupes 11 ou 33% d'incorporation : OGM NK603 33%; OGM NK603 11%; contrôle quasi isogénique 11%; contrôle quasi isogénique 33%; 6 variétés commerciales 33%	Covance Laboratories BPL	Consommation et croissance pondérale, Observations cliniques, pathologie clinique, hématologie et coagulation, chimie sanguine, analyse d'urine, histopathologie, poids des organes	13 semaines 10 rats: 20 rats pour croissance pondérale et consommation.
(Hammond <i>et al.</i> , 2006b) Results of a 90-day safety assurance study with rats fed grain from corn borer-protected corn. Food and chem. Tox. 44 1092-1099	Maïs MON810 tolérant à des lépidoptères	Cry1Ab	Grain	20 rats/groupe/sexe Hébergement individuel rats CrI:CD (SD) IGS BR Charles River	10 groupes 11 ou 33% d'incorporation : OGM MON810 33%; OGM MON810 11%; contrôle quasi isogénique 11%; contrôle quasi isogénique 33%; 6 variétés commerciales 33%.	Covance Laboratories BPL	Consommation et croissance pondérale, Observations cliniques, pathologie clinique, hématologie et coagulation, chimie sanguine, analyse d'urine, histopathologie, poids des organes	13 semaines 10 rats: 20 rats pour croissance pondérale et consommation.
(Hammond <i>et al.</i> , 2006a) Results of a 90-day safety assurance study with rats fed grain from rootworm-protected corn. Food and chem. tox. 44 147-160	Maïs MON863 tolérant à des coléoptères	Cry3Bb1 NptII	Grain	20 rats/groupe/sexe Hébergement individuel rats CrI:CD (SD) IGS BR Charles River	10 groupes 11 ou 33% d'incorporation : OGM MON863 33%; OGM MON863 11%; contrôle quasi isogénique 11%; contrôle quasi isogénique 33%; 6 variétés commerciales 33%	Covance Laboratories BPL	Consommation et croissance pondérale, Observations cliniques, pathologie clinique, hématologie et coagulation, chimie sanguine, analyse d'urine, histopathologie, poids des organes	13 semaines 10 rats: 20 rats pour croissance pondérale et consommation.

Références Articles	Espèce événement caractère	Protéines exprimées	Mat. végétal	Condition Rats	Groupes Traitements	centre investigateur	Paramètres toxicologiques observés	périodes et effectifs
(Hammond <i>et al.</i> , 2008) Safety assesment of SDA soybean oil: Results of a 28-day study and a 90-day/one generation reproduction feeding study in rats. Regul. Tox.and Pharmacol. 52, 311-323.	Soja riche en SDA (acide gras omega-3)	Delta 6 et delta 15 desaturases	huile de soja	25 rats/groupe/sexe rats Crl:CD (SD)IGS BR Charles River	4 groupes 1,5 g ou 4g/kg/jour huile soja GM SDA 1,5g; huile soja GM SDA 4g; huile de soja contrôle 4g; huile de poisson (Menhaden).	Covance Laboratories BPL	Consommation et croissance pondérale, hématologie, chimie sanguine, poids des organes, histopathologie	13 semaines 10 rats: 20 rats pour croissance pondérale et consommation.
(He <i>et al.</i> , 2008) Comparison of grain from corn rootworm resistant T DAS -59122-7 maize with non-T maize grain in a 90 day feeding study in sprague dawley rats. Food and Chem. Toxicol. 46 1994-2002.	maïs DAS-59122 tolérant à des lépidoptères et au glufosinate d'ammonium	Cry34Ab1 Cry35Ab1 PAT	Farine de maïs	10 rats/groupe/sexe Hébergement individuel rats Sprague Dawley Peking university	5 groupes 50 ou 70% d'incorporation : OGM 59122 50%; OGM 59122 70%; contrôle quasi isogénique 50%; contrôle quasi isogénique 70%; 43,3% farine maïs (régime témoin)	Experimental animal center of Peking University, Beijing, China	Consommation et croissance pondérale, hématologie, chimie sanguine, poids des organes et pathologie	13 semaines 10 rats
(He <i>et al.</i> , 2009) A 90-day toxicology study of transgenic lysine-rich maize grain(Y642) in sprague Dawley rats, Food and Chem. Toxicol. 47 425-432.	Maïs riche en lysine (Y642)	sb401 (gène de p. de terre)		10 rats/groupe/sexe Hébergement individuel rats Sprague Dawley Peking university	5 groupes 30 ou 76% d'incorporation : OGM Y642 30%; OGM Y642 76%; contrôle quasi isogénique 30%; contrôle quasi isogénique 6%; 43,3% farine maïs (régime témoin)	Contrôle négatif supplémentaire	Consommation et croissance pondérale, hématologie, chimie sanguine, poids des organes et pathologie	13 semaines 10 rats
(Healy <i>et al.</i> , 2008) Results of a 13-week safety assurance study with rats fed grain from corn rootworm-protected, glyphosate-tolerant MON87017 corn. Food and chem. Toxicol. 46 2517-2524.	mais MON87017 tolérant à des lépidoptères et au glyphosate	CryBb1 CP4EPSPS	Grain	20 rats/groupe/sexe Hébergement individuel rats Crl:CD (SD)IGS BR Charles River	9 groupes 11 ou 33% d'incorporation : OGM MON87017 33%; OGM MON87017 11%; contrôle quasi isogénique 33%; 6 variétés commerciales 33%	Covance Laboratories BPL	Consommation et croissance pondérale, Observations cliniques, pathologie clinique, hématologie et coagulation, chimie sanguine, analyse d'urine, histopathologie, poids des organes	13 semaines 10 rats: 20 rats pour croissance pondérale et consommation.
(MacKenzie <i>et al.</i> , 2007) Thirteen week feeding study with transgenic maize grain containing event DAS 01507-1 in sprague Dawley rats. Food and Chem. Tox. 45 551-562.	Maïs 1507 tolérant à des lépidoptères et glufosinate d'ammonium	Cry1FPAT	Grain moulu	12 rats/groupe/sexe Hébergement individuel rats Crl:CD (SD)IGS BRCharles River	5 groupes 11 ou 33% d'incorporation : OGM 1507 33%; OGM1507 11%; variété commerciale 33P66 33%; variété commerciale 33J56 33%; 33P66 11%= 33J56 22%	Dupont Haskell Global Centers for Health and environmental Sciences; BPL	Observations cliniques, ophtalmologie, évaluation comportement, pathologie clinique, hématologie et coagulation, chimie sanguine, analyse d'urine, croissance et pathologie anatomique	13 semaines 12 rats
(Malley <i>et al.</i> , 2007) Subchronic feeding study of DAS-59122-7 maize grain in Spague-Dawley rats, Food and Chem. Tox. 45 1277-1292.	mais DAS-59122 tolérant à des lépidoptères et au glufosinate d'ammonium	Cry34Ab1 Cry35Ab1 PAT	Grain	12 rats/groupe/sexe Hébergement individuel rats Crl:CD (SD)IGS BR Charles River	5 groupes 35% d'incorporation : OGM DAS-59122 35%; contrôle quasi isogénique 35%; 1 variété commerciale 35%; 2 régimes commerciaux contenant 35% de grain de maïs		Consommation et croissance pondérale, Observations cliniques, ophtalmologie, comportement, pathologie clinique, hématologie et coagulation, chimie sanguine, analyse d'urine, histopathologie, poids des organes	13 semaines 12 rats

Références Articles	Espèce événement caractère	Protéines exprimées	Mat. végétal	Condition Rats	Groupes Traitements	centre investigateur	Paramètres toxicologiques observés	périodes et effectifs
(Poulsen <i>et al.</i> , 2007a) A 90-day safety study in Wistar rats fed genetically modified rice expressing snowdrop lectin <i>Galanthus nivalis</i> (GNA) Food and Chem. Tox. 45 350-363.	Riz GNA	Lectine de <i>Galanthus nivalis</i>	Farine de riz	16 rats/sexe/groupe rats Wistar M&B breeding center, Danemark 2 rats par cage	2 groupes : 60% riz GNA (OGM); 60% riz contrôle	Danemark	Consommation et Croissance pondérale, Biochimie sanguine, hématologie, immunologie, infectiologie, poids des organes et histopathologie	13 semaines 16 rats
(Poulsen <i>et al.</i> , 2007b) Safety testing of GM rice expressing PHA-E lectin using a new animal test design Food and chem. Tox. 45 364-377.	Riz PHA-E	Forme agglutinine E de la lectine de <i>Phaseolus vulgaris</i> (PHA-E)	Farine de riz	16 rats/sexe/groupe rats Wistar M&B breeding center, Danemark 2 rats par cage	3 groupes : 60% riz PHA-E (OGM); 60% riz PHA-E + lectine 0,1%; 60% riz contrôle;	Danemark	Consommation et Croissance pondérale, Biochimie sanguine, hématologie, immunologie, infectiologie, poids des organes et histopathologie	13 semaines 16 rats
(Schroder <i>et al.</i> , 2007) A 90-day safety study of genetically modified rice expression Cry1Ab protein (<i>Bacillus thuringiensis</i> toxin) in Wistar rats. Food and Chem. Tox. 45 339-349.	Riz KMD1	Cry1Ab	Farine de riz	16 rats/sexe/groupe rats Wistar M&B breeding center, Danemark 2 rats par cage	2 groupes : 60% riz KMD1 (OGM); 60% riz contrôle	Danemark	Consommation et Croissance pondérale, Biochimie sanguine, hématologie, immunologie, infectiologie, poids des organes et histopathologie	
(Wang <i>et al.</i> , 2002) Toxicological evaluation of transgenic rice flour with a synthetic <i>cry1Ab</i> gene from <i>Bacillus thuringiensis</i> . J. Sci. Food Agric. 82 738-744.	Riz KMD1	Cry1Ab	Farine de riz	10 rats/sexe/groupe 2 animaux/cage Rats Sprague Dawley	4 groupes : 64% riz KMD1 (OGM); 32% riz KMD1 (OGM); 16% riz KMD1 (OGM); 64% riz contrôle	Chine	Consommation et croissance pondérale, hématologie, chimie sanguine, poids des organes et pathologie	13 semaines 10 rats

**ANNEXE 2 : Tableau de synthèse de la revue bibliographique
(analyse statistique)**

Tableau de synthèse des méthodologies statistiques des études de toxicité sub-chronique de 90 jours réalisées chez les rats à partir d'aliments issus ou consistant à des OGM publiées dans la littérature.

Référence Article	Espèce	Groupes	valeurs atypiques	Test 1 : Normalité données	Test 2 : Homo scedasticité	croissance poids+Conso	Test de différence (hors données de croissance pondérale)	Multiplicité des tests	Données Historiques, VC
(Appenzeller <i>et al.</i> , 2008) Subchronic feeding study of herbicide-tolerant soybean DP-356043-5 in Sprague-Dawley rats.	Soja 356043 tolérant à des herbicides.	6 groupes (20% de tourteau et 1,5% de coques): OGM non traité, OGM traité glyphosate chlorimuron thifensulfuron, ISO ¹⁷ , 3 VC ¹⁸ .	Obs. clin ¹⁹	Selon Shapiro and Wilk, 1965	Selon Levene, 1960	Analyse visuelle des courbes + ANOVA valeurs finales	(1) <u>variables continues</u> : Si Test 1+2 non S+ ²⁰ : ajustement p-Value ²¹ . Si S+ : contraste linéaire. (2) <u>variables qualitatives</u> , ANOVA avec ajustement . Si ANOVA S+, test non paramétrique (test de rang) (Dunn, 1964) . (3) <u>données d'incidence</u> (quantitatif) analysées avec test de Fisher avec ajustement . Pour les comparaisons S+, comparaison (non statistique) à des données historiques.	(1) Dunn (Dunn, 1964). (2) Miliken and Johnson, 1984 (3) Bonferroni-Holm (Holm 1979).	Données historiques (2000 et 2006) 3 groupes de VC utilisés pour étudier la variabilité naturelle
(Appenzeller <i>et al.</i> , 2009a) Subchronic feeding study with GM stacked trait lepidopteran and coleopteran resistant (DAS-1507-1xDAS-59122-7) maize grain.	Mais 1507x59122 tolérant à des lépidoptères et herbicide.	6 groupes 34% d'incorporation : OGM 1507x59122; ISO; 3 VC.	Obs. Clin	Selon Shapiro and Wilk, 1965	Selon Levene, 1960 Contraste (inter et intra groupe)	Analyse visuelle des courbes + ANOVA valeurs finales	Si Test 1 et Test 2 non S+, ANOVA+ ajustement. Si S+, contraste linéaire avec p ajusté (TYPE 1 (Dunn, 1964). Comparaison Groupe Test/Iso, puis Test/ autres groupes. Pour mesures des pathologies cliniques, ANOVA ou Test de Dunn.	Test de Dunn +Snedecor and Cochran	Données historiques cités
(Appenzeller <i>et al.</i> , 2009b) Subchronic feeding study of grain from herbicide tolerant maize DP-98140-6 in Sprague Dawley rats.	Mais 98140 tolérance glufosinate d'ammonium	6 groupes 35-38% d'incorporation: OGM 98140 non traité; OGM 98140 traité herbicide; ISO; 3 VC.	Obs. Clin	Selon Shapiro and Wilk, 1965	Selon Levene, 1960	Analyse visuelle des courbes + ANOVA valeurs finales	(1) <u>variables continues</u> Si Test 1 et Test 2 non S+, ANOVA+ ajustement. Si S+, contraste linéaire avec p ajusté. Comparaison Groupe Test/Iso, puis Test/ autres groupes.. Pour mesures des pathologies cliniques, ANOVA ou Test de Dunn. (2) <u>Pour les variables qualitatives</u> , ANOVA avec ajustement. Si S+, test non paramétrique (test de rang) (Dunn, 1964). ou test de Fisher avec ajustement.	(1) Snedecor and Cochran, 1967. Dempster et al, 1984 Miliken and Johnson, 1984 Hochking, 1985 (2) Bonferroni-Holm	Données historiques (2000 et 2006) voir Appenzeller ,2008

¹⁷ ISO : Contrôle quasi-isogénique

¹⁸ VC : groupe d'animaux nourris à base de Variétés Commerciales de la même espèce végétale.

¹⁹ Obs. Clin : Observations cliniques des rats présentant des valeurs atypiques pour certains paramètres. Exclusion de l'étude si examen clinique non-satisfaisant.

²⁰ S+ : Test statistiquement significatif (par rapport à α , risque de première espèce). Non S+ : Non significatif.

²¹ Ajust. : Ajustement du P-value (compte tenu de la multiplicité des tests)

Référence Article	Espèce	Groupes	valeurs atypiques	Test 1 : Normalité données	Test 2 : Homoscedasticité testée	croissance poids+Conso	Test de différence (hors données de croissance pondérale)	multiplicité des tests	données Historiques, VC
(Dryzga <i>et al.</i> , 2007) Evaluation of the safety and nutritional equivalence of a GM cottonseed meal in a 90 day dietary toxicity study in rats.	Cotonnier tolérance à des lépidoptères et glufosinate d'ammonium	5 groupes 10% d'incorporation OGM Widestrike; ISO; 3 VC	Obs. Clin +test de Grubbs 1969	NP ²²	Test de Bartlett ($\alpha=0.01$)	Analyse visuelle des courbes + ANOVA valeurs finales	ANOVA entre le groupe test et les 4 contrôles (Steel and Torrie, 1960). <u>Variables qualitatives et d'évaluation sensitive</u> , test z de proportion comparant le coton WIDESTRIKE à la moyenne des proportions des 4 groupes contrôles <u>Variables discrètes</u> , ANOVA avec contraste selon le régime et le sexe Pour pouvoir retenir le sexe comme facteur, il faut interaction sexe*groupe non S+. Si S+, analyse effectuée séparément selon le sexe.	Méthode de Benjamini et Hochberg après ANOVA pour contrôle du FDR. ²³	VC utilisées dans l'analyse statistique
(Hammond <i>et al.</i> , 2004) Results of a 13 week safety assurance study with rats fed grain from glyphosate tolerant corn.	Maïs NK603 tolérant au glyphosate	10 groupes : OGM 33%; OGM 11%; ISO 11%; ISO 33%; 6 VC 33%	NP	NP	Selon Levene, 1960 ($\alpha=0.01$)	Comparaison Hebdomadaire entre les 10 groupes.	<u>Variables quantitatives</u> : comparaison Test 33% à Iso 33% et Test 11% à Iso 11%, pour chaque sexe, ANOVA avec contraste/ régime. comparaison Test 33% à moyenne des VC avec T test (1degré de liberté)	NP	VC servent de contrôle de référence
(Hammond <i>et al.</i> , 2008) Safety assesment of SDA soybean oil: Results of a 28-day study and a 90-day/one generation reproduction feeding study in rats.	Soja riche en SDA (acide gras omega-3)	4 groupes 1,5 g ou 4g/kg/jour huile soja GM SDA 1,5g; huile soja GM SDA 4g; huile de soja ISO 4g; huile de poisson (Menhaden).	NP	NP	NP	ANOVA si S+ alors test de Dunnett	Test de comparaisons paramétrique entre contrôle et le groupe SDA aux deux doses : ANOVA (Snedecor and Cochran). Si différence significative alors Test de Dunnett. Si significativité entre dose forte du groupe SDA et contrôle pour les paramètres sanguins alors comparaison au groupe de référence (huile de poisson).	NP	1 groupe de VC sert de contrôle de référence
(He <i>et al.</i> , 2008) Comparison of grain from corn rootworm resistant T DAS -59122-7 maize with non-T maize grain in a 90 day feeding study in sprague dawley rats.	maïs DAS-59122 tolérant à des lépidoptères et au glufosinate d'ammonium	5 groupes : OGM 50%; OGM 70%; ISO 50%; ISO 70%; VC 43,3% farine maïs.	Obs. Clin		Selon Levene, 1960 ($\alpha=0.01$)	NP	ANOVA 1) comparaison OGM / VC. Puis 2) comparaison OGM / ISO pour chaque dose.	NP	Une VC sert de contrôle de référence
(He <i>et al.</i> , 2009) A 90-day toxicology study of transgenic lysine-rich maize grain(Y642) in sprague Dawley rats,	Maïs riche en lysine (Y642)	5 groupes : OGM Y642 30%; OGM Y642 76%; ISO 30%; ISO 76%; VC 43,3% farine maïs	NP	NP	Selon Levene, 1960 ($\alpha=0.01$)	NP	ANOVA 1) comparaison OGM / VC. Puis 2) comparaison OGM / ISO pour chaque dose..	NP	Une VC sert de contrôle de référence

²² NP : Non précisé²³ FDR : False Discovery rate (taux de faux positifs)

Référence Article	Espèce	Groupes	valeurs atypiques	Test 1 : Normalité données	Test 2 : Homoscedasticité testée	croissance poids+Conso	Test de différence (hors données de croissance pondérale)	multiplicité des tests	données Historiques, VC
(He <i>et al.</i> , 2009) A 90-day toxicology study of transgenic lysine-rich maize grain(Y642) in sprague Dawley rats,	Maïs riche en lysine (Y642)	5 groupes 30 ou 76% d'incorporation : OGM Y642 30%; OGM Y642 76%; contrôle quasi isogénique 30%; contrôle quasi isogénique 6%; 43,3% farine maïs (régime témoin)	NP	NP	Selon Levene,1960 ($\alpha=0.01$)	NP	ANOVA entre les 5 groupes. Si différence S+, comparaisons entre doses.	NP	Une VC sert de contrôle de référence
(Healy <i>et al.</i> , 2008) Results of a 13-week safety assurance study with rats fed grain from corn rootworm-protected, glyphosate-tolerant MON87017 corn.	Mais MON87017 tolérant à des lépidoptères et au glyphosate	9 groupes 11 ou 33% d'incorporation : OGM MON87017 33%; OGM MON87017 11%; contrôle quasi isogénique 33%; 6 VC 33%	Obs. Clin	NP	NP	Analyse visuelle des courbes Comparaison Hebdomadaire entre les 10 groupes.	ANOVA entre les groupes. Si résultats significatif : test de Dunnett entre le groupe OGM et le groupe contrôle.	NP	VC servent de contrôle de référence
(MacKenzie <i>et al.</i> , 2007) Thirteen week feeding study with transgenic maize grain containing event DAS 01507-1 in sprague Dawley rats.	Maïs 1507 tolérant à des lépidoptères et glufosinate d'ammonium	5 groupes : OGM 1507 33%; OGM1507 11%; ISO 33%; ISO 11% VC 33%;	Obs. Clin	Selon Shapiro and Wilk, 1965	Selon Levene, 1960	Analyse visuelle des courbes + ANOVA valeurs finales	Si Test 1 et Test 2 non S+, ANOVA. Si S+, contraste linéaire avec p ajusté selon TYPE 1 (Dunn, 1964). Comparaison Groupe OGM/ISO, puis OGM aux autres groupes.	Test de Dunn	Une VC sert de contrôle de référence
(Malley <i>et al.</i> , 2007) Subchronic feeding study of DAS-59122-7 maize grain in Spague-Dawley rats,	mais DAS-59122 tolérant à des lépidoptères et au glufosinate d'ammonium	5 groupes: OGM 35%; ISO 35%; 1 VC contrôle 35%; 2 VC	NP	Selon Shapiro and Wilk, 1965	Selon Levene, 1960		Si test 1 et test 2 non S+ : ANOVA avec p ajusté selon Snedecor and Cochran 1967, Dempster et al 1984. Si test 1 ou test 2 S+ : Contraste linéaire de Dunn (Dunn, 1964) Contraste supplémentaire pour chaque catégorie de variable : test de Fisher. Comparaison OGM à ISO et OGM à VC contrôle	Test de Bonferroni_Holm	Données historiques pour comparaison supplémentaire. 2 VC servent de contrôle de référence

Référence Article	Espèce	Groupes	valeurs atypiques	Test 1 : Normalité données	Test 2 : Homoscedasticité testée	croissance poids+Conso	Test de différence (hors données de croissance pondérale)	multiplicité des tests	données Historiques, VC
(Poulsen <i>et al.</i> , 2007a) A 90-day safety study in Wistar rats fed genetically modified rice expressing snowdrop lectin Galanthus nivalis (GNA)	Riz GNA	2 groupes : OGM 60% contrôle 60%	Obs. Clin	Testé, transformation si nécessaire (méthode NP)	observation des résiduelles, modèle linéaire	NP.	ANOVA. Si S+, T test. Si test 1 S- mais test 2 S+ : test non paramétrique (test de Kruskall-Wallis suivi d'un test de Wilcoxon si Kruskall-Wallis S+).	NP	NP
(Poulsen <i>et al.</i> , 2007b) Safety testing of GM rice expressing PHA-E lectin using a new animal test design	Riz PHA-E	3 groupes : OGM 60% OGM 60% + lectine 0,1%; Contrôle 60%;	Obs. Clin	Testé, transformation si nécessaire (méthode NP)	observation des résiduelles, modèle linéaire	Analyse visuelle des courbes Comparaison Hebdomadaire entre les 3 groupes.	ANOVA. Si S+, T test. Si test 1 S- mais test 2 S+ : test non paramétrique (test de Kruskall-Wallis suivi d'un test de Wilcoxon si Kruskall-Wallis S+).	NP	NP
(Schroder <i>et al.</i> , 2007) A 90-day safety study of genetically modified rice expression Cry1Ab protein (Bacillus thuringiensis toxin) in Wistar rats.	Riz KMD1	2 groupes : OGM 60% Contrôle 60%;	Obs. Clin	Testé (méthode NP)	Testé (NP)	ANOVA avec mesures répétés pour un facteur. Si S+, LS means test ou Duncan test	ANOVA, si F test S+ : test LS means pour comparer les deux groupes ISO et OGM. Si test 1 S+, test non paramétrique (Test de Kruskal Wallis suivi d'un test de Wilcoxon)	NP	NP
(Wang <i>et al.</i> , 2002) Toxicological evaluation of transgenic rice flour with a synthetic cry1Ab gene from bacillus thuringiensis.	Riz KMD1	4 groupes : OGM 64% OGM 32% OGM 16% Contrôle 64%	Obs. Clin	NP	NP	NP	LS Means test	NP	NP

ANNEXE 3 : Les mesures d'influence

L'influence des observations dans un modèle tel que le modèle linéaire peut s'exercer sur la partie fixe du modèle $X\beta$ mais également sur la partie aléatoire $Z\gamma$ générant des biais sur ces deux composantes du modèle.

Un nombre important d'indicateurs sont proposés dans la littérature. Les plus importants sont présentés ci-dessous.

- **La statistique PRESS**

Cet indicateur global mesure l'écart entre les valeurs observées y_i et leurs prédictions dans un modèle où l'observation i a été écartée $\hat{y}_{(-i)}$:

$$\text{PRESS} = \sum_i (y_i - \hat{y}_{(-i)})^2$$

- **Le coefficient D de Cook**

Cet indicateur mesure l'influence de l'observation i sur l'estimation des effets fixes β

$$\text{Cook } D_i = \frac{r_i^2}{\text{rang}(X)} \frac{h_{ii}}{(1-h_{ii})}$$

Où r_i^2 représente le résidu studentisé $r_i^2 = \frac{\hat{\varepsilon}_i}{\sqrt{\hat{\sigma}^2(1-h_{ii})}}$ et h_{ii} représente le terme diagonal de la matrice

$$\text{de projection } H = X(X'X)^{-1}X'$$

Une grande hétérogénéité entre les valeurs de cet indice peut traduire une influence importante des observations ayant les valeurs sur cet indice les plus élevées. Des valeurs supérieures à 1 traduisent une importance excessive.

Cet indice peut également être utilisé pour mesurer l'influence des observations sur les estimations des paramètres aléatoires γ .

- **Le coefficient covratio**

Cet indice mesure l'influence de l'observation i sur la matrice de variance-covariance des estimations des effets fixes β

$$\text{covratio}(\beta) = \frac{|\hat{\text{var}}(\hat{\beta}_{(-i)})|}{|\hat{\text{var}}(\hat{\beta})|}$$

Où $|\hat{\text{var}}(\hat{\beta}_{(-i)})|$ désigne le déterminant de la matrice de variance-covariance des estimations de effets fixes en l'absence de l'observation i .

Une valeur inférieure à 1 indique une influence prépondérante de l'observation i sur les variances-covariances des estimations des effets fixes.

Cet indice peut également être utilisé pour mesurer l'influence des observations sur les variances des estimations des paramètres aléatoires γ .

- **L'indice de distance de vraisemblance restreinte ("Restricted Likelihood Distance")**

Cet indice égal à deux fois l'écart entre les log vraisemblances avec et sans l'observation i est une mesure globale de l'influence de l'observation i (Cook and Weisberg 1982)²⁴. Des valeurs élevées pour une observation i traduit une contribution excessive à la vraisemblance apportée par cette observation.

²⁴ Cook, R. D. and Weisberg, S. (1982), *Residuals and Influence in Regression*, New York: Chapman and Hall.

ANNEXE 4 : Le False Discovery Rate (FDR)

Définitions

Le contrôle du risque d'erreur de première espèce, lorsque l'on réalise simultanément un très grand nombre de tests statistiques sur le même ensemble de données, est un problème majeur (Hochberg and Tamhane, 1987). D'importants développements théoriques ont été réalisés depuis une vingtaine d'années. De nouveaux critères ont été définis et de nouvelles procédures ont été proposées. Cette note présente une synthèse des développements autour du False Discovery Rate (FDR) proposé par Y. Benjamini et Y. Hochberg en 1995 (Benjamini and Hochberg, 1995). Elle ne se veut pas exhaustive mais décrit succinctement les principales procédures utilisées.

Le contexte

Si m hypothèses sont testées simultanément sur le même ensemble de données. Le tableau ci-dessous détaille la classification des hypothèses nulles selon qu'elles sont vraies ou fausses (respectivement m_0 , m_1) et selon qu'elles sont acceptées ou rejetées par les tests (respectivement W , R). Parmi les R hypothèses nulles rejetées, V sont rejetées à tort (« faux positifs »). Parmi les W hypothèses nulles non rejetées, T le sont à tort (« faux négatifs »).

		Hypothèses nulles		Total
		Acceptées	rejetées	
Hypothèses nulles	vraies	U	V	m_0
	fausses	T	S	m_1
Total		W	R	m

Trois critères sont habituellement définis qui permettent d'estimer différentes erreurs commises sur les m tests.

Le « *per Comparison Error Rate* » : $PCER = \frac{E[V]}{m}$

Le « *Family Wise Error Rate* » : $FWER = \Pr[V \geq 1]$

Le « *False Discovery Rate* » : $FDR = E\left[\frac{V}{R} | R > 0\right] \cdot \Pr[R > 0]$

Le PCER est le taux d'erreur que l'on assume habituellement sur l'ensemble des m tests lorsque chacun est réalisé au même niveau de risque de 1^{ère} espèce α . Si toutes les hypothèses nulles sont vraies ($m_0 = m$, « hypothèse nulle complète ») alors $PCER = \alpha$. On obtient alors en moyenne $\alpha\%$ de faux positifs.

Le FWER est le risque de rejeter au moins une hypothèse nulle à tort sur l'ensemble des m tests réalisés. On montre que FWER est plus élevé que PCER et ce d'autant plus que le nombre de tests est élevé. On a l'inégalité suivante entre les taux d'erreurs :

$$FWER \leq 1 - (1 - PCER)^m$$

Le FDR est l'espérance du taux de faux positifs. Sous l'hypothèse nulle complète le FDR est égal au FWER. Si au moins une hypothèse nulle est fautive (« hypothèse nulle partielle ») le FDR est plus faible que le FWER.

Le FDR et le FWER ne contrôlent pas les mêmes risques. Choisir le FDR suppose implicitement que l'on préfère contrôler la proportion de faux positifs plutôt que la probabilité d'avoir au moins un faux positif.

Le FDR offre une solution acceptable au contrôle du risque de faux positifs lorsque le nombre m de tests réalisés est très élevé. Dans ce cas le FWER tend vers 1 et son contrôle se traduit par une diminution dramatique de la puissance des tests (accroissement des faux négatifs).

La procédure de Benjamini Hochberg

Benjamini et Hochberg ont proposé dans leur article de 1995 une procédure séquentielle ascendante qui

contrôle le FDR à un niveau au plus égal à $\frac{m_0}{m} \alpha^{25}$ où α est le niveau de risque au sens du FDR, assumé *a priori*. Les m tests statistiques doivent être supposés indépendants.

La procédure, notée « BH » par la suite, est la suivante :

²⁵Le FDR est rigoureusement égal au seuil $\frac{m_0}{m} \alpha$ dans le cas de statistiques de tests continus.

Soit $p(1) \leq p(2) \leq \dots \leq p(k) \leq \dots \leq p(m-1) \leq p(m)$ les degrés de signification (« pvalues ») des m hypothèses testées classées par valeurs croissantes.

$$\hat{k} = \max \left\{ k \mid p(k) \leq \frac{k}{m} \alpha \right\}$$

On détermine

On rejette les hypothèses nulles $H_{0(1)}, H_{0(2)}, \dots, H_{0(\hat{k})}$

A chaque test est associé un degré de signification ajusté :

$$p^*(m) = p(m)$$

$$p^*(k) = \min \left\{ p(k+1) ; \frac{m}{k} p(k) \right\} \quad k = m-1, \dots, 2, 1$$

Ces degrés de signification permettent leur comparaison directe au niveau α du risque consenti.

La procédure « adaptative »

En partant du constat que la procédure BH contrôle le FDR à un niveau trop faible par rapport au risque consenti α , Benjamini et Hochberg, 2000 ont proposé une approche en deux étapes basée sur une estimation \hat{m}_0 du nombre m_0 d'hypothèses nulles vraies obtenu à partir d'une méthode graphique (Schweder et Spjøtvoll, 1982).

Il suffit ensuite d'appliquer la procédure de Benjamini et Hochberg au seuil $\frac{m}{\hat{m}_0} \alpha$.

Les degrés de signification ajustés s'en déduisent :

$$\tilde{p}(k) = \min \left\{ \tilde{p}(k+1) ; \frac{\hat{m}_0}{k} p(k) \right\} \quad k = m-1, \dots, 2, 1 \text{ et } \tilde{p}(m) = \frac{\hat{m}_0}{m} p(m)$$

Les auteurs montrent que sous l'hypothèse de tests indépendants la procédure contrôle le FDR au niveau α tout en étant plus puissante que la procédure HB.

Tests statistiques dépendants

Benjamini et Yekutieli, 2001 ont montré que la procédure BH contrôle également le FDR à un niveau au plus égal à $\frac{m_0}{m} \alpha$ lorsque les tests sont corrélés positivement.

Ces mêmes auteurs ont proposé une adaptation de la procédure BH en cas de dépendance quelconque

entre les tests en remplaçant le seuil $\frac{k}{m} \alpha$ par : $\frac{k}{m} \left(\sum_{i=1}^m \frac{1}{i} \right)^{-1} \alpha$

Les degrés de signification ajustés s'écrivent :

$$p^{**}(k) = \min \left\{ p^{**}(k+1) ; \frac{m}{k} \left(\sum_{i=1}^m \frac{1}{i} \right) p(k) \right\} \quad k = m-1, \dots, 2, 1 \text{ et } p^{**}(m) = \left(\sum_{i=1}^m \frac{1}{i} \right) p(m)$$

Cette procédure (notée « BY ») contrôle également le FDR à un niveau au plus égal à $\frac{m_0}{m} \alpha$

Le Positive False Discovery Rate (PFDR)

J. Storey, (2002a, 2002b) propose d'utiliser l'espérance du taux de faux positifs conditionnellement au rejet d'au moins une hypothèse nulle parmi les m hypothèses testées :

$$\text{Le « positive False Discovery Rate » : } p\text{FDR} = E \left[\frac{V}{R} \mid R > 0 \right]$$

La valeur de ce critère est supérieure à celle du FDR mais les deux critères sont asymptotiquement équivalents lorsque le nombre m de tests tend vers l'infini ($\Pr(R > 0)$ tend vers 1). Il présente par contre l'inconvénient de ne pas pouvoir être contrôlé sous l'hypothèse nulle complète (pFDR=1 dans ce cas).

J. Storey démontre que, lorsque que les m tests statistiques sont indépendants et identiquement distribués sous l'hypothèse nulle, le PFDR s'interprète comme la probabilité *a posteriori* d'hypothèse nulle vraie conditionnellement aux résultats des tests statistiques :

$$pFDR = \Pr(H_0 \text{ vraie} | X \in \Gamma) = \pi_0 \frac{\Pr(X \in \Gamma | H_0 \text{ vraie})}{\Pr(X \in \Gamma)} \quad [1]$$

où X est la statistique de test, Γ désigne la zone de rejet, π_0 est la probabilité que l'hypothèse nulle soit vraie (i.e la proportion d'hypothèses nulles vraies) et $\Pr(X \in \Gamma | H_0 \text{ vraie})$ représente le risque de première espèce du test.

Cette interprétation du pFDR permet à Storey de proposer une méthode d'estimation du pFDR (procédure « ST ») basée sur la distribution marginale des degrés de signification p_k associés aux m statistiques x_k .

$$p\hat{FDR}(\gamma) = \hat{\pi}_0 \frac{\gamma}{\#\{p_k \leq \gamma\}/m} \quad [2]$$

où γ est le risque de 1^{ère} espèce assumé au niveau des m tests et $\#\{p_k \leq \gamma\}/m$ est la proportion d'hypothèses nulles rejetées à ce niveau de risque.

Cette estimation nécessite d'estimer la proportion d'hypothèses nulles vraies π_0 que Storey obtient en exprimant l'estimation de π_0 comme une fonction d'un paramètre de filtrage λ :

$$\hat{\pi}_0(\lambda) = \frac{\#\{p_k > \lambda | k = 1, \dots, m\}}{m(1-\lambda)} \quad 0 \leq \lambda \leq 1 \quad [3]$$

où $\#\{p_k > \lambda | k = 1, \dots, m\}$ représente le nombre de degrés de signification supérieurs à λ .

$\hat{\pi}_0$ est alors estimé en posant un modèle de régression spline cubique sur $\hat{\pi}_0(\lambda)$ en fonction de λ et en posant $\hat{\pi}_0 = \hat{\pi}_0(0,9)$.

L'estimation obtenue est biaisée supérieurement et entraîne en conséquence un biais positif sur le pFDR et donc au final une règle de décision conservative au niveau des tests (trop d'hypothèses nulles conservées pour un niveau de pFDR fixé a priori).

D'autres méthodes sont possibles pour estimer $\hat{\pi}_0$.

C. Dalmaso *et al.*, 2005 proposent une classe d'estimateurs de π_0 basée sur l'espérance mathématique d'une transformée du degré de signification p : $\varphi(p) = [-\ln(1-p)]^\nu$ ($\nu=1$ ou 2). L'estimateur obtenu LBE (« Local Based Estimator ») est comparable à l'estimateur ST avec un biais légèrement supérieur, une variance moindre et une erreur quadratique moyenne souvent inférieure.

Une méthode alternative (Allison *et al.*, 2002) consiste à modéliser la distribution des degrés de signification à l'aide d'un mélange d'une loi uniforme \mathbf{U} pondérée par π_0 pour les hypothèses nulles vraies et d'une distribution de densité de probabilité appropriée \mathbf{G} pondérée par $(1-\pi_0)$ pour les hypothèses nulles fausses. La Loi Beta semble notamment être un choix de fonction \mathbf{G} particulièrement adapté (J. A. Osborne, 2006).

Les qvalues²⁶

J. Storey, 2002a définit la notion de qvalue comme une mesure analogue au degré de signification ajusté appliquée au pFDR (FDR²⁷). Étant donné un test donné X_k de degré de signification p_k la qvalue est la plus petite valeur du FDR pour laquelle le test peut être déclaré significatif à un niveau de risque global α .

$$qvalue(p_k) = \min_{p_k \leq \alpha} \{ \hat{FDR}(\alpha) \}$$

La procédure suivante permet de calculer la qvalue associée à chaque test.

Soit $p(1) \leq p(2) \leq \dots \leq p(k) \leq \dots \leq p(m-1) \leq p(m)$ les m degrés de signification des m hypothèses testées classées par valeurs croissantes.

Les « qvalues » associées aux m pvalues sont définies de la façon suivante :

$$qvalue(p(m)) = \text{MIN}\{1; \hat{\pi}_0 \cdot p(m)\}$$

$$k = m-1, m-2, \dots, 2, 1 : qvalue(p(k)) = \text{MIN}\left\{qvalue(p(k+1)); \hat{\pi}_0 \cdot \frac{p(k)}{k/m}\right\}$$

²⁶ Le terme anglais n'ayant pas de traduction française pour l'instant il est conservé dans le présent document.

²⁷ On considère le nombre de tests suffisamment grand pour qu'il y ait au moins une hypothèse nulle rejetée. Le pFDR est alors équivalent au FDR.

où $\hat{\pi}_0$ est l'estimation de la proportion d'hypothèses nulles vraies parmi les m hypothèses testées.

La fixation a priori d'une valeur seuil α au FDR permet d'identifier les tests statistiques significatifs assurant un taux de faux positifs au maximum égal à α . Ce sont les tests vérifiant : $qvalue(p(k)) \leq \alpha$.

Conclusion

En règle générale les procédures sont classées de la façon suivante sur le plan de leur capacité à détecter des faux positifs :

ST \geq adaptative \geq HB \geq BY

En particulier on déduit de [2] que la procédure ST avec un niveau de contrôle du FDR égal à α est équivalente (c'est-à-dire rejette les mêmes hypothèses nulles) à la procédure BH avec un niveau de contrôle du FDR égal à α/π_0 . Ce gain de puissance est dû au fait que la procédure ST utilise l'information apportée par la distribution des degrés de signification pour estimer π_0 ce que ne fait pas la procédure BH.

Lorsque le nombre de tests réalisés est faible les estimations des probabilités π_0 ont des dispersions importantes. Il est alors préférable d'utiliser les procédures BH ou BY pour lesquelles π_0 n'est pas estimé.

Lorsque les tests sont dépendants la procédure adaptée est BY. Cette procédure s'avère toutefois être très conservatrice et peu utilisée en pratique.

La procédure BH peut être utilisée lorsque les statistiques de tests sont dépendantes au sein de petits groupes indépendants les uns des autres (Storey, 2002a).

Elle peut également être utilisée dans le cas de tests bilatéraux lorsque l'on peut poser une structure de corrélation commune ρ sur les m statistiques de tests, la procédure BH contrôlant alors le FDR (Anat Reiner-Benaim, 2007). Le même auteur conjecture que le contrôle du FDR devrait être assuré lorsque l'on remplace les corrélations ρ_{ij} entre couples de statistiques de tests (X_i, X_j) par une corrélation moyenne $\bar{\rho}$.

L'utilisation des q values

Storey (Storey, 2002a, 2003) propose de fixer a priori le risque de première espèce γ au niveau des tests individuels et d'en déduire le FDR minimum pour ce niveau de risque (cf. formule [2]). Ce FDR fonction de γ , son intervalle de confiance (estimé par des méthodes de Bootstrap (Westfall and Young, 1993) ainsi que l'estimation de l'espérance de faux positifs sont les paramètres d'intérêt.

Cette démarche va à l'encontre de l'attitude classique qui consiste à se fixer a priori la valeur du FDR à un niveau α et d'en déduire la zone de rejet pour chacune des statistiques de tests.

La démarche de Storey s'inscrit dans le cadre d'une démarche « exploratoire » où les qvalues jouent le rôle d'indicateurs permettant de sélectionner les tests significatifs à un niveau de risque contrôlé au sens du FDR contrairement aux degrés de signification.

A contrario l'attitude classique de détermination des zones de rejet des tests pour un niveau fixé a priori du FDR s'inscrit dans une démarche « confirmatoire » analogue à celle adoptée lorsque l'on recherche les tests significatifs avec un contrôle a priori du FW.

Bibliographie

Allison, D., Gadbury, G., Heo, M., Fernandez, J., Lee, C.-K., Prolla, T. and Weindruch, R. (2002). A mixture model approach for the analysis of microarray gene expression data. *Computational Statistics & Data Analysis* **39**, 1-20.

Anat Reiner-Benaim (2007). FDR control by the BH procedure for two-sided correlated tests with implications to gene expression data analysis. *Biometrical Journal* **49** 1, 107-126.

Benjamini, Y., Hochberg, Y., (2000) "On the « adaptative » control of the false discovery rate", *J. Edu. Behav. Statist.*, **25**, 60-83.

Benjamini, Y., Yekutieli, D., (2001) "The control of the false discovery rate in multiple testing under dependency", *Ann. Statist.* **29**, 1165-1188.

Dalmasso C., Broët P, Moreau T. (2005) "A simple procedure for estimating the false discovery rate". *Bioinformatics*, **25**, 660-668.

Osborne, J. A. (2006) "Estimating the False Discovery Rate using SAS®". SUGI 31

Schweder, T., Spjøtvoll, E. (1982) "Plots of p-values to evaluate many tests simultaneously", *Biometrika*, **69**, 3, pp. 493-502.

ANNEXE 5 : Paramètres étudiés dans l'étude MON810

N° de variable	Type de paramètres	abréviations	Paramètres	Unité	Période	Propriété
Var1	urinaire	BILI	Bilirubine		P1+P2	Qual
Var2	urinaire	GRAVITY	Gravité urine		P1+P2	Quant
Var3	urinaire	PROT	Protéine dans Urine	mg/dL	P1+P2	Qual
Var4	urinaire	GLUCOSE	Glucose urine	mg/dL	P1+P2	Qual
Var5	urinaire	BLOOD	Sang dans urines		P1+P2	Qual
Var6	urinaire	PH	pH		P1+P2	Quant
Var7	urinaire	KETONE	Cétones	mg/dL	P1+P2	Qual
Var8	urinaire	UROBILI	Urobilinogène	EU/dL	P1+P2	Qual
Var9	Chimie Sang	Nasg	Sodium Sanguin	mmol/L	P1+P2	Quant
Var10	Chimie Sang	AGRATIO	Ratio Alb/Glob		P1+P2	Quant
Var11	Chimie Sang	ALBUMIN	Albumine	g/dL	P1+P2	Quant
Var12	Chimie Sang	ALKPHOSPH	Phosphatase Alcaline	U/L	P1+P2	Quant
Var13	Chimie Sang	BUN	Urée Sanguine	mg/dL	P1+P2	Quant
Var14	Chimie Sang	CaSg	Calcium sanguin	mg/dL	P1+P2	Quant
Var15	Chimie Sang	ChlorideSg	Chloride sanguin	mmol/L	P1+P2	Quant
Var16	Chimie Sang	CreatSg	Créatinine sanguine	mg/dL	P1+P2	Quant
Var17	Chimie Sang	DIRBILI	Bilirubine directe	mg/dL	P1+P2	Quant
Var18	Chimie Sang	GAMMAGT	Gamma glutamine transférase	U/L	P1+P2	Qual
Var19	Chimie Sang	GLOBULINE	Globuline	g/dL	P1+P2	Quant
Var20	Chimie Sang	glycemie	Glucose dans sang	mg/dL	P1+P2	Quant
Var21	Chimie Sang	SGOTAST	Aspartate aminotransferase	U/L	P1+P2	Quant
Var22	Chimie Sang	SGPTALT	Alanine Aminotransferase	U/L	P1+P2	Quant
Var23	Chimie Sang	TOTBILI	Bilirubine Totale	mg/dL	P1+P2	Quant
Var24	Chimie Sang	phosphsg	Phosphate sang	mg/dL	P1+P2	Quant
Var25	Chimie Sang	Ksg	Potassium sang	mmol/L	P1+P2	Quant
Var26	Chimie Sang	TOTPROT	Protéines Totales	g/dL	P1+P2	Quant
Var27	Hematologique	APTTSEC	Temps de céphaline activée	secondes	P2	Quant
Var28	Hematologique	PTSEC	Temps de prothrombine	secondes	P2	Quant
Var29	Hematologique	ABSBASO	Basophiles absous	.10 ³ /μL	P1+P2	Quant
Var30	Hematologique	PLT	Taux Plaquettes	.10 ³ /μL	P1+P2	Quant
Var31	Hematologique	ABSEOS	Eosinophiles absolus	.10 ³ /μL	P1+P2	Quant
Var32	Hematologique	ABSLUC	Large Unstained Cells Absolus	.10 ³ /μL	P1+P2	Quant
Var33	Hematologique	ABSLYMPH	Lymphocytes absolus	.10 ³ /μL	P1+P2	Quant
Var34	Hematologique	ABSMONO	Monocytes absolus	.10 ³ /μL	P1+P2	Quant
Var35	Hematologique	ABSNEUT	Neutrophiles absolus	.10 ³ /μL	P1+P2	Quant
Var36	Hematologique	BASO	Basophiles	%	P1+P2	Quant
Var37	Hematologique	EOS	Eosinophiles	%	P1+P2	Quant
Var38	Hematologique	HCT	Hématocrite	%	P1+P2	Quant
Var39	Hematologique	HGB	Hémoglobine	g/dL	P1+P2	Quant
Var40	Hematologique	LUC	Large Unstained Cells	%	P1+P2	Quant
Var41	Hematologique	LYMPH	Lymphocytes	%	P1+P2	Quant
Var42	Hematologique	MCH	mean corpuscular hemoglobin	pg	P1+P2	Quant

N° de variable	Type de paramètres	abréviations	Paramètres	Unité	Période	Propriété
Var43	Hématologique	MCHC	mean corpuscular hemoglobin concentration	g/dL	P1+P2	Quant
Var44	Hématologique	MCV	Volume corpusculaire moyen	fL	P1+P2	Quant
Var45	Hématologique	RBC	Hématies	$\cdot 10^6/\mu\text{L}$	P1+P2	Quant
Var46	Hématologique	MONO	Monocytes	%	P1+P2	Quant
Var47	Hématologique	NEUT	Neutrophiles	%	P1+P2	Quant
Var48	Hématologique	WBC	Leucocytes	$\cdot 10^3/\mu\text{L}$	P1+P2	Quant
Var49	urinaire	CALCIUM	Calcium Urine	mg/dL	P1+P2	Quant
Var50	urinaire	CHLORIDE	Chloride	mmol/L	P1+P2	Quant
Var51	urinaire	PHOSPH	Phosphate urine	mg/dL	P1+P2	Quant
Var52	urinaire	CREAT	Creatinine	mg/dL	P1+P2	Quant
Var53	urinaire	POTASSIUM	Potassium Urine	mmol/L	P1+P2	Quant
Var54	urinaire	CREATCLEARANCE	Clairance de la créatinine		P1+P2	Quant
Var55	urinaire	NAKRATIO	Ratio Sodium/Potassium		P1+P2	Quant
Var56	urinaire	SODIUM	Sodium urine	mmol/L	P1+P2	Quant
Var57	urinaire	UPROT	Protéine dans Urine	mg/dL	P1+P2	Quant
Var58	urinaire	VOLTOTAL	Volume urinaire	mL/16hrs	P1+P2	Quant
Var59	Poids Organes	ADRENAL	Glande surrénale	g	P2	Quant
Var60	Poids Organes	BRAIN	Cerveau	g	P2	Quant
Var61	Poids Organes	HEART	Cœur	g	P2	Quant
Var62	Poids Organes	KIDNEY	Rein	g	P2	Quant
Var63	Poids Organes	LIVER	Foie	g	P2	Quant
Var64	Poids Organes	OVARY	Ovaire	g	P2	Quant
Var65	Poids Organes	SPLEEN	Rate	g	P2	Quant
Var66	Poids Organes	TERMINALBODY WEIGHT	Poids corporel final	g	P2	Quant
Var67	Poids Organes	TESTIS	Testicules	g	P2	Quant

ANNEXE 6 : Programme SAS

```

/*=====
Transformation des données
=====*/

data mon_so.totalid;
set mon_so.total5_sans_atypique;
array Vart[67] vart1-vart67;
array svar[67] svar1-svar67;
array BVAR[67] BVAR1-BVAR67;
array var[67] var1-var67;
do i=1 to 67;
if svar[i] = 1 then Vart[i]=log(var[i]-BVAR[i]);
if svar[i] = -1 then Vart[i]=(1+(var[i]/BVAR[i]))**BVAR[i];
if svar[i] = 0 then Vart[i]=var[i];
end;
run;

/*=====
Test de différences entre les groupes OGM et ISO
=====*/

/*-----
Modèle mixte pour les variables quantitatives sur les deux périodes ;
-----*/
%macro mixedcontraste;
ods html;
data general;
variable="PPPPP"; output;
run;
%do v=9 %to 17;
ods graphics on;
Proc mixed data=mon_so.totalid covtest method=REML;
title1 "Modèle mixte avec analyse de l'influence des observations (rats)";
title2 "modèle 'Diètes-Doses-Périodes' ";
by sexe ;
where group<5;
class id dose periode diet group ;
model vart&v= diet dose diet*dose periode diet*periode dose*periode diet*dose*periode / residual
|
nfluence(iter=5 effect=id(diet*dose) estimates);
random id(diet*dose);
lsmeans diet*dose*periode diet*dose diet*periode diet / alpha=.10 cl;

* Estimation des contrastes entre régimes par dose et par période;
estimate "O|d=11%&p1" diet -1 +1 diet*dose -1 +1 0 0 diet*periode -1 +1 0 0
diet*dose*periode -1 +1 0 0 0 0 0 / alpha=.10 cl;
estimate "O|d=33%&p1" diet -1 +1 diet*dose 0 0 -1 +1 diet*periode -1 +1 0 0
diet*dose*periode 0 0 0 -1 +1 0 0 / alpha=.10 cl;
estimate "O|d=11%&p2" diet -1 +1 diet*dose -1 +1 0 0 diet*periode 0 0 -1 +1
diet*dose*periode 0 0 -1 +1 0 0 0 / alpha=.10 cl;
estimate "O|d=33%&p2" diet -1 +1 diet*dose 0 0 -1 +1 diet*periode 0 0 -1 +1 diet*dose*periode 0 0 0 0 0 -1
+1
/ alpha=.10 cl;

* Comparaisons des régimes par dose, par période et marginalement;
estimate "O|d11%" diet -1 +1 diet*dose -1 +1 0 0 / alpha=.10 cl;
estimate "O|d33%" diet -1 +1 diet*dose 0 0 -1 +1 / alpha=.10 cl;
estimate "O|p1" diet -1 +1 diet*periode -1 +1 0 0 / alpha=.10 cl;
estimate "O|p2" diet -1 +1 diet*periode 0 0 -1 +1 / alpha=.10 cl;
estimate "O|" diet -1 +1 / alpha=.10 cl;

* Estimation des évolutions régimes|dose entre périodes (période 1 - période 2);

```

```

estimate "OP1OP2|d11%" periode +1 -1 diet*periode 0 +1 0 -1 dose*periode +1 -1 0 0
      diet*dose*periode 0 +1 0 -1 0 0 0 0 / alpha=.10 cl;
estimate "IP1IP2|d11%" periode +1 -1 diet*periode +1 0 -1 0 dose*periode +1 -1 0 0
      diet*dose*periode +1 0 -1 0 0 0 0 0 / alpha=.10 cl;
estimate "OP1OP2|d33%" periode +1 -1 diet*periode 0 +1 0 -1 dose*periode 0 0 +1 -1
      diet*dose*periode 0 0 0 0 0 +1 0 -1 / alpha=.10 cl;
estimate "IP1IP2|d33%" periode +1 -1 diet*periode +1 0 -1 0 dose*periode 0 0 +1 -1
      diet*dose*periode 0 0 0 0 +1 0 -1 0 / alpha=.10 cl;

* Comparaisons des évolutions 'OGM-ISO' entre périodes (période 1 - période 2);
estimate "D*P|d11" diet*periode -1 +1 +1 -1 dose*periode*diet -1 +1 +1 -1 0 0 0 0 / alpha=.10 cl;
estimate "D*P|d33" diet*periode -1 +1 +1 -1 dose*periode*diet 0 0 0 0 -1 +1 +1 -1 / alpha=.10 cl;
ods output estimates=Est;
ods output covparms=var;
ods output tests3=tests;

run;
quit;
data num;
do sexe="F","M";
do contraste=1 to 15;
output;
end;
end;
run;
data est2;
merge num est;
by sexe;
run;
data Res;
set var;
where covparm="Residual" ;
VAR_Res=Estimate;
run;
data ID;
set var;
where covparm="id(dose*diet)" ;
VAR_ID=Estimate;
run;
data VAR2;
merge ID RES;
by sexe;
keep sexe VAR_ID VAR_Res Rho;
Rho=VAR_ID/(VAR_ID+VAR_Res);
run;

data num1;
do sexe="F","M";
do T=1 to 7;
output;
end;
end;
run;
data tests2;
merge num1 tests;
by sexe;
run;
proc transpose data=tests2 out=tests2 name=test
prefix=T;
by sexe;
run;
data tests2;
set tests2;
where test="ProbF";
keep sexe T1-T7;
run;
data tout;

```

```

merge est2 var2 tests2;
by sexe;
Variable="Var&v";
run;
data general;
set general tout;
run;
%end;
%mend;
%mixedcontraste;

%macro mixedcontraste2;
ods html;
data general2;
variable="PPPPP"; output;
run;
%do v=19 %to 58;
ods graphics on;
Proc mixed data=mon_so.totalid covtest method=REML;
  title1 "Modèle mixte avec analyse de l'influence des observations (rats)";
  title2 "modèle 'Diètes-Doses-Périodes' ";
  by sexe ;
  where group<5;
  class id dose periode diet group ;
  model vart&v= diet dose diet*dose periode diet*periode dose*periode diet*dose*periode/ residual
          influence(iter=5 effect=id(diet*dose) estimates);

  random id(diet*dose);
  lsmeans diet*dose*periode diet*dose diet*periode diet / alpha=.10 cl;
* estimation des contrastes entre régimes par dose et par période;
  estimate "OI|d=11%&p1" diet -1 +1 diet*dose -1 +1 0 0 diet*periode -1 +1 0 0
          diet*dose*periode -1 +1 0 0 0 0 0 0 / alpha=.10 cl;
  estimate "OI|d=33%&p1" diet -1 +1 diet*dose 0 0 -1 +1 diet*periode -1 +1 0 0
          diet*dose*periode 0 0 0 0 -1 +1 0 0 / alpha=.10 cl;
  estimate "OI|d=11%&p2" diet -1 +1 diet*dose -1 +1 0 0 diet*periode 0 0 -1 +1
          diet*dose*periode 0 0 -1 +1 0 0 0 0 / alpha=.10 cl;
  estimate "OI|d=33%&p2" diet -1 +1 diet*dose 0 0 -1 +1 diet*periode 0 0 -1 +1
          diet*dose*periode 0 0 0 0 0 0 -1 +1 / alpha=.10 cl;

* Comparaisons des régimes par dose, par période et marginalement;
  estimate "OI|d11%" diet -1 +1 diet*dose -1 +1 0 0 / alpha=.10 cl;
  estimate "OI|d33%" diet -1 +1 diet*dose 0 0 -1 +1 / alpha=.10 cl;
  estimate "OI|p1" diet -1 +1 diet*periode -1 +1 0 0 / alpha=.10 cl;
  estimate "OI|p2" diet -1 +1 diet*periode 0 0 -1 +1 / alpha=.10 cl;
  estimate "OI" diet -1 +1 / alpha=.10 cl;

* Estimation des évolutions régimes|dose entre périodes (période 1 - période 2);
  estimate "OP1OP2|d11%" periode +1 -1 diet*periode 0 +1 0 -1 dose*periode +1 -1 0 0
          diet*dose*periode 0 +1 0 -1 0 0 0 0 / alpha=.10 cl;
  estimate "IP1IP2|d11%" periode +1 -1 diet*periode +1 0 -1 0 dose*periode +1 -1 0 0
          diet*dose*periode +1 0 -1 0 0 0 0 0 / alpha=.10 cl;
  estimate "OP1OP2|d33%" periode +1 -1 diet*periode 0 +1 0 -1 dose*periode 0 0 +1 -1
          diet*dose*periode 0 0 0 0 0 +1 0 -1 / alpha=.10 cl;
  estimate "IP1IP2|d33%" periode +1 -1 diet*periode +1 0 -1 0 dose*periode 0 0 +1 -1
          diet*dose*periode 0 0 0 0 +1 0 -1 0 / alpha=.10 cl;

* Comparaisons des évolutions 'OGM-ISO' entre périodes (période 1 - période 2);
  estimate "D*P|d11" diet*periode -1 +1 +1 -1 dose*periode*diet -1 +1 +1 -1 0 0 0 0 / alpha=.10 cl;
  estimate "D*P|d33" diet*periode -1 +1 +1 -1 dose*periode*diet 0 0 0 0 -1 +1 +1 -1 / alpha=.10 cl;
  ods output estimates=Est;
  ods output covparms=var;
  ods output tests3=tests;
run;
quit;
data num;
do sexe="F","M";

```

```
do contraste=1 to 15;
output;
end;
end;
run;
data est2;
merge num est;
by sexe;
run;
data Res;
set var;
where covparm="Residual" ;
VAR_Res=Estimate;
run;
data ID;
set var;
where covparm="id(dose*diet)" ;
VAR_ID=Estimate;
run;
data VAR2;
merge ID RES;
by sexe;
keep sexe VAR_ID VAR_Res Rho;
Rho=VAR_ID/(VAR_ID+VAR_Res);
run;

data num1;
do sexe="F","M";
do T=1 to 7;
output;
end;
end;
run;
data tests2;
merge num1 tests;
by sexe;
run;
proc transpose data=tests2 out=tests2 name=test
prefix=T;
by sexe;
run;
data tests2;
set tests2;
where test="ProbF";
keep sexe T1-T7;
run;
data tout;
merge est2 var2 tests2;
by sexe;
Variable="Var&v";
run;
data general2;
set general2 tout;
run;
%end;

%mend;
%mixedcontraste2;
```

```
/*-----
Modèle mixte pour les variables quantitatives sur une période
-----*/
```

```
%macro mixedcontraste3;
ods html;
data general3;
```

```

variable="PPPPPP"; output;
run;
%do v=59 %to 67;
ods graphics on;
Proc mixed data=mon_so.totalid covtest method=REML;
  title1 "Modèle mixte avec analyse de l'influence des observations (rats)";
  title2 "modèle 'Diètes-Doses-Périodes' ";
  by sexe ;
  where group<5;
  class id dose diet group ;
  model vart&v= diet dose diet*dose / residual
          influence(iter=5 effect=id(diet*dose) estimates);

  lsmeans diet*dose diet / alpha=.10 cl;

* Comparaisons des régimes par dose, par période et marginalement;
estimate "O-I|d=11%" diet -1 +1 diet*dose -1 +1 0 0 / alpha=.10 cl;
estimate "O-I|d=33%" diet -1 +1 diet*dose 0 0 -1 +1 / alpha=.10 cl;
estimate "O-I" diet -1 +1 / alpha=.10 cl;

ods output estimates=Est;
ods output covparms=var;
ods output tests3=tests;

run;
quit;
data num;
do sexe="F","M";
do contraste=1 to 3;
output;
end;
end;
run;
data est2;
merge num est;
by sexe;
run;
data Res;
set var;
where covparm="Residual" ;
VAR_Res=Estimate;
run;
data ID;
set var;
where covparm="id(dose*diet)" ;
VAR_ID=Estimate;
run;
data VAR2;
merge ID RES;
by sexe;
keep sexe VAR_ID VAR_Res Rho;
Rho=VAR_ID/(VAR_ID+VAR_Res);
run;

data num1;
do sexe="F","M";
do T=1 to 3;
output;
end;
end;
run;
data tests2;
merge num1 tests;
by sexe;
run;
proc transpose data=tests2 out=tests2 name=test
prefix=T;

```

```

by sexe;
run;
data tests2;
set tests2;
where test="ProbF";
keep sexe T1-T3;
run;
data tout;
merge est2 var2 tests2;
by sexe;
Variable="Var&v";
run;
data general3;
set general3 tout;
run;
%end;
%mend;
%mixedcontraste3;

```

```

data general1; set general; where contraste=1 or contraste=2 or contraste=3 or contraste=4; run;
data general2; set general2; where contraste=1 or contraste=2 or contraste=3 or contraste=4; run;
data general3; set general3; where contraste=1 or contraste=2 ; run;
data mon_so.tabfinale;
  set general1 general2 general3;
run;

```

```

/* -----
Tests non paramétriques appliqués à toutes les données AVEC les données atypiques
-----*/

```

```

data mon_so.total5_non_para;
set mon_so.total5;
ID=Kcompress(Kstrcat(group,sexe,animal));
if group=1 or group=2 then diet="ISOnp";
if group=3 or group=4 then diet="OGMnp";
if group=1 or group=3 then dose="11np%";
if group=2 or group=4 then dose="33np%";
run;

ods graphics on;
proc npar1way data=mon_so.total5_non_para wilcoxon median
  plots=(wilcoxonboxplot medianplot);
  class diet;
  var var1-var65;
  where group<5;
  by sexe periode dose;
  ods output WilcoxonTest=Wil;
run;
ods graphics off;

```

```

/* -----
Tests non paramétriques appliqués à toutes les données SANS les données atypiques
-----*/

```

```

data mon_so.total5_sans_atypique_non_para;
set mon_so.total5_sans_atypique;
ID=Kcompress(Kstrcat(group,sexe,animal));
if group=1 or group=2 then diet="ISOnpsda";
if group=3 or group=4 then diet="OGMnpsda";
if group=1 or group=3 then dose="11%npsda";
if group=2 or group=4 then dose="33%npsda";
run;

ods graphics on;
proc npar1way data=mon_so.total5_sans_atypique_non_para wilcoxon median

```

```

plots=(wilcoxonboxplot medianplot);
class diet;
var var1-var65;
where group<5;
by sexe periode dose;
ods output WilcoxonTest=Wil3;
run;
ods graphics off;

/*****
Calcul des moyennes et des effectifs par groupe
*****/

%let g1=1;
%let g2=3;
proc sort data=mon_so.totalid;
  by periode sexe;
run;
proc ttest data=mon_so.totalid;
  class group;
  var vart01-vart67;
  by periode sexe;
  where group=&g1 or group=&g2;
  ods output statistics=Test;
run;

data test&g1;
set test;
where class="&g1";
keep periode--mean stddev;
rename mean=M&g1;
rename stddev=STD&g1;
rename N=N&g1;
drop class;
run;
data test&g2;
set test;
where class="&g2";
keep periode--mean stddev;
rename mean=M&g2;
rename stddev=STD&g2;
rename N=N&g2;
drop class;
run;
proc sort data=test&g1;
by periode sexe variable;
run;
proc sort data=test&g2;
by periode sexe variable;
run;
data moyenne1;
merge test&g1 test&g2;
by periode sexe variable;
run;
*****,
*****,
%let g1=2;
%let g2=4;
proc sort data=mon_so.totalid ;
by periode sexe;
run;
proc ttest data=mon_so.totalid ;
class group;
var vart01-vart65;
by periode sexe;

```

```

where group=&g1 or group=&g2;
ods output statistics=Test;
run;

data test&g1;
set test;
where class="&g1";
keep periode--mean stddev;
rename mean=M&g1;
rename stddev=STD&g1;
rename N=N&g1;
drop class;
run;
data test&g2;
set test;
where class="&g2";
keep periode--mean stddev;
rename mean=M&g2;
rename stddev=STD&g2;
rename N=N&g2;
drop class;
run;
proc sort data=test&g1;
by periode sexe variable;
run;
proc sort data=test&g2;
by periode sexe variable;
run;
data moyenne2;
merge test&g1 test&g2;
by periode sexe variable;
run;
*****
*****
data mon_so.moyenne;
merge moyenne1 moyenne2;
by periode sexe variable;
run;

data tabfinale;
set mon_so.tabfinale;
if Var_ID=. then Var_ID=0;
if contraste=1 then output;
run;

proc sql;
create table mon_so.ogmres as
select a.periode,a.sexe,a.variable,a.N1,a.N3,a.N2,a.N4,a.M1,a.M3,a.M2,a.M4,b.Var_ID,b.Var_RES
from moyenne as a,
tabfinale as b
where a.Variable=b.variable
and a.sexe=b.sexe;
quit;
run;

/*****
Analyse pour la multiplicité des tests;
*****/

* Tri préalable par valeurs croissantes de probt;
* -----;
proc sort data=mon_so.Tabfinale2 out=w_Tabfinale2; by probt; run;

```

```
ods graphics on;
proc multtest inpvalues(Probt)=w_Tabfinale2
  plots(signonly=0.5)=(all rawprob(unpack)adjusted(unpack) pbytest(notestname vref=0.1)) seed=518498000
  fdr pfdr(positive) out=pvalues_aj;
run;
ods graphics off;

* Calcul de l'Espérance de faux positifs en fonction de la pvalue;
* -----;
proc sort data=pvalues_aj; by probt; run;
data w_pvalues_aj;
  set pvalues_aj;
  EpFDR=_n_*pFDR_p;
  nb_tests_sig=_n_;
run;
proc print data=w_pvalues_aj;
  var probt pFDR_p EpFDR nb_tests_sig;
  where probt le 0.15;
run;

* Représentation graphique du pFDR en fonction des pvalues;
* -----;
options ls=75 ps=1000;
goptions reset=all;
symbol1 i=none color=blue value=dot;
symbol2 i=join color=black value=dot;
symbol3 i=join color=black value=dot;
symbol4 i=join color=black value=dot;
symbol5 i=join color=red value=none;

* Graphique des pFDR(Storey) en fonction des pvalues;
* -----;
axis1 label=(angle=90 h=2 'pFDR (Storey)') minor=(n=5) order=(0 to 1 by 0.1);
axis2 label=(h=2 'p-values') minor=(n=5) order=(0 to 1 by 0.1);
proc gplot data=w_pvalues_aj;
  plot pfdr_p*probt=5/haxis=axis2 vaxis=axis1;
run; quit;

* Graphique des FDR(HB) en fonction des pvalues;
* -----;
axis1 label=(angle=90 h=2 'FDR (BH)') minor=(n=5) order=(0 to 1 by 0.1);
axis2 label=(h=2 'p-values') minor=(n=5) order=(0 to 1 by 0.1);
proc gplot data=w_pvalues_aj;
  plot fdr_p*probt=5/haxis=axis2 vaxis=axis1;
run; quit;

* Nombre de Faux positifs en fonction de la qvalue;
* -----;
axis1 label=(angle=90 h=2 'nombre de tests significatifs');
axis2 label=(h=2 'qvalues');
proc gplot data=w_pvalues_aj;
  plot nb_tests_sig*pfdr_p=5/haxis=axis2 vaxis=axis1 href=0.10;
  where probt le 0.5;
run;
quit;

* L'Espérance des Faux positifs en fonction de la pvalue;
* -----;
axis1 label=(angle=90 h=2 'FDR(HB)');
axis2 label=(h=2 'p-values');
proc gplot data=w_pvalues_aj;
  plot FDR_p*probt=5/haxis=axis2 vaxis=axis1;
  where probt le 0.5;
run;
quit;
```

* L'Espérance des Faux positifs en fonction du nombre de tests significatifs;

```

* -----;
axis1 label=(angle=90 h=2 'Espérance de Faux positifs');
axis2 label=(h=2 'nombre de tests significatifs');
proc gplot data=w_pvalues_aj;
  plot Epfd*nb_tests_sig=5/haxis=axis2 vaxis=axis1 href=47;
  where probt le 0.15;
run;
quit;

```

```

/*=====
Analyse de la puissance
=====*/

```

```

data mon_so.ogmres;
merge mon_so.ogmres mon_so.com (keep=sd_com) mon_so.hist(keep=sd_hist);
run;

```

```

data mon_so.equiv_new ;
set mon_so.ogmres;
Var_TOT=Var_RES+Var_ID; /*Var_TOT correspond à la variance totale */
tdiff13=abs(M1-M3)/sqrt(Var_TOT*(1/n1+1/n3));
pdiff13=2*sdf('t',tdiff13,n1+n3-2);
tdiff24=abs(M2-M4)/sqrt(Var_TOT*(1/n2+1/n4));
pdiff24=2*sdf('t',tdiff24,n2+n4-2);
tequivH13=tdiff13-sd_hist/sqrt(Var_TOT*(1/N1+1/N3));
pequivH13=cdf('t',tequivH13,n1+n3-2);
tequivH24=tdiff24-sd_hist/sqrt(Var_TOT*(1/n2+1/n4));
pequivH24=cdf('t',tequivH24,n2+n4-2);
tequivS13=tdiff13-sd_com/sqrt(Var_TOT*(1/N1+1/N3));
pequivS13=cdf('t',tequivS13,n1+n3-2);
tequivS24=tdiff24-sd_com/sqrt(Var_TOT*(1/n2+1/n4));
pequivS24=cdf('t',tequivS24,n2+n4-2);
run;

```

```

data d.puissance_COM_hist;
set d.Equiv_new;
s=sqrt(Var_TOT); /* s correspond à l'écart type de la variance totale*/
n13=N1+N3;
n24=N2+N4;
if n13=20 then deltaL13= s * 1.156;
if n13=19 then deltaL13= s * 1.189;
if n13=18 then deltaL13= s * 1.225;
if n13=39 then deltaL13= s * 0.811;
if n13=40 then deltaL13= s * 0.801;
if n24=20 then deltaL24= s * 1.156;
if n24=19 then deltaL24= s * 1.189;
if n24=18 then deltaL24= s * 1.225;
if n24=39 then deltaL24= s * 0.811;
if n24=40 then deltaL24= s * 0.801;
ncpS13=(sd_com)**2/(Var_TOT*(1/n1+1/n3));
ncpS24=(sd_com)**2/(Var_TOT*(1/n2+1/n4));
CriticalValue13 = FINV(1-0.1, 1,N1+N3-2, 0);
CriticalValue24 = FINV(1-0.1, 1,N2+N4-2, 0);
PowerS13 = SDF('F', CriticalValue13,1, N1+N3-2, ncpS13);
PowerS24 = SDF('F', CriticalValue24,1, N2+N4-2, ncpS24);

```

*Calcul du plus petit écart détectable avec une puissance de 80%

```

* -----;
DTI8013=0;
dti8024=0;
do i=1 to 500; /* Détransformation */
  x3=normal(0)*(Var_TOT**0.5)+M3+deltaL13;
  if svar=1 then y3=exp(x3)+bvar;

```

```

if svar=0 then y3=x3;
if svar=-1 and bvar ne 0 then y3=bvar*((x3**(1/bvar))-1);
if svar=-1 and bvar=0 then y3=x3;
DTI8013=y3/500+DTI8013;

x4=normal(0)*(Var_TOT**0.5)+M4+deltaL24;
if svar=1 then y4=exp(x4)+bvar;
if svar=0 then y4=x4;
if svar=-1 and bvar ne 0 then y4=bvar*((x4**(1/bvar))-1);
if svar=-1 and bvar=0 then y4=x4;
dti8024=y4/500+dti8024;
end;

```

* Calcul du delta détransformé pour les données commerciales

```

*-----;
DTIS13=0;
DTIS24=0;
do i=1 to 500;
x3=normal(0)*(Var_TOT**0.5)+M3+sd_com;
if svar=1 then y3=exp(x3)+bvar;
if svar=0 then y3=x3;
if svar=-1 and bvar ne 0 then y3=bvar*((x3**(1/bvar))-1);
if svar=-1 and bvar=0 then y3=x3;
DTIS13=y3/500+DTIS13;

x4=normal(0)*(var_TOT**0.5)+M4+sd_com;
if svar=1 then y4=exp(x4)+bvar;
if svar=0 then y4=x4;
if svar=-1 and bvar ne 0 then y4=bvar*((x4**(1/bvar))-1);
if svar=-1 and bvar=0 then y4=x4;
DTIS24=y4/500+DTIS24;
end;
run;

data d.puissance_COM_hist;
set d.puissance_COM_hist;
DTIH13=DTIH13-MTI3; /*MTI3 correspond à la moyenne détransformée du groupe 3*/
DTIH24=DTIH24-MTI4;
DTIS13=DTIS13-MTI3;
DTIS24=DTIS24-MTI4;
dti8013=dti8013-MTI3;
dti8024=dti8024-MTI4;
run;

```

* Calcul de la puissance pour un effectif de 20 individus

```

*-----;
data d.puissance_COM_hist_n20;
set d.Equiv_new;
ncpS13=(sd_com)**2/(var_TOT*0.1);
ncpS24=(sd_com)**2/(var_TOT*0.1);
CriticalValue13 = FINV(1-0.1, 1,40-2, 0);
CriticalValue24 = FINV(1-0.1, 1,40-2, 0);
PowerS13 = SDF('F', CriticalValue13,1, 40-2, NCPS13);
PowerS24 = SDF('F', CriticalValue24,1, 40-2, NCPS24);
run;

```

```

/*=====
Tests d'équivalence
=====*/

```

```

data d.equiv_com_hist2;
set d.Equiv_com_hist;
tc13=sqrt(FINV(0.80,1,n1+n3-2,0));
tc24=sqrt(FINV(0.80,1,n2+n4-2,0));

```

```

erg13=abs(M1+M3)+tc13*sqrt(Var_TOT*(1/n1+1/n3));
erg24=abs(M2+M4)+tc24*sqrt(Var_TOT*(1/n2+1/n4));
/*Détransformation de l'erg pour pouvoir le comparer avec l'écart type historique */
ergDTIH13=0;
ergDTIH24=0;
do i=1 to 500;
  x3=normal(0)*(Var_TOT**0.5)+M3+erg13;
  if svar=1 then y3=exp(x3)+bvar;
  if svar=0 then y3=x3;
  if svar=-1 and bvar ne 0 then y3=bvar*((x3**(1/bvar))-1);
  if svar=-1 and bvar=0 then y3=x3;
  ergDTIH13=y3/500+ergDTIH13;

  x4=normal(0)*(Var_TOT**0.5)+M4+erg24;
  if svar=1 then y4=exp(x4)+bvar;
  if svar=0 then y4=x4;
  if svar=-1 and bvar ne 0 then y4=bvar*((x4**(1/bvar))-1);
  if svar=-1 and bvar=0 then y4=x4;
  ergDTIH24=y4/500+ergDTIH24;
end;
run;

* Calcul des équivalences pour un effectif de 20 individus
*-----;
data d.equiv_com_hist3;
set d.Equiv_com_hist;
tc13=sqrt(FINV(0.80,1,40-2,0));
tc24=sqrt(FINV(0.80,1,40-2,0));
erg13=abs(M1+M3)+tc13*sqrt(Var_TOT*(1/20+1/20));
erg24=abs(M2+M4)+tc24*sqrt(Var_TOT*(1/20+1/20));
/*Détransformation de l'erg pour pouvoir le comparer avec l'écart type historique */
ergDTIH13=0;
ergDTIH24=0;
do i=1 to 500;
  x3=normal(0)*(Var_TOT**0.5)+M3+erg13;
  if svar=1 then y3=exp(x3)+bvar;
  if svar=0 then y3=x3;
  if svar=-1 and bvar ne 0 then y3=bvar*((x3**(1/bvar))-1);
  if svar=-1 and bvar=0 then y3=x3;
  ergDTIH13=y3/500+ergDTIH13;

  x4=normal(0)*(Var_TOT**0.5)+M4+erg24;
  if svar=1 then y4=exp(x4)+bvar;
  if svar=0 then y4=x4;
  if svar=-1 and bvar ne 0 then y4=bvar*((x4**(1/bvar))-1);
  if svar=-1 and bvar=0 then y4=x4;
  ergDTIH24=y4/500+ergDTIH24;
end;
run;

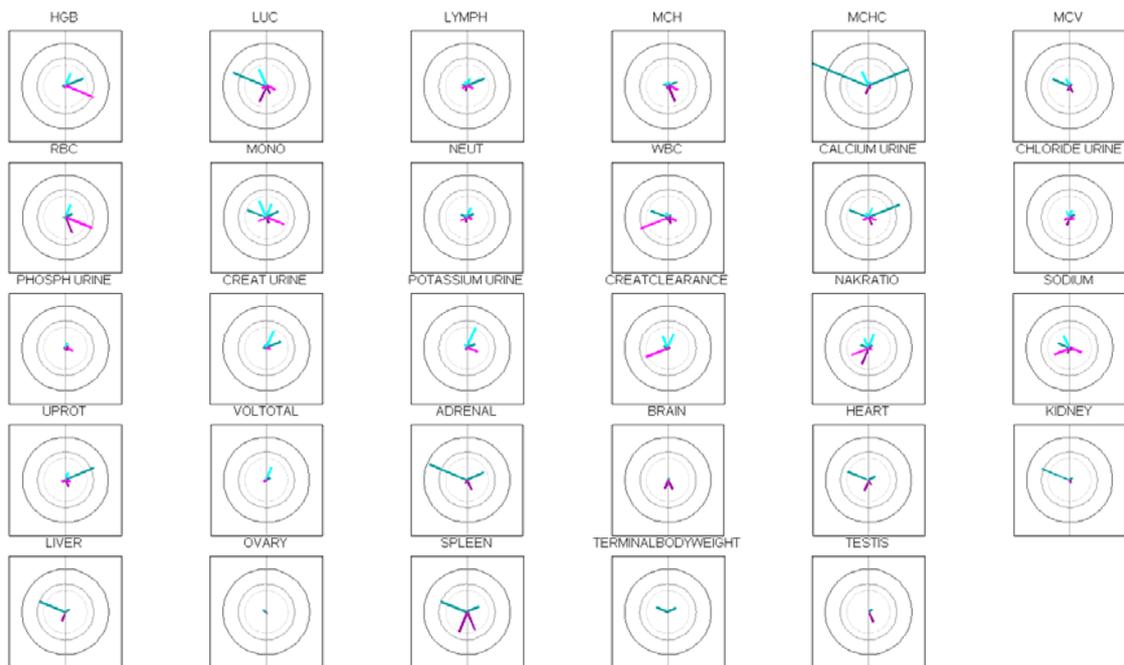
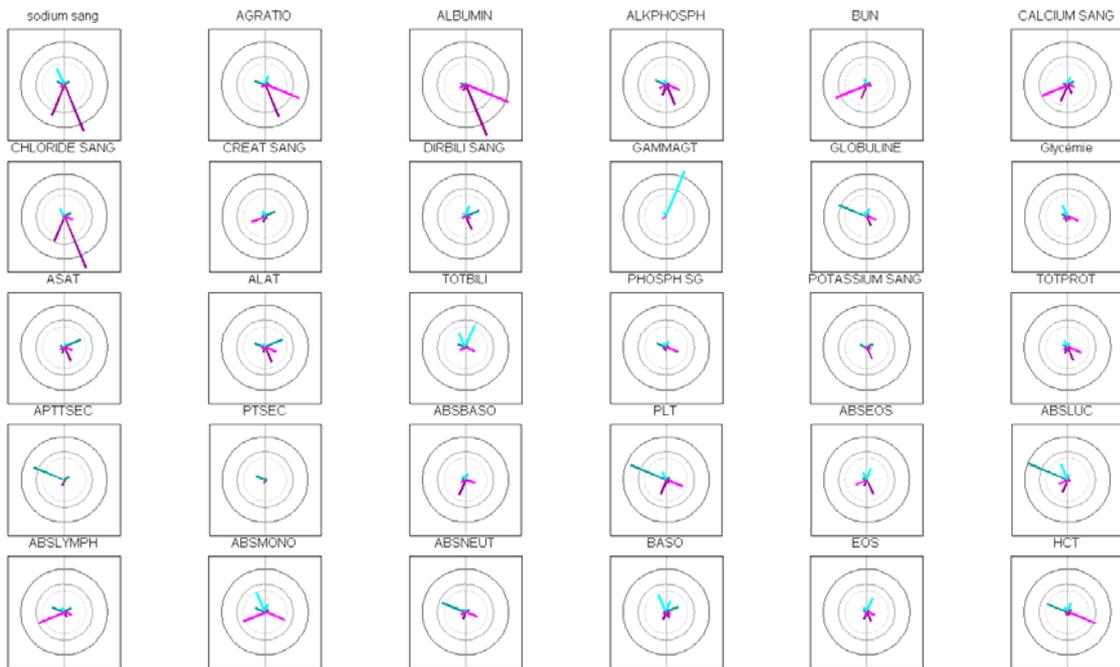
```

ANNEXE 7 : Comparaison des tests paramétriques et non paramétriques pour les variables quantitatives

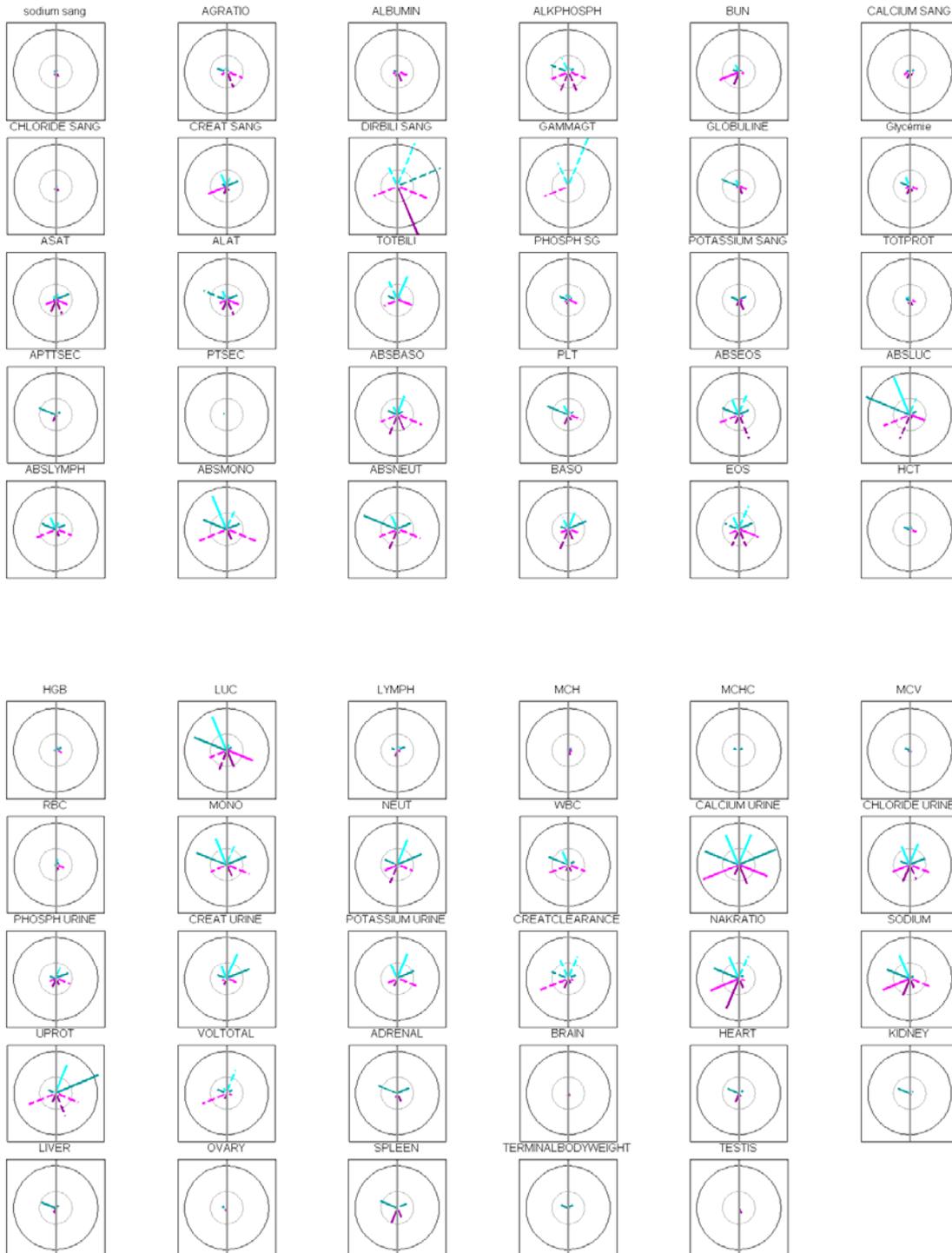
Les pValues significatives au risque 10% sont surlignées en vert pour les tests non paramétriques et en rouge pour les paramétriques

Paramètre	sexe	periode	non paramétrique		paramétrique		
			pValue 1-3	pValue 24	pValue 1-3	p Value 2-4	
Var02	GRAVITY	M	1	0,081	0,762	0,192	0,708
Var09	SODIUM_Sang	F	2	0,820	0,084	0,457	0,035
Var09	SODIUM_Sang	M	2	0,308	0,138	0,551	0,007
Var10	AGRATIO	M	1	0,850	0,049	0,696	0,025
Var10	AGRATIO	M	2	0,623	0,025	0,874	0,029
Var11	ALBUMIN	M	1	0,818	0,004	0,998	0,010
Var11	ALBUMIN	M	2	0,909	0,003	0,913	0,004
Var13	BUN	F	1	0,473	0,017	0,543	0,036
Var14	CALCIUM_Sang	F	1	0,880	0,074	0,778	0,066
Var15	CHLORIDE_Sang	F	2	0,939	0,115	0,954	0,067
Var15	CHLORIDE_Sang	M	2	0,672	0,047	0,444	0,004
Var19	GLOBULINE	F	2	0,060	0,969	0,047	0,850
Var21	SGOTAST	M	2	0,064	0,226	0,149	0,204
Var23	TOTBILI	M	1	0,168	0,583	0,078	0,363
Var23	TOTBILI	M	2	0,077	0,168	0,783	0,854
Var27	APTTSEC	F	2	0,081	0,307	0,037	0,530
Var29	ABSBASO	F	2	0,847	0,081	0,724	0,191
Var30	PLT	F	2	0,104	0,186	0,022	0,209
Var32	ABSLUC	F	1	0,160	0,080	0,176	0,377
Var32	ABSLUC	F	2	0,095	0,448	0,015	0,243
Var33	ABSLYMPH	F	1	0,473	0,038	0,621	0,064
Var34	ABSMONO	F	1	0,363	0,081	0,109	0,094
Var35	ABSNEUT	F	2	0,290	0,571	0,077	0,480
Var36	BASO	F	2	0,583	0,046	0,770	0,426
Var38	HCT	M	1	0,252	0,072	0,370	0,045
Var39	HGB	M	1	0,325	0,045	0,251	0,046
Var39	HGB	M	2	0,086	0,909	0,133	0,948
Var40	LUC	F	1	0,062	0,207	0,153	0,579
Var40	LUC	F	2	0,103	0,255	0,028	0,175
Var43	MCHC	F	1	0,081	0,820	0,194	0,979
Var43	MCHC	F	2	0,003	0,622	0,002	0,439
Var43	MCHC	M	2	0,025	0,676	0,016	0,732
Var45	RBC	M	1	0,206	0,066	0,239	0,047
Var48	WBC	F	1	0,473	0,026	0,586	0,051
Var49	CALCIUM	M	2	0,236	0,212	0,029	0,444
Var52	CREAT	M	1	0,070	0,734	0,151	0,666
Var53	POTASSIUM	M	1	0,059	0,345	0,133	0,314
Var54	CREATCLEARANCE	F	1	0,426	0,066	0,254	0,076
Var57	UPROT	M	2	1,000	0,273	0,044	0,462
Var59	ADRENAL	F	2	0,038	0,860	0,016	0,698
Var61	HEART	F	2	0,053	0,344	0,098	0,295
Var62	KIDNEY	F	2	0,072	0,925	0,046	0,933
Var63	LIVER	F	2	0,120	0,474	0,060	0,378
Var65	SPLEEN	F	2	0,045	0,126	0,058	0,104
Var65	SPLEEN	M	2	0,465	0,096	0,261	0,124

ANNEXE 8 : p-Value pour le test de comparaison pour les différentes variables



ANNEXE 9 : Taille d'effet observée pour les différentes variables



ANNEXE 10 : Puissance

Paramètre	sexe	periode	Ecart-type des données historiques	Ecart-type des données commerciales		880% : Delta détectable pour une puissance de 80%		Puissance données commerciales n=20	
				Puissance	Delta D2	Régime 11%	Régime 33%		
Var02	GRAVITY	F	1	0,01	0,626	0,014	0,020	0,019	0,884
Var02	GRAVITY	F	2	0,01	0,626	0,015	0,027	0,020	0,884
Var02	GRAVITY	M	1	0,01	0,562	0,008	0,011	0,012	0,830
Var02	GRAVITY	M	2	0,01	0,562	0,01	0,015	0,015	0,830
Var06	pH	F	1		0,785	0,428	0,482	0,449	0,969
Var06	pH	F	2		0,785	0,431	0,471	0,426	0,969
Var06	pH	M	1		0,719	0,387	0,400	0,375	0,942
Var06	pH	M	2		0,719	0,381	0,429	0,410	0,942
Var09	SODIUM_Sang	F	1	8,60	0,631	1,882	2,323	2,313	0,888
Var09	SODIUM_Sang	F	2	7,00	0,631	1,942	2,439	2,473	0,888
Var09	SODIUM_Sang	M	1	9,00	0,530	2,199	2,973	3,132	0,798
Var09	SODIUM_Sang	M	2	9,00	0,530	2,486	3,609	3,575	0,798
Var10	AGRATIO	F	1	0,10	0,696	0,311	0,357	0,341	0,929
Var10	AGRATIO	F	2	0,10	0,696	0,32	0,394	0,368	0,929
Var10	AGRATIO	M	1	0,10	0,608	0,175	0,228	0,206	0,870
Var10	AGRATIO	M	2	0,10	0,608	0,178	0,222	0,230	0,870
Var11	ALBUMIN	F	1	0,40	0,737	0,37	0,377	0,400	0,950
Var11	ALBUMIN	F	2	0,40	0,737	0,326	0,403	0,396	0,950
Var11	ALBUMIN	M	1	0,40	0,753	0,246	0,231	0,235	0,957
Var11	ALBUMIN	M	2	0,30	0,753	0,225	0,250	0,244	0,957
Var12	ALKPHOSPH	F	1	99,50	0,602	25,136	34,734	33,393	0,865
Var12	ALKPHOSPH	F	2	53,00	0,602	15,735	21,068	18,174	0,865
Var12	ALKPHOSPH	M	1	113,50	0,740	31,976	34,399	34,917	0,952
Var12	ALKPHOSPH	M	2	73,60	0,740	18,691	18,986	17,579	0,952
Var13	BUN	F	1	21,60	0,761	2,302	2,615	2,303	0,960
Var13	BUN	F	2	22,20	0,761	2,6	2,787	2,702	0,960
Var13	BUN	M	1	22,20	0,717	1,723	1,976	2,082	0,941
Var13	BUN	M	2	26,40	0,717	2,179	2,501	2,479	0,941
Var14	CALCIUM_Sang	F	1	0,40	0,686	0,386	0,433	0,435	0,924
Var14	CALCIUM_Sang	F	2	0,40	0,686	0,703	0,815	0,825	0,924
Var14	CALCIUM_Sang	M	1	0,40	0,608	0,349	0,415	0,500	0,869
Var14	CALCIUM_Sang	M	2	0,40	0,608	0,689	0,834	0,882	0,869
Var15	CHLORIDE_Sang	F	1	4,50	0,765	2,145	2,202	2,262	0,962
Var15	CHLORIDE_Sang	F	2	4,90	0,765	1,685	2,049	2,002	0,962
Var15	CHLORIDE_Sang	M	1	3,90	0,862	2,428	2,477	2,380	0,989
Var15	CHLORIDE_Sang	M	2	4,90	0,862	2,387	2,244	2,069	0,989
Var16	Creat sang	F	1	0,10	0,677	0,075	0,086	0,084	0,918
Var16	Creat sang	F	2	0,10	0,677	0,071	0,088	0,074	0,918
Var16	Creat sang	M	1	0,10	0,586	0,058	0,078	0,076	0,851
Var16	Creat sang	M	2	0,10	0,586	0,064	0,077	0,078	0,851
Var17	DIRBILI	F	1	0,01	0,541	0,044	0,060	0,053	0,810
Var17	DIRBILI	F	2	0,01	0,541	0,04	0,059	0,058	0,810
Var17	DIRBILI	M	1	0,01	0,869	0,047	0,042	0,043	0,991
Var17	DIRBILI	M	2	0,02	0,869	0,045	0,040	0,041	0,991
Var19	GLOBULINE	F	1	0,20	0,634	0,235	0,290	0,292	0,890
Var19	GLOBULINE	F	2	0,10	0,634	0,276	0,306	0,309	0,890
Var19	GLOBULINE	M	1	0,30	0,627	0,207	0,258	0,255	0,884
Var19	GLOBULINE	M	2	0,30	0,627	0,212	0,275	0,268	0,884
Var20	Glycémie	F	1	21,60	0,689	15,319	20,267	18,718	0,937
Var20	Glycémie	F	2	16,20	0,711	31,369	34,860	36,061	0,937
Var20	Glycémie	M	1	21,60	0,862	15,507	11,558	12,979	0,989
Var20	Glycémie	M	2	18,00	0,862	37,963	33,113	33,147	0,989
Var21	SGOTAST	F	1	20,20	0,772	24,074	26,899	25,360	0,965
Var21	SGOTAST	F	2	16,80	0,772	20,81	24,658	21,742	0,965
Var21	SGOTAST	M	1	31,30	0,673	14,782	19,975	18,423	0,916
Var21	SGOTAST	M	2	25,30	0,673	11,833	13,127	14,273	0,916
Var22	SGPTALT	F	1	8,70	0,906	14,519	11,101	11,593	0,996
Var22	SGPTALT	F	2	7,90	0,906	18,489	14,468	15,722	0,996
Var22	SGPTALT	M	1	11,40	0,728	6,224	7,001	7,026	0,946
Var22	SGPTALT	M	2	18,00	0,728	7,982	7,193	8,468	0,946

La dernière colonne correspond au calcul de la puissance pour les données commerciales avec un effectif de 20 individus. Dans les colonnes puissance : les ronds verts, oranges et rouges indiquent respectivement, une puissance supérieure ou égale à 80%, entre 70 et 80% et strictement inférieure à 70%.

Paramètre	sexe	periode	Ecart-type des données historiques	Ecart-type des données commerciales		880% : Delta détectable pour une puissance de 80%		Puissance données commerciales n=20		
				Puissance	Delta D2	Régime 11%	Régime 33%			
Var23	TOTBILI	F	1	●	0,689	0,034	0,037	0,039	●	0,925
Var23	TOTBILI	F	2	●	0,689	0,031	0,038	0,039	●	0,925
Var23	TOTBILI	M	1	●	0,907	0,033	0,027	0,028	●	0,996
Var23	TOTBILI	M	2	●	0,907	0,032	0,028	0,027	●	0,996
Var24	PHOSPH_Sang	F	1	●	0,774	0,851	0,819	0,838	●	0,965
Var24	PHOSPH_Sang	F	2	●	0,774	1,216	1,414	1,332	●	0,965
Var24	PHOSPH_Sang	M	1	●	0,781	0,752	0,796	0,815	●	0,968
Var24	PHOSPH_Sang	M	2	●	0,781	1,214	1,314	1,206	●	0,968
Var25	POTASSIUM_Sang	F	1	●	0,40	0,689	0,385	0,447	●	0,925
Var25	POTASSIUM_Sang	F	2	●	0,50	0,689	1,028	1,245	●	0,925
Var25	POTASSIUM_Sang	M	1	●	0,50	0,763	0,48	0,566	●	0,961
Var25	POTASSIUM_Sang	M	2	●	0,50	0,763	1,114	1,024	●	0,961
Var26	TOTPROT	F	1	●	0,40	0,670	0,371	0,431	●	0,914
Var26	TOTPROT	F	2	●	0,40	0,670	0,416	0,569	●	0,914
Var26	TOTPROT	M	1	●	0,40	0,682	0,322	0,339	●	0,921
Var26	TOTPROT	M	2	●	0,30	0,682	0,398	0,502	●	0,921
Var27	APTTSEC	F	2	●	0,782	1,672	1,852	1,681	●	0,968
Var27	APTTSEC	M	2	●	0,911	1,722	1,489	1,566	●	0,996
Var28	PTSEC	F	2	●	1,80	0,807	0,582	0,558	●	0,976
Var28	PTSEC	M	2	●	1,20	0,719	0,541	0,691	●	0,941
Var29	ABSBASO	F	1	●	0,855	0,016	0,014	0,015	●	0,988
Var29	ABSBASO	F	2	●	0,855	0,027	0,024	0,024	●	0,988
Var29	ABSBASO	M	1	●	0,556	0,022	0,030	0,035	●	0,844
Var29	ABSBASO	M	2	●	0,577	0,028	0,038	0,037	●	0,844
Var30	PLT	F	1	●	141,50	0,617	123,272	156,629	●	0,877
Var30	PLT	F	2	●	140,90	0,617	108,401	128,135	●	0,877
Var30	PLT	M	1	●	153,00	0,527	142,407	188,161	●	0,816
Var30	PLT	M	2	●	136,20	0,548	128,378	187,135	●	0,816
Var31	ABSEOS	F	1	●	0,750	0,069	0,064	0,081	●	0,956
Var31	ABSEOS	F	2	●	0,750	0,058	0,054	0,070	●	0,956
Var31	ABSEOS	M	1	●	0,706	0,071	0,083	0,083	●	0,946
Var31	ABSEOS	M	2	●	0,728	0,063	0,072	0,073	●	0,946
Var32	ABSLUC	F	1	●	0,558	0,053	0,074	0,078	●	0,826
Var32	ABSLUC	F	2	●	0,558	0,059	0,081	0,081	●	0,826
Var32	ABSLUC	M	1	●	0,523	0,065	0,095	0,099	●	0,811
Var32	ABSLUC	M	2	●	0,543	0,062	0,101	0,093	●	0,811
Var33	ABSLYMPH	F	1	●	0,802	2,776	2,404	2,632	●	0,975
Var33	ABSLYMPH	F	2	●	0,802	1,79	1,837	1,723	●	0,975
Var33	ABSLYMPH	M	1	●	0,645	3,775	4,432	4,877	●	0,912
Var33	ABSLYMPH	M	2	●	0,667	2,475	3,086	3,010	●	0,912
Var34	ABSMONO	F	1	●	0,703	0,087	0,090	0,097	●	0,933
Var34	ABSMONO	F	2	●	0,703	0,084	0,098	0,102	●	0,933
Var34	ABSMONO	M	1	●	0,592	0,131	0,180	0,197	●	0,874
Var34	ABSMONO	M	2	●	0,614	0,11	0,144	0,148	●	0,874
Var35	ABSNEUT	F	1	●	0,991	0,774	0,464	0,454	●	1,000
Var35	ABSNEUT	F	2	●	0,991	0,805	0,453	0,485	●	1,000
Var35	ABSNEUT	M	1	●	0,650	0,457	0,591	0,560	●	0,915
Var35	ABSNEUT	M	2	●	0,672	0,467	0,593	0,538	●	0,915
Var36	BASO	F	1	●	0,10	0,418	0,103	0,203	●	0,665
Var36	BASO	F	2	●	0,10	0,418	0,32	0,525	●	0,665
Var36	BASO	M	1	●	0,10	0,637	0,107	0,132	●	0,907
Var36	BASO	M	2	●	0,10	0,659	0,165	0,228	●	0,907
Var37	EOS	F	1	●	0,60	0,677	0,668	0,629	●	0,918
Var37	EOS	F	2	●	0,80	0,677	0,512	0,700	●	0,918
Var37	EOS	M	1	●	0,70	0,634	0,396	0,609	●	0,905
Var37	EOS	M	2	●	0,90	0,656	0,501	0,587	●	0,905
Var38	HCT	F	1	●	2,40	0,734	1,639	1,703	●	0,949
Var38	HCT	F	2	●	2,10	0,734	1,771	1,724	●	0,949
Var38	HCT	M	1	●	2,40	0,711	1,535	1,743	●	0,948
Var38	HCT	M	2	●	2,50	0,733	1,692	1,752	●	0,948
Var39	HGB	F	1	●	0,70	0,809	0,575	0,599	●	0,977
Var39	HGB	F	2	●	0,70	0,809	0,612	0,590	●	0,977
Var39	HGB	M	1	●	0,70	0,744	0,495	0,640	●	0,962
Var39	HGB	M	2	●	0,70	0,765	0,623	0,621	●	0,962

Paramètre	sexe	periode	Ecart-type des données historiques	Ecart-type des données commerciales		880% : Delta détectable pour une puissance de 80%		Puissance données commerciales n=20	
				Puissance	Delta D2	Régime 11%	Régime 33%		
Var40	LUC	F	1	0,40	0,594	4,906	4,879	5,563	0,859
Var40	LUC	F	2	0,40	0,594	15,337	11,081	18,442	0,859
Var40	LUC	M	1	0,40	0,641	5,81	8,627	6,759	0,909
Var40	LUC	M	2	0,40	0,663	12,738	18,037	15,087	0,909
Var41	LYMPH	F	1	4,60	0,711	3,054	3,579	3,423	0,938
Var41	LYMPH	F	2	4,40	0,711				0,938
Var41	LYMPH	M	1	5,40	0,716	3,425	3,847	3,860	0,951
Var41	LYMPH	M	2	6,30	0,738		4,684		0,951
Var42	MCH	F	1	0,60	0,792	0,548	0,541	0,561	0,971
Var42	MCH	F	2	0,60	0,792	0,583	0,556	0,522	0,971
Var42	MCH	M	1	0,50	0,594	0,539	0,736	0,683	0,876
Var42	MCH	M	2	0,80	0,616	0,569	0,738	0,712	0,876
Var43	MCHC	F	1	1,40	0,809	0,515	0,528	0,516	0,977
Var43	MCHC	F	2	1,00	0,809	0,581	0,529	0,572	0,977
Var43	MCHC	M	1	1,10	0,630	0,442	0,560	0,557	0,902
Var43	MCHC	M	2	1,10	0,652	0,479	0,653	0,588	0,902
Var44	MCV	F	1	1,80	0,720	1,657	1,946	1,957	0,942
Var44	MCV	F	2	2,10	0,720	1,832	1,961	2,144	0,942
Var44	MCV	M	1	1,80	0,658	1,912	2,297	2,330	0,932
Var44	MCV	M	2	3,00	0,680	1,531	1,795	1,897	0,932
Var45	RBC	F	1	0,40	0,783	0,339	0,360	0,371	0,969
Var45	RBC	F	2	0,40	0,783	0,331	0,348	0,367	0,969
Var45	RBC	M	1	0,40	0,650	0,345	0,436	0,447	0,915
Var45	RBC	M	2	0,50	0,672	0,369	0,441	0,427	0,915
Var46	MONO	F	1	0,80	0,683	0,806	0,856	1,048	0,922
Var46	MONO	F	2	0,80	0,683	1,055	1,171	1,241	0,922
Var46	MONO	M	1	0,80	0,827	0,952	0,899	0,962	0,986
Var46	MONO	M	2	0,90	0,846	1,093	0,896	0,903	0,986
Var47	NEUT	F	1	4,30	0,723	3,356	4,698	3,796	0,943
Var47	NEUT	F	2	4,00	0,723	6,049	6,629	6,611	0,943
Var47	NEUT	M	1	5,30	0,749	3,931	4,715	4,652	0,964
Var47	NEUT	M	2	5,90	0,770	5,472	5,472	5,670	0,964
Var48	WBC	F	1	2,20	0,815	3,403	2,509	3,364	0,979
Var48	WBC	F	2	2,40	0,815	2,39	2,194	2,287	0,979
Var48	WBC	M	1	2,40	0,622	4,467	5,440	6,680	0,897
Var48	WBC	M	2	2,40	0,644	3,109	3,695	3,597	0,897
Var49	CALCIUM	F	1		0,514	7,749	12,068	10,286	0,781
Var49	CALCIUM	F	2		0,514	7,188	11,523	11,610	0,781
Var49	CALCIUM	M	1		0,831	3,131	3,030	2,862	0,983
Var49	CALCIUM	M	2		0,831	2,973	3,200	2,917	0,983
Var50	CHLORIDE	F	1		0,581	29,529	42,788	47,170	0,847
Var50	CHLORIDE	F	2	27,32	0,581	23,389	54,188	31,872	0,847
Var50	CHLORIDE	M	1		0,503	27,189	38,235	45,958	0,770
Var50	CHLORIDE	M	2	20,41	0,503	27,316	45,713	44,299	0,770
Var51	PHOSPH	F	1		0,739	88,982	114,832	98,077	0,951
Var51	PHOSPH	F	2		0,739	89,206	128,352	95,203	0,951
Var51	PHOSPH	M	1		0,693	55,097	68,244	60,824	0,928
Var51	PHOSPH	M	2		0,693	53,268	72,086	65,726	0,928
Var52	CREAT	F	1		0,585	49,015	67,976	66,901	0,851
Var52	CREAT	F	2		0,585	47,518	78,631	66,507	0,851
Var52	CREAT	M	1		0,635	32,648	44,064	38,521	0,890
Var52	CREAT	M	2		0,635	38,473	56,173	46,227	0,890
Var53	POTASSIUM	F	1		0,672	65,551	84,588	75,720	0,915
Var53	POTASSIUM	F	2	27,32	0,672	68,822	90,166	78,776	0,915
Var53	POTASSIUM	M	1		0,566	49,938	63,579	66,377	0,834
Var53	POTASSIUM	M	2	40,82	0,566	53,31	70,769	67,749	0,834
Var54	CREATCLEARANCE	F	1		0,702	0,208	0,280	0,246	0,944
Var54	CREATCLEARANCE	F	2		0,702	0,129	0,147	0,143	0,944
Var54	CREATCLEARANCE	M	1		0,477	0,153	0,293	0,281	0,740
Var54	CREATCLEARANCE	M	2		0,477	0,089	0,154	0,152	0,740
Var55	NAKRATIO	F	1		0,643	0,119	0,176	0,159	0,896
Var55	NAKRATIO	F	2		0,643	0,133	0,162	0,169	0,896
Var55	NAKRATIO	M	1		0,761	0,123	0,107	0,138	0,968
Var55	NAKRATIO	M	2		0,782	0,127	0,118	0,131	0,968

Paramètre	sexe	periode	Ecart-type des données historiques	Ecart-type des données commerciales		880% : Delta détectable pour une puissance de 80%		Puissance données commerciales n=20	
				Puissance	Delta D2	Régime 11%	Régime 33%		
Var56	SODIUM	F	1		0,609	18,307	23,115	24,333	0,870
Var56	SODIUM	F	2	27,32	0,587	18,149	30,452	25,238	0,870
Var56	SODIUM	M	1		0,690	20,576	20,968	22,658	0,926
Var56	SODIUM	M	2	20,41	0,690	19,436	26,136	25,708	0,926
Var57	UPROT	F	1		0,845	24,657	24,743	23,072	0,990
Var57	UPROT	F	2		0,845	24,993	24,017	22,691	0,990
Var57	UPROT	M	1		0,474	69,309	109,187	103,927	0,736
Var57	UPROT	M	2		0,474	59,773	110,650	109,010	0,736
Var58	VOLTOTAL	F	1	2,61	0,592	2,881	4,842	4,750	0,857
Var58	VOLTOTAL	F	2	5,00	0,592	3,355	4,200	4,320	0,857
Var58	VOLTOTAL	M	1	2,05	0,587	3,2	3,737	4,512	0,852
Var58	VOLTOTAL	M	2	6,15	0,587	2,878	3,325	3,773	0,852
Var59	ADRENAL	F	2		0,956	0,013	0,007	0,009	0,956
Var59	ADRENAL	M	2		0,983	0,012	0,008	0,007	0,983
Var60	BRAIN	F	2		0,893	0,087	0,070	0,073	0,893
Var60	BRAIN	M	2		0,924	0,095	0,065	0,072	0,924
Var61	HEART	F	2		0,799	0,119	0,114	0,119	0,799
Var61	HEART	M	2		0,972	0,222	0,119	0,151	0,972
Var62	KIDNEY	F	2		0,872	0,201	0,188	0,172	0,872
Var62	KIDNEY	M	2		0,892	0,438	0,383	0,362	0,892
Var63	LIVER	F	2		0,968	1,367	0,818	1,033	0,968
Var63	LIVER	M	2		0,807	2,185	2,309	2,265	0,807
Var64	OVARY	F	2		0,941	0,021	0,017	0,017	0,941
Var65	SPLEEN	F	2		0,784	0,073	0,089	0,081	0,784
Var65	SPLEEN	M	2		0,925	0,136	0,113	0,123	0,925
Var66	TERMINALBODYWEIGHT	F	2		0,848	27,251	20,571	24,068	0,848
Var66	TERMINALBODYWEIGHT	M	2		0,899	60,087	47,555	55,379	0,899
Var67	TESTIS	M	2		0,893	0,292	0,247	0,247	0,893

ANNEXE 11 : Transformation des variables quantitatives

Svar=type de transformation (0 : pas de transformation, 1 : transformation logarithmique, -1 : transformation puissance), bvar : le paramètre de transformation.

Obs	variable	Namevar	svar	bvar
1	Var02	GRAVITY	1	0.99
2	Var06	pH	0	0.00
3	Var09	SODIUM_Sang	1	117.52
4	Var10	AGRATIO	1	-1.12
5	Var11	ALBUMIN	-1	1.18
6	Var12	ALKPHOSPH	1	6.08
7	Var13	BUN	1	0.11
8	Var14	CALCIUM_Sang	1	8.24
9	Var15	CHLORIDE_Sang	-1	3.13
10	Var16	Creat sang	0	0.00
11	Var17	DIRBILI	0	0.00
12	Var19	GLOBULINE	1	-3.15
13	Var20	Glycémie	1	48.30
14	Var21	SGOTAST	1	43.40
15	Var22	SGPTALT	1	9.80
16	Var23	TOTBILI	0	0.00
17	Var24	PHOSPH_Sang	1	2.01
18	Var25	POTASSIUM_Sang	1	3.13
19	Var26	TOTPROT	1	3.39
20	Var27	APTTSEC	-1	1.69
21	Var28	PTSEC	1	9.79
22	Var29	ABSBASO	1	-0.02
23	Var30	PLT	1	141.26
24	Var31	ABSEOS	1	-0.02

Obs	variable	Namevar	svar	bvar
25	Var32	ABSLUC	-1	0.00
26	Var33	ABSLYMPH	1	1.19
27	Var34	ABSMONO	1	-0.18
28	Var35	ABSNEUT	-1	0.00
29	Var36	BASO	-1	0.01
30	Var37	EOS	1	-1.03
31	Var38	HCT	-1	2.76
32	Var39	HGB	-1	4.24
33	Var40	LUC	1	0.00
34	Var41	LYMPH	-1	6.37
35	Var42	MCH	-1	1.22
36	Var43	MCHC	-1	6.45
37	Var44	MCV	1	37.90
38	Var45	RBC	-1	1.89
39	Var46	MONO	-1	0.15
40	Var47	NEUT	1	0.12
41	Var48	WBC	1	2.87
42	Var49	CALCIUM	-1	0.00
43	Var50	CHLORIDE	-1	0.10
44	Var51	PHOSPH	-1	0.48
45	Var52	CREAT	-1	0.36
46	Var53	POTASSIUM	-1	0.71
47	Var54	CREATCLEARANCE	1	-0.26
48	Var55	NAKRATIO	-1	0.01
49	Var56	SODIUM	1	-22.48
50	Var57	UPROT	-1	0.00
51	Var58	VOLTOTAL	1	-1.65
52	Var59	ADRENAL	1	-0.14
53	Var60	BRAIN	-1	15.52

Obs	variable	Namevar	svar	bvar
54	Var61	HEART	1	0.50
55	Var62	KIDNEY	1	-0.71
56	Var63	LIVER	1	4.22
57	Var64	OVARY	1	-0.05
58	Var65	SPLEEN	1	0.22
59	Var66	TERMINALBODYWEIGHT	1	74.14
60	Var67	TESTIS	1	-0.44

Annexe 12 : Démarche de l'analyse de l'insuffisance de puissance

Hématologie

La prise en compte du D2 exprimé à partir des moyennes des variétés commerciales, souvent inférieur aux SD historique, conduit à 55% de mesures avec une puissance supérieure à 70%, dont 20% avec une puissance supérieure à 80%.

Sur la base des données historiques disponibles, 53% des mesures ont une puissance supérieure à 80%. Dans ces conditions, il revenait au toxicologue de juger si la taille de l'effet retenu, en regard d'une puissance estimée insuffisante pour 47% des mesures, était de nature à occulter un possible effet toxique.

La démarche peut être illustrée à partir de quelques exemples.

Ainsi, pour le paramètre **monocytes en %**, aucune des 4 mesures²⁸ n'a une puissance supérieure à 80% avec la borne des données historiques SD, alors que c'est le cas pour 2 mesures sur 4, sur la base du D2. L'examen de la borne D2 (0.8 à 1.1), par rapport à celle des données historiques (0.8 à 0.9), explique cette différence de puissance en regard de la taille de l'effet détectable à 80% (0.86 à 1.2).

Au-delà de ce chiffre de 80% de puissance appliqué à une borne, c'est le sens biologique de la taille de l'effet que le toxicologue doit juger. La taille de 0.86 à 1.2 ne soulève pas d'interrogation compte tenu 1) de la nature du paramètre, et 2) des variations spontanées.

L'exemple des **plaquettes** pour lesquelles la puissance à 80% est atteinte en période 2 chez les femelles uniquement, est celle d'un paramètre qui varie de façon sensible chez l'Homme où les valeurs normales vont de 150 000 à 400 000 /mm³, comme chez le rat où ce même facteur de 2.5 peut être appliqué entre les valeurs minima et maxima chez des animaux témoins. Dans ces conditions, une taille d'effet qui excède légèrement un SD historique ne peut à elle seule remettre en cause la pertinence de l'évaluation qui prend en compte un ensemble de paramètres hématologiques.

Pour le paramètre **APTT** (temps de céphaline activée en seconde) où la puissance est supérieure à 90% chez les mâles, mais inférieure à 80% chez les femelles, la taille de l'effet détectable à 80% est de 1.49 à 1.85 alors que D2 varie entre 1.67 et 1.72. Cette différence de 0.18 et 0.13 seconde, à l'origine des puissances inférieures à 80% chez les femelles, n'a aucun sens en terme de signification d'un effet toxique.

Les taux de **basophiles** apparaît comme étant le seul paramètre pour lequel la puissance est insuffisante quelle que soit la borne considérée. Compte tenu de la faible contribution quantitative de cette sous population à la formule leucocytaire et de ses variations spontanées (0.0% à 0.6%), les valeurs des moyennes pour les groupes OGM, ISO, et variétés commerciales s'inscrivant dans une fourchette de 0.02 à 0.06% et d'absence de variations ayant une signification biologique, le toxicologue note cette insuffisance de puissance liée aux caractéristiques de cette sous population leucocytaire, sans conséquence pour l'interprétation toxicologique.

En conclusion, pour les paramètres dont la puissance est inférieure à 80%, l'examen des bornes D2 ou SD proposées pour juger de la taille de l'effet, associé à la prise en compte des variations "spontanées" mini-maxi de chacun des paramètres et de la cohérence des variations dans une approche multi-paramètres, permet d'estimer, qu'à l'exception des basophiles, aucune mesure ne présente une taille détectable susceptible d'induire un jugement erroné par rapport aux tests de différences mis en œuvre.

Biochimie sanguine

L'analyse des tableaux de puissance conduit à 41% de mesures avec une puissance supérieure à 70% (dont 11% sont supérieures à 80%) en prenant comme seuil le D2, et 55% de mesures avec une puissance supérieure à 80% avec le SD.

Ces chiffres doivent être rapprochés de la signification de la taille de l'effet qui a été prise en compte avec le même raisonnement que celui considéré pour les paramètres hématologiques.

Pour l'ensemble **Albumine, globuline, AGratio**, les profils se répartissent comme suit :

- Albumine, la puissance est supérieure à 70% avec D2 et à 80% avec SD pour les 4 mesures.

²⁸ Pour la plupart des paramètres, la puissance est calculée quatre fois pour chaque paramètre correspondant aux conditions de l'expérience et des comparaisons (2 sexes et 2 périodes).

- Globuline, la puissance est inférieure à 70% avec D2 pour les 4 mesures et supérieure à 80% pour 2 des 4 mesures avec SD.

Le toxicologue conclut qu'avec 4 mesures de puissance supérieure à 80% pour l'albumine et 2 pour les globulines (versus SD historiques), et en considérant la taille de l'effet détectable (tableau 1), la puissance est jugée suffisante.

Tableau 1 : Taille d'effet détectable comparée aux bornes de tolérance (SD, D2) des paramètres concentration en albumine et globuline dans le sang.

Paramètres (g/dL)	Taille de l'effet détectable	SD historique	D2 variétés commerciales	Mini-Maxi historique
Albumine	0.23 à 0.40	0.30 à 0.40	0.23 à 0.37	2.4 à 4.7
Globuline	0.26 à 0.31	0.10 à 0.30	0.21 à 0.28	2.9 à 4.9

AG ratio, pour aucune mesure la puissance est supérieure à 70% ou 80% selon la borne (tableau 2).

Tableau 2 : Taille d'effet détectable comparée aux bornes de tolérance (SD, D2) du paramètre rapport de la concentration de l'albumine sur la concentration de la globuline dans le sang.

Paramètre (g/dL)	Taille de l'effet détectable	SD historique	D2 variétés commerciales	Mini-Maxi historique
AG Ratio	0.21 à 0.40	0.10	0.18 à 0.32	0.5 à 1.2

Ce paramètre est calculé à partir des mesures précédentes, il n'appelle pas d'interrogations particulières du toxicologue, en regard notamment des autres paramètres hépatiques, du poids du foie et des reins, et des examens macroscopiques et microscopiques de ces organes cibles.

Pour les paramètres biochimiques de la **cible hépatique**, l'ALT considérée comme le paramètre majeur d'une hépatotoxicité a une puissance supérieure à 70% pour les 4 mesures versus D2 et pour 2 sur 4 mesures à plus de 80% versus SD. Pour AST et ALK, 2 sur 4 des mesures ont une puissance supérieure à 70% versus D2 et 4 sur 4 une puissance supérieure à 80% versus SD. Seule interrogation, la bilirubine avec 2 sur 4 des mesures supérieure à 80% de puissance versus D2 mais 0 sur 4 versus SD.

Si l'on considère par ailleurs le poids du foie pour lequel la puissance est supérieure à 80% et l'absence de signes histologiques, le toxicologue n'a pas de réserve sur la taille de l'effet détectable en regard des puissances.

Pour les paramètres biochimiques de la **cible rénale**, la puissance de BUN (concentration en urée dans le sang en mg/dl) est supérieure à 70% pour les 4 mesures versus D2 et supérieure à 80% pour 2 sur 4 mesures versus SD. Pour la créatinine, c'est 0 sur 4 versus D2 mais 4 sur 4 supérieures à 80% versus SD.

Le cas de la créatinine est intéressant, puisque selon la borne considérée, la taille de l'effet conduit à une puissance supérieure à 80% dans tous les cas (SD) ou inférieure à 70% dans tous les autres cas (D2).

Le tableau suivant illustre les raisons de cette différence (tableau 3).

Tableau 3 : Taille d'effet détectable comparé aux bornes de tolérance (SD, D2) du paramètre concentration en créatinine dans le sang.

Paramètre (mg/dL)	Taille de l'effet détectable	SD historique	D2 variétés commerciales	Mini-Maxi historique
Créatinine	0.07 à 0.09	0.10	0.06 à 0.08	3.0 à 10.0

La taille de l'effet détectable étant inférieure au SD historique pris comme référence, la puissance est donc supérieure à 80%. Elle ne l'est pas en regard de l'autre borne (D2), uniquement parce que la taille de l'effet détectable excède de 0.01 mg/dL la limite supérieure de la borne. Pour ce cas, qui n'est pas isolé dans l'ensemble de l'analyse, il est nécessaire de comparer, d'une part l'écart à la borne D2 qui est à l'origine du manque de puissance et d'autre part, la taille d'effet détectable par rapport aux variations naturelles de ce

paramètre chez les animaux témoins. Dans le cas présent, l'écart de 0.01 mg/dL est sans signification toxicologique.

Dans ce cas également, le poids des reins a une puissance supérieure à 80% et il n'y a pas d'altérations histologiques.

Le toxicologue n'a donc pas de réserve sur la taille de l'effet détectable des paramètres biochimiques sanguins en regard des puissances.

Biochimie urinaire

Pour ces paramètres, la puissance est généralement en dessous des 80% versus SD et 70% versus D2.

Compte tenu des éléments considérés dans les analyses précédentes, les tailles d'effets détectables apparaissent suffisantes pour éviter toute conclusion erronée.

Parmi d'autres, l'exemple du volume urinaire et de la densité illustrent cette assertion (tableau 4).

Tableau 4 : Taille d'effet détectable comparé aux bornes de tolérance (SD, D2) du paramètre de volume et de densité urinaires.

Paramètre (mL/16 heures)	Taille de l'effet détectable	SD historique	D2 variétés commerciales	Mini-Maxi historique
Volume urinaire	3.3 à 4.8	2.0 à 6.0	2.9 à 3.4	1.0 à 21.0*
Densité	0.01 à 0.02	0.01°	0.008 à 0.015	1.01 à 1.06*

* Après surcharge hydrique de 10mL/kg pc

° SD historique de 0.014 via une autre source (WIL control data, female rats from 565 animaux)

La taille de l'effet détectable s'inscrit à l'intérieur des SD historiques. Elle est supérieure au D2 des variétés commerciales. L'ensemble de ces données n'est toutefois pas de nature à suggérer une réserve quant à la puissance du test, sur la base de l'effet détectable.

De plus, c'est la prise en compte d'un faisceau convergent d'événements pouvant laisser supposer une toxicité rénale, qui est à la base du jugement du toxicologue.

Poids des organes

Les mesures effectuées sur les principaux organes correspondent à des puissances égales ou supérieures à 80% à l'exception de la rate chez les femelles (avec 78% de puissance).

S'agissant d'un paramètre dont les modifications, en relation avec les paramètres hématologiques et biochimiques qui lui sont liés, peuvent résulter d'une toxicité ciblée, la puissance satisfaisante des tests apparaît être un élément favorable pour étayer le jugement du toxicologue.



Agence nationale de sécurité sanitaire
de l'alimentation, de l'environnement et du travail
27-31 avenue du général Leclerc
94701 Maisons-Alfort Cedex
www.anses.fr