Section 4
Effets sur la santé

Essai n° 488:
Essais de mutations génétiques des cellules somatiques et germinales de rongeurs transgéniques



25 juin 2025

Lignes directrices de l'OCDE pour les essais de produits chimiques

OCDE
DES POLITIQUES MEILLEURES POUR UNE VIE MEILLEURE

OECD/OCDE 488

Adoptée :
25 juin 2025

LIGNE DIRECTRICE DE L'OCDE POUR LES ESSAIS DE PRODUITS CHIMIQUES

Essais de mutations génétiques des cellules somatiques et germinales de rongeurs transgéniques

1. INTRODUCTION

1. Les Lignes directrices (LD) de l'OCDE pour les essais de produits chimiques sont réexaminées périodiquement à la lumière des progrès scientifiques, de l'évolution des besoins réglementaires et des considérations relatives au bien-être animal. La Ligne directrice n° 488 a été initialement adoptée en 2011, puis révisée en 2013, 2020 et 2022. Elle a évolué pour intégrer l'analyse des mutations dans les tissus somatiques et les cellules germinales. Afin d'assurer l'harmonisation avec les autres Lignes directrices de l'OCDE pour les essais de génotoxicité, la révision de 2022 comprenait trois éléments de même importance pour évaluer si une réponse est positive ou négative. Tout d'abord, la réponse du produit chimique testé était évaluée afin de déterminer s'il y avait une augmentation statistiquement significative par rapport au témoin négatif concomitant. Deuxièmement, la réponse était analysée pour évaluer un effet lié à la concentration/dose. Troisièmement, les distributions des témoins négatifs historiques étaient utilisées pour évaluer la pertinence biologique. Comme mentionné plus haut, ces trois considérations étaient traitées avec une importance égale. Toutefois, des analyses récentes suggèrent que la troisième considération (c'est à-dire la comparaison avec les témoins historiques) devrait être nuancée et assouplie. En particulier, les distributions ne peuvent servir d'indicateur supplétif de la variation biologique normale que lorsqu'elles sont d'une qualité suffisante pour servir à cet effet. Les modifications présentées ici répondent à ces questions en révisant le texte de la section « Évaluation et interprétation des résultats » et en mettant à jour l'annexe associée.

2. Un document contenant un aperçu à la fois des essais de toxicologie génétique et des récents changements qui ont été apportés à ces Lignes directrices a été développé (1). Des informations complémentaires relatives aux principaux changements apportés à ces Lignes directrices ont aussi été publiées (2).

3. Les LD de l'OCDE pour les essais s'appliquent à toute une série d'essais de mutations in vitro permettant de détecter les mutations chromosomiques et/ou génétiques. Les LD s'appliquent aussi à plusieurs effets mesurés de génotoxicité *in vivo* (à savoir les aberrations chromosomiques, la formation de micronoyaux, la synthèse d'ADN non programmée et la rupture de brins d'ADN) mais elles ne mesurent pas les mutations génétiques. Alors que le test des comètes et le test de synthèse d'ADN non programmée sont des indicateurs permettant la détection de lésions pré-mutagènes, et que l'essai du *Pig-a* est limité au système hématopoïétique, les essais de mutations génétiques chez les rongeurs transgéniques (RTG)

© OCDE (2025)

L'OCDE autorise l'utilisation de ce contenu aux conditions décrites sur le site : http://www.oecd.org/fr/conditionsdutilisation.

répondent à ce besoin d'essais *in vivo* qui soient à la fois pratiques et exploitables à grande échelle, pour mesurer des mutations génétiques dans n'importe quel tissu.

4. Les données des essais de mutations génétiques chez les RTG ont fait l'objet d'un examen complet, ex. (3) (4) (5). Les modèles couverts par la présente Ligne directrice utilisent des rats et des souris transgéniques qui comportent de multiples copies de vecteurs navettes de phages intégrés dans leurs chromosomes. Les transgènes contiennent des gènes rapporteurs pour la détection de différents types de mutations induites *in vivo* par les produits chimiques testés. L'objectif de l'essai de mutations génétiques chez les RTG est d'identifier les substances qui induisent des mutations dues à des dommages à l'ADN dans le tissu qui est étudié.

5. On détecte les mutations qui se produisent chez un rongeur en récupérant le transgène et en analysant le phénotype du gène rapporteur chez un hôte bactérien démuni du gène rapporteur. Les essais de mutations chez les RTG permettent de mesurer les mutations induites dans des gènes génétiquement neutres présents dans la quasi-totalité des tissus du rongeur. Ces essais permettent donc de surmonter un certain nombre d'obstacles qui entravent actuellement l'étude de la mutation génétique *in vivo* dans des gènes endogènes (quantité limitée de tissus appropriés pour l'analyse, sélection positive/négative par rapport aux mutations, par exemple).

6. Les données disponibles démontrent que les transgènes répondent aux mutagènes de manière approximativement similaire aux gènes endogènes, notamment en ce qui concerne la détection de substitutions d'une base par une autre, les décalages du cadre de lecture et les délétions et insertions (3).

7. Les ateliers internationaux sur les essais de génotoxicité (IWGT) ont approuvé l'utilisation des essais pour la détection *in vivo* des mutations génétiques chez les RTG et ont recommandé un protocole de mise en application de ces essais (6) (7). Le document référencé (5) donne une analyse plus détaillée des avantages induits par l'emploi de ce protocole (8). La présente LD est basée sur ces recommandations pour l'évaluation des mutations génétiques dans les tissus somatiques et inclut des recommandations mises à jour pour l'évaluation des mutations génétiques dans les celles germinales males.

8. L'essai de mutations génétiques chez les RTG utilise un régime de traitement identique à celui de l'étude de toxicité à doses répétées (LD 407), c'est-à-dire une administration de 28 jours, ce qui offre l'option de combiner les deux essais en une seule étude, à condition que le fait de réaliser l'autopsie le jour suivant la fin du traitement pour les deux essais n'impacte pas de manière défavorable la récupération des mutations. Des données sont aussi requises pour s'assurer que la performance de l'étude de toxicité à doses répétées n'est pas affectée par l'utilisation d'une souche transgénique à la place de souches de rongeurs traditionnelles. De plus, il est possible d'intégrer à l'essai RTG d'autres effets mesurés de génotoxicité, tels que l'évaluation des micronoyaux et des mutations *Pig-a* (9). Le fait de combiner les études doit être basé sur le besoin d'investigation d'effets mesurés spécifiques, sur la base d'informations existantes ou d'exigences réglementaires spécifiques.

9. Les définitions des termes clés figurent en annexe 1.

2. CONSIDÉRATIONS INITIALES

10. Les essais de mutations chez les RTG pour lesquels il existe un nombre de données suffisant pour permettre de donner un avis favorable à leur utilisation dans le cadre de la présente LD sont les suivants : souris bactériophage *lacZ* (MutaMouse) ; souris et rat *gpt* delta (*gpt* et Spi$^-$*)* ; souris et rat bactériophage *lacI* (Big Blue®), mis en œuvre dans des conditions non sélectives. Dans ces essais, les mutations sont mesurées dans des gènes bactériens (*lacI*, *lacZ* et *gpt)*, insérés dans un vecteur lambda. De plus, les mutations peuvent être mesurées dans le gène *cII* du bactériophage dans les modèles Big Blue® et MutaMouse et les gènes *red/gam* dans le modèle delta *gpt* sous séléction Spi$^-$. Le modèle de plasmide *lacZ* a été supprimé de cette Ligne directrice parce qu'il n'est plus utilisé en routine. Des

© OCDE (2025)

méthodes pour l'identification des mutants sous conditions sélectives sont disponibles (voir paragraphes 17) et doivent être utilisées préférentiellement. Dans les modèles de RTG, le paramètre utilisé pour évaluer la mutagenèse est normalement la fréquence des mutants ; toutefois, une analyse moléculaire des mutations peut, si nécessaire, fournir des informations supplémentaires (voir paragraphes 57-58).

11. Ces essais de mutations génétiques chez les RTG (3) sont particulièrement pertinents pour évaluer le risque mutagène dans la mesure où les réponses qu'ils apportent dépendent du métabolisme *in vivo*, de la pharmacocinétique, des processus de réparation de l'ADN et de la synthèse d'ADN de translésion, bien que ces paramètres varient selon les espèces, selon les tissus et selon les types de dommages subis par l'ADN. Un essai *in vivo* de mutations génétiques sert à analyser plus en détail un effet mutagène détecté par un système *in vitro* et à étudier le mode d'action sous-jacent d'autres essais *in vivo*, tel qu'un résultat positif dans une étude de cancérogénèse. Outre le fait qu'elles soient mises en cause dans l'apparition de cancers, les mutations génétiques constituent un indicateur pertinent pour prédire les pathologies non cancéreuses des tissus somatiques dues à des facteurs mutagènes (10) (11) ainsi que les pathologies transmises par le biais des cellules germinales (12).

12. Si des éléments prouvent que le produit chimique testé ou un métabolite pertinent n'atteindra aucun des tissus d'intérêt, il n'est pas opportun d'avoir recours à l'essai de mutations génétiques chez les RTG.

13. Avant d'utiliser la Ligne directrice sur un mélange pour générer des données avec pour objectif recherché l'application réglementaire, on considérera si, et si oui pourquoi, elle peut fournir des résultats adéquats dans cet objectif. De telles considérations ne sont pas nécessaires quand les exigences réglementaires stipulent que le mélange doit être testé.

3. PRINCIPE DE LA MÉTHODE D'ESSAI

14. Dans les essais décrits au paragraphe 10, le gène cible est d'origine bactérienne ou bactériophagique, et sa récupération à partir de l'ADN génomique du rongeur se fait par introduction du transgène dans un vecteur de bactériophages lambda. Cette procédure implique l'extraction de l'ADN génomique du tissu d'intérêt du rongeur, le traitement *in vitro* de l'ADN génomique (à savoir l'encapsidation des vecteurs lambda) et la détection ultérieure des mutations subies par les hôtes bactériens dans des conditions appropriées. Les essais emploient des transgènes neutres non-transcrits, faciles à retrouver à partir de la plupart des tissus.

15. L'essai de base de mutations chez RTG implique d'administrer un produit chimique au rongeur pendant un certain temps. Toute voie d'administration appropriée est autorisée, y compris l'implantation (essais de dispositifs médicaux, par exemple). La période totale de traitement d'un animal est appelée la période d'administration. L'administration est généralement suivie d'une période, antérieure à la mise à mort, au cours de laquelle le produit chimique testé n'est pas administré et au cours de laquelle les lésions d'ADN non réparées se fixent sous forme de mutations stables. Dans la littérature, cette période est désignée sous diverses appellations telles que la période de manifestation, la période de fixation ou la période d'expression ; la fin de cette période marque le moment de l'échantillonnage (6) (8). Après une mise à mort de l'animal dans des conditions humaines, les tissus sont rapidement collectés et congelés, après quoi ils peuvent être conservés à une température inférieure ou égale à - 70°C +/- 5°C jusqu'à ce que l'ADN génomique soit isolé du/des tissus(s) d'intérêt puis purifié. Les tissus peuvent être collectés à partir d'animaux moribonds tués dans des conditions humaines au cours de la dernière semaine de traitement, puis conservés à une température inférieure ou égale à - 70°C +/- 5°C ; leur analyse sera conduite au cas par cas, si nécessaire.

16. Les données relatives à un seul tissu par animal, issues d'encapsidations/de ligatures multiples sont généralement agrégées, et la fréquence des mutants est généralement évaluée par le biais d'un

© OCDE (2025)

nombre total d'unités formant plage ou d'unités formant colonie par animal compris entre 10^5 et 10^7. En cas de recours à des méthodes de sélection positive, le nombre total d'unités formant plage ou formant colonie est déterminé à l'aide d'un ensemble séparé de plages non sélectives.

17. Les méthodes de sélection positive ont été développées pour que l'on puisse détecter plus facilement les mutations du gène *gpt* [souris et rat *gpt* delta, phénotype *gpt*⁻ (13) (14) (15)] et du gène *lacZ* [MutaMouse (16) (17) (18)] ; tandis qu'il n'y a pas de méthode de sélection positive disponible pour les mutations du gène *lacI* chez les animaux Big Blue® chez lesquels les mutations sont détectées par le biais d'une méthode non sélective qui permet d'identifier les mutants en générant des plages de couleur (bleue). Les méthodes de sélection positive permettent également de détecter les mutations du gène *cII* du bactériophage lambda utilisé comme vecteur navette [souris ou rat Big Blue®, MutaMouse) (19)] et les délétions sur les gènes lambda *red* et *gam* [sélection Spi⁻ chez souris et rat *gpt* delta) (14) (15) (20)]. On obtient la fréquence des mutants en divisant le nombre de plages contenant les mutations qui se produisent dans le transgène par le nombre total de plages retrouvés dans un même échantillon d'ADN. Dans les études de mutations chez les RTG, la fréquence des mutants est le paramètre observé. De plus, une fréquence de mutation peut être déterminée sous forme de fraction de cellules portant des mutations indépendantes ; ce calcul nécessite de corriger l'expansion clonale par séquençage des mutants retrouvés (paragraphes 57-58).

18. Les mutations notées dans les essais de mutations *lacI, lacZ, cII* et *gpt* consistent essentiellement en substitutions d'une base par une autre, en décalages du cadre de lecture et en petites insertions/délétions, qui ne peuvent être distinguées dans l'essai phénotypique. La part relative de ces types de mutations dans les mutations spontanées est similaire à celle observée pour le gène endogène *Hprt*. On observe des délétions importantes uniquement avec les essais de sélection du gène Spi⁻ dans le *gpt* delta (3). Les mutations qui nous préoccupent sont les mutations *in vivo* chez la souris ou le rat. Les mutations *in vitro* et *ex vivo*, qui peuvent se produire lors de la récupération des phages, de la réplication ou de la réparation sont relativement rares et peuvent, dans certains systèmes, être spécifiquement identifiées ou exclues par l'hôte bactérien ou le système de sélection positive (3) (4).

4. DESCRIPTION DE LA MÉTHODE D'ESSAI

4.1 Préparations

4.1.1 Sélection de l'espèce animale

19. On dispose actuellement de modèles de détection des mutations génétiques chez les souris et chez les rats transgéniques. Les modèles de souris et de rats sont tous deux acceptables. La justification de l'utilisation du modèle doit considérer : (i) la compétence du laboratoire avec le modèle ; (ii) la disponibilité de données historiques dans les tissus investigués ; (iii) les différences toxicologiques connues entre les espèces pour la substance étudiée (par exemple, si l'on étudie le mécanisme de la cancérogenèse d'une tumeur constatée uniquement chez une espèce de rongeur, pour établir une corrélation avec une étude de toxicité chez une espèce spécifique ou si l'on sait que le métabolisme chez une espèce de rongeur est plus représentatif du métabolisme humain) ; et (iv) l'espèce préférée utilisée dans d'autres études toxicologiques dans le cas où la combinaison avec l'essai RTG est envisagée.

4.1.2. Conditions d'encagement et d'alimentation

20. Toutes les procédures de laboratoire se conformeront aux normes locales en vigueur en matière de protection animale. Pour les rongeurs, la température de la salle d'expérimentation animale est idéalement de 22°C (±3°C). Bien que l'humidité relative doivent être d'au moins 30 % sans dépasser 70 %, en dehors des phases de nettoyage de la salle, l'objectif est de 50-60 %. La lumière est artificielle, avec

© OCDE (2025)

une séquence quotidienne de 12 heures de lumière suivies de 12 heures d'obscurité. Les aliments classiques pour animaux de laboratoire peuvent être utilisés et l'eau potable est fournie à satiété. Le choix des aliments peut être influencé par la nécessité d'assurer une bonne incorporation du produit chimique dans la nourriture si l'administration se fait par cette voie. Les animaux sont regroupés en petits groupes du même sexe ; ils peuvent être placés séparément si cela est justifié sur le plan scientifique. Pour les cages en groupe, chaque cage ne doit pas accueillir plus de cinq animaux. En outre, les cages doivent être conformes aux normes de bien-être animal [*par exemple*, Directive 2010/63/UE]. Selon les recommandations du comité compétent pour le soin et l'utilisation des animaux (ou de recommandations équivalentes), les cages doivent être à fond plein à chaque fois que c'est possible et dotées d'un enrichissement environnemental approprié.

4.1.3. Préparation des animaux

21. De jeunes adultes en bonne santé et sexuellement matures (âgés de 8-12 semaines au début du traitement) sont utilisés quand des données sur les cellules germinales sont requises (paragraphe 33). Pour les études portant sur les tissus somatiques, des animaux plus jeunes (par exemple 4-6 semaines au début du traitement) sont acceptables si cela est justifié du point de vue scientifique ou du bien-être animal. Les animaux sont répartis au hasard dans les groupes témoins et les groupes de traitement. Les animaux sont identifiés individuellement selon une méthode sans cruauté, la moins invasive possible (par exemple, bagage, étiquetage, pose d'une puce électronique ou identification biométrique, en évitant l'entaillage des oreilles ou la phalangectomie). Les animaux sont gardés dans leurs cages pendant au moins cinq jours avant le commencement de l'étude, de façon à les acclimater aux conditions du laboratoire. Au début de l'étude, il convient que la variation pondérale des animaux soit minimale et ne dépasse pas ± 20 % du poids moyen de chaque sexe. La sélection du sexe à utiliser dépend de l'exigence ou non de données sur les cellules germinales (paragraphe 33) et/ou de la spécificité de l'exposition humaine à la substance chimique pour un sexe en particulier (paragraphe 34).

4.1.4. Préparation des doses

22. Les produits chimiques testés solides sont dissous, mis en suspension dans des solvants ou véhicules appropriés (voir paragraphe 23) ou incorporés dans les aliments ou dans l'eau de boisson avant d'être administrés aux animaux. Les produits chimiques testés liquides peuvent être administrés directement ou dilués avant d'être administrés. En cas d'exposition par inhalation, les produits chimiques testés peuvent être administrés sous forme de gaz, de vapeur ou d'aérosol solide ou liquide en fonction de leurs propriétés physico-chimiques. D'autres voies d'exposition doivent être scientifiquement justifiées. Il faut utiliser des préparations fraîches sauf si l'on dispose de données qui démontrent la stabilité des préparations dans les conditions du stockage.

4.2. Conditions expérimentales

4.2.1. Solvant/véhicule

23. Il convient que le solvant/véhicule n'ait pas d'effets toxiques aux volumes de dose administrés et ne soit pas suspecté de réaction chimique avec le produit chimique testé. Il est recommandé d'envisager en premier lieu l'utilisation de solvants/véhicules aqueux chaque fois que c'est possible. Parmi les exemples de solvants/véhicules compatibles couramment utilisés figurent notamment l'eau, le sérum physiologique, les solutions de méthylcellulose, les solutions de carboxyméthylcellulose sodique, l'huile d'olive et l'huile de maïs (21). Le recours à des solvants/véhicules inhabituels fait l'objet de justification par des données de référence faisant état de leur compatibilité. En l'absence de données historiques ou

© OCDE (2025)

publiées significatives montrant qu'aucune mutation ou autre effet délétère n'est induit par un solvant/véhicule inhabituel sélectionné, une étude initiale devra être réalisée afin d'établir l'acceptabilité du témoin pour le solvant/véhicule.

4.2.2. Témoins positifs

24. Des animaux témoins positifs sont normalement inclus simultanément dans l'essai. Cette étape peut être évitée une fois que le laboratoire d'essai a apporté la preuve de ses compétences pour la conduite de l'essai et établi la plage des témoins historiques pour le tissu examiné (voir paragraphes 28-32). Dans cette situation, pour assurer la continuité des compétences dans la détection de l'augmentation de la fréquence des mutants, les laboratoires doivent de temps en temps (au moins une fois/an) effectuer des tests supplémentaires sur des tissus provenant d'animaux traités par des mutagènes, tel que décrit au paragraphe 27. Lorsqu'aucun groupe témoin positif n'est utilisé simultanément, les tissus d'animaux témoins positifs traités précédemment doivent être pris en compte dans chaque étude pour confirmer la fiabilité de la méthode. Ces échantillons doivent provenir de la même espèce, d'un âge similaire et des mêmes tissus d'intérêt, être stockés dans de bonnes conditions (voir paragraphe 53) et générer des fréquences de mutation qui soient consistantes avec les expériences précédentes.

25. Quand des témoins positifs sont inclus simultanément dans l'essai, il n'est pas nécessaire de les administrer par la même voie ou selon la même durée que celles du produit chimique testé ; toutefois, il convient d'avoir la certitude que les témoins positifs induisent des mutations dans un ou plusieurs tissu(s) d'intérêt pour le produit chimique testé. Les substances chimiques utilisées comme témoins positifs doivent produire, de façon fiable, un accroissement détectable de la fréquence des mutations par rapport au niveau spontané. Les doses des substances testées utilisés comme témoins positifs sont sélectionnées de manière à produire des effets faibles ou modérés qui permettent d'évaluer de manière critique les performances et la sensibilité de l'essai. Des exemples de substances utilisées comme témoins positifs et de certains de leurs tissus cibles figurent dans le tableau 1. Des substances autres que celles données dans le tableau 1 peuvent être sélectionnées si cela est justifié sur le plan scientifique.

© OCDE (2025)

Tableau 1. Exemples de substances utilisées comme témoins positifs et de certains de leurs tissus cibles

Substance et n° CAS	Caractéristiques	Tissus/types cellulaires cibles de la mutation	
		Rat	Souris
N-nitroso-N-éthylurée [n° CAS : 759-73-9]	Mutagène à action directe	Foie, estomac glandulaire, duodénum, jéjunum, moelle osseuse, rate, poumon, épithélium nasal, rein, vessie, cellules germinales testiculaires	Foie, secteur gastrique antérieur, estomac glandulaire, duodénum, colon, moelle osseuse, rate, poumon, épithélium nasal, rein, cellules folliculaires de la granulosa, cellules germinales testiculaires, sperme
Carbamate d'éthyle (uréthane) [n° CAS : 51-79-6]	Mutagène, nécessite d'être métabolisé mais ne produit que des effets faibles		Foie, moelle osseuse, rate, secteur gastrique antérieur, intestin grêle, poumon
Toluène-2,4-diamine [n° CAS : 95-80-7]	Mutagène, nécessite d'être métabolisé, positif également dans l'essai Spi$^-$	Foie	Foie
Benzo[a]pyrène [n° CAS : 50-32-8]	Mutagène, nécessite d'être métabolisé	Foie, secteur gastrique glandulaire, duodénum, jéjunum, moelle osseuse, rate, poumon, épithélium nasal, rein, vessie, épiploon	Foie, secteur gastrique antérieur, estomac glandulaire, duodénum, jéjunum, colon, moelle osseuse, glandes mammaires, cœur, poumon, rein, vessie, cellules germinales testiculaires, sperme

4.2.3. Témoins négatifs

26. Des témoins négatifs (voir paragraphe 28) traités de manière identique aux groupes traités, sont inclus pour chaque tissu et moment d'échantillonnage (cependant, voir le paragraphe 23 concernant les solvants ou véhicules inhabituels. En l'absence de données significatives observées ou publiées montrant que le solvant/véhicule choisi n'induit pas d'effets délétères ou mutagènes, une étude initiale est conduite afin d'établir afin d'établir l'acceptabilité du témoin contenant le solvant/véhicule.

5. VÉRIFICATION DES COMPÉTENCES DU LABORATOIRE

5.1. Vérification des compétences

27. Afin d'établir qu'il possède une expérience suffisante pour mener à bien l'essai avant de l'utiliser en routine, le laboratoire doit avoir démontré sa capacité à reproduire les résultats attendus d'après les données publiées (3) (22) concernant à la fois les fréquences de mutants et la récupération des transgènes de l'ADN génomique (par exemple efficacité de l'encapsidation). On utilisera un minimum de deux produits

© OCDE (2025)

chimiques témoins positifs (y compris des réponses faibles induites par des doses faibles de témoins positifs) tels que ceux énumérés au tableau 1 (voir paragraphe 25), et avec des témoins de véhicule/solvant compatibles (voir paragraphe 23). Au début, la compétence du laboratoire doit être démontrée dans deux tissus au moins, de préférence un pour les tissus à division lente tel que le foie, et un tissu à division rapide tels que la moelle osseuse, le secteur gastrique glandulaire ou le duodénum (7) (23). Si des évaluations de cellules germinales sont prévues, elles feront aussi partie des investigations relatives à la vérification des compétences du laboratoire. Les doses utilisées dans le cadre de ces expériences doivent produire des augmentations reproductibles qui sont fonction de la dose administrée, et démontrer la sensibilité et la plage dynamique du système d'essai sur le tissu en question. Cette exigence ne s'applique pas aux laboratoires possédant déjà une expérience, c'est-à-dire qui disposent d'une base de données historiques telle que définie aux paragraphes 28-32. Préalablement à la réalisation d'une étude dans un tissu qui n'aura pas été examiné auparavant, un laboratoire (y compris un laboratoire expérimenté) devra établir sa compétence dans les techniques d'extraction d'ADN et de récupération des transgènes spécifiques du tissu étudié, afin d'établir des fréquences de mutation et d'efficacité de l'encapsidation probables.

5.2. Données des contrôles historiques

28. Au cours des épreuves de compétence, le laboratoire doit établir pour chaque tissu évalué :

- une plage et une distribution des témoins positifs historiques, et

- une plage et une distribution des témoins négatifs historiques (*c'est-à-dire* animaux traités avec le solvant/véhicule ou animaux non traités).

29. Lors de l'acquisition initiale de données en vue d'établir une distribution des témoins négatifs historiques, les données des témoins négatifs concomitants doivent être cohérentes avec les données publiées, lorsqu'elles existent. Puis, à mesure que de nouvelles données expérimentales viennent étoffer la plage de distribution des témoins historiques, les données des témoins négatifs concomitants doivent idéalement se situer entre les seuils inférieurs et supérieurs de cette distribution (voir paragraphe 32). La base des données historiques du laboratoire, relative aux témoins négatifs doit être compilée, analysée et régulièrement mise à jour selon les recommandations de la littérature (par exemple (23) (58) ; voir aussi annexe 2). Cela comprend : la prise en compte du nombre de données minimum pour établir une distribution robuste (un minimum de 30 animaux est souhaitable) ; la fréquence de mise à jour et des méthodes permettant de s'assurer que les données les plus récentes et les plus pertinentes sont utilisées pour l'acceptation de l'essai et l'évaluation des données (voir paragraphe 61). Des déviations significatives de ces recommandations doivent être justifiées. Les laboratoires doivent avoir recours à des méthodes de contrôle de la qualité telles que des graphiques statistiques qui sont adaptés pour la distribution des données [par exemple cartes C ou cartes X-barre (24) (25) (26) (27)], et pas à de simples fourchettes de distribution des données de contrôle, pour identifier la variabilité des données et montrer que la méthodologie est « sous contrôle » dans leur laboratoire.

30. Si, au cours des expériences visant à vérifier sa compétence (comme décrit au paragraphe 29), le laboratoire n'a pas réalisé un nombre suffisant d'expériences pour établir une distribution des témoins négatifs statistiquement robuste (voir paragraphe 27), on peut accepter que la distribution soit construite au cours des premiers tests de routine. Cette approche devra suivre les recommandations établies dans la littérature [(24); annexe 2)] et les résultats des témoins négatifs obtenus lors de ces expériences devront être cohérents avec les données publiées des témoins négatifs.

31. Toute modification du protocole expérimental doit être étudiée à la lumière de ses répercussions sur la cohérence des nouvelles données avec celles de la base de données des témoins historiques.

© OCDE (2025)

Seules des incohérences majeures justifient l'établissement d'une nouvelle base de données des témoins historiques, après confirmation de la différence de distribution des données par un jugement scientifique solide (voir paragraphe 29) (24) (25) (26) (27). Lors de la constitution de cette nouvelle base de données, le laboratoire n'a pas forcément besoin d'une base de données de témoins négatifs complète pour autoriser la conduite d'un essai, à condition qu'il puisse apporter la preuve que les valeurs des témoins négatifs concomitants sont cohérentes avec la précédente base de données ou avec les données publiées correspondantes.

32. Les données des témoins négatifs désignent la fréquence des mutations par tissu chez chaque animal. Les témoins négatifs concomitants se situent normalement entre les seuils inférieurs et supérieurs de la distribution des données historiques des témoins négatifs contenues dans la base de données du laboratoire. Lorsque les données des témoins négatifs concomitants se situent en dehors de ces limites de contrôle, leur inclusion dans la distribution des témoins historiques peut être acceptable, à condition que ces données ne soient pas exagérément extrêmes (c'est-à-dire identifiées comme tel par un test de valeurs aberrantes) et qu'il soit prouvé que le système d'essai est « sous contrôle » (voir paragraphe 29) et qu'il n'y a pas eu de défaillance technique ou humaine.

6. DEROULEMENT DE L'ESSAI

6.1. Nombre et sexe des animaux

33. Un nombre suffisamment important d'animaux est prévu afin d'obtenir l'efficacité statistique nécessaire pour détecter une fréquence de mutants au moins multipliée par deux. Chaque groupe contient au moins cinq animaux; toutefois, si l'efficacité statistique est insuffisante, on utilise un plus grand nombre d'animaux, selon les besoins. Quand la conception de l'étude requiert des données sur cellules germinales, des animaux mâles sont alors utilisés car il n'est pas possible de collecter un nombre suffisant de cellules germinales femelles pour conduire l'essai de mutations chez les RTG (28).

34. Quand seules des cellules somatiques sont nécessaires, ces études peuvent être conduites avec l'un ou l'autre sexe, car la réponse des mutations est similaire pour les animaux mâles et femelle. Si l'exposition humaine est spécifique pour un sexe, par exemple dans le cas de certains produits pharmaceutiques, l'essai sera pratiqué sur des animaux du sexe approprié. Des données (incluant par exemple des données d'une étude préliminaire de détermination des doses) démontrant des différences importantes entre mâles et femelles (par exemple différence de toxicité systémique, de métabolisme, de biodisponibilité, etc.) encouragent l'utilisation des deux sexes. Si un essai de mutations TGR est réalisé en tant que suivi d'essais mettant en évidence des tumeurs ou des résultats toxicologiques positifs, la sélection de l'espèce et du sexe doit se baser sur l'espèce et le sexe de l'étude initiale.

6.2. Période d'administration

35. Sachant que les mutations s'accumulent au fil des traitements, un régime de doses répétées est nécessaire, à raison de traitements quotidiens pendant 28 jours. Cette démarche est généralement jugée acceptable à la fois pour produire une accumulation suffisante de mutations par des mutagènes faibles et pour fournir une durée d'exposition adéquate pour la détection des mutations dans les organes à prolifération lente. D'autres régimes de traitement peuvent être adaptés pour certaines évaluations, et il convient que ces autres calendriers de dosage soient scientifiquement justifiés dans le protocole. Les traitements ne sont pas plus courts que le temps nécessaire à l'induction complète de l'ensemble des

© OCDE (2025)

enzymes impliquées dans les métabolismes mis en jeux, et les traitements de plus courte durée peuvent nécessiter le recours à des moments d'échantillonnage multiples, adaptés aux organes dont les taux de prolifération sont différents. Quel que soit le cas de figure, toutes les informations disponibles (par exemple celles sur la toxicité générale ou le métabolisme et la pharmacocinétique) sont exploitées lorsqu'il s'agit de justifier un protocole, en particulier lorsque ce dernier dévie des recommandations standard décrites précédemment. S'ils permettent certes d'accroître la sensibilité, les traitements de plus de 28 jours sont expliqués clairement et justifiés, car des traitements de longue durée peuvent engendrer une augmentation apparente de la fréquence de mutants par expansion clonale (7).

© OCDE (2025)

6.3. Niveaux de doses

36. Toutes données existantes, de toxicité ou de toxicocinétique sont prises en compte pour la détermination des niveaux de doses. Si une étude préliminaire de détermination des concentrations est conduite parce que les données adéquates déjà disponibles sont insuffisantes pour permettre de guider la sélection des doses, l'étude est alors réalisée par le même laboratoire, en utilisant les mêmes souche (des animaux non transgéniques peuvent être utilisés), sexe, voie d'exposition que ceux qui seront utilisés dans l'étude principale.

37. Dans l'essai principal, pour obtenir des informations sur la relation dose-réponse, une étude complète comporte un groupe témoin négatif (voir paragraphe 26) et au moins trois niveaux de doses du produit chimique testé, espacés comme il se doit, sauf si la dose limite a été utilisée (voir paragraphe 40). En dehors des situations où l'essai limite est applicable, la dose la plus élevée doit être choisie dans le but d'induire des effets toxiques, mais pas la mort ni de graves souffrances. Cette définition est tirée de la LD 407 et est utilisée ici pour faciliter l'intégration de l'essai chez les RTG à d'autres études à doses répétées, et ainsi maximiser les informations toxicologiques obtenues. Lorsque la dose la plus élevée est limitée par d'autres facteurs, comme l'appétence pour l'administration dans les aliments ou l'eau de boisson (et lorsque l'administration par gavage n'est pas possible) ou l'explosivité pour les produits chimiques d'essai administrés par inhalation, la dose la plus élevée sera la dose maximale possible. Par la suite, il convient de sélectionner une séquence de doses décroissantes en vue de mettre en évidence une éventuelle relation dose-réponse. Des intervalles de dose d'un facteur de de deux à quatre sont fréquemment utilisés pour déterminer les doses décroissantes lorsque les effets génotoxiques sont évalués. Avec la majeure partie des produits chimiques testés, les niveaux de doses employés sont compris entre « toxicité maximale » et « faible toxicité » voire « absence de toxicité ». Lorsque d'autres effets de toxicité sont intégrés dans l'étude, d'autres écarts entre les doses peuvent être envisagés. Il est souvent préférable d'ajouter un quatrième groupe d'essai plutôt que d'espacer excessivement les doses.

38. Les produits chimiques testés ayant une activité biologique spécifique à des niveaux de doses faibles et non toxiques (telles que les hormones et les mitogènes) et les substances qui peuvent présenter des propriétés de saturation toxicocinétique ou induire des processus de détoxification, pouvant entrainer une diminution de l'exposition après une administration à long terme peuvent être considérés comme des exceptions aux critères de détermination des doses et sont évalués au cas par cas.

39. Il convient de s'assurer que la dose la plus élevée qui a été identifiée dans l'étude préliminaire de détermination des concentrations n'entraine de toxicité excessive dans aucun tissu d'intérêt car cela pourrait empêcher d'obtenir un nombre de cellules suffisant pour extraire une quantité et une qualité d'ADN adéquates à la récupération des transgènes pour l'analyse des mutations. Si tel est le cas, on peut envisager l'inclusion d'un groupe de dose supplémentaire, rapproché de la dose la plus élevée pour assurer la disponibilité des trois groupes de traitement requis pour l'analyse des mutations. Pour les études de gavage par voie orale, l'utilisation d'un autre véhicule ou un dosage fractionné (deux traitements ou plus dans la journée, séparés par 2-3 heures au maximum) peuvent être envisagés, avec justification, pour minimiser les effets entrainant une toxicité excessive.

6.4. Essai limite

40. Si les essais préliminaires de détermination des concentrations ou les données disponibles issues de souches de rongeurs apparentées indiquent qu'un régime de traitement égal ou supérieur à la dose limite (voir ci-dessous) n'engendre pas d'effets toxiques observables, et si la génotoxicité n'est pas escomptée sur la base des études de génotoxicité *in vitro* ou sur la base des données relatives aux

© OCDE (2025)

substances structurellement apparentées, une étude complète utilisant trois niveaux de doses peut ne pas être considérée comme nécessaire. Dans ce cas, un seul niveau de dose (égal à la dose limite), peut s'avérer suffisant. Pour une période d'administration de 28 jours (soit 28 traitements quotidiens), cette dose limite est de 1000 mg/kg de poids corporel/jour. Pour les périodes d'administration d'une durée inférieure ou égale à 14 jours, la dose limite est de 2000 mg/kg de poids corporel/jour (les calendriers de dosage autres que 28 traitements quotidiens font l'objet d'une justification scientifique dans le protocole ; voir paragraphe 35). En cas d'exposition par inhalation, les concentrations de l'essai limite sont respectivement de 20 mg/L, 5 mg/L ou 20 000 ppm pour les vapeurs, les poussières/brouillards (aérosols) et les gaz.

6.5. Administration des doses

41. La voie retenue doit permettre l'exposition du/des tissu(s) d'intérêt. La voie de préférence est la voie d'exposition humaine anticipée et d'autres voies d'exposition doivent être justifiées. Par conséquent, les voies d'exposition telles que voie alimentaire, eau de boisson, topique, sous-cutanée, intraveineuse, orale (par gavage), inhalation, intratrachéale ou implantation, peuvent être choisies si elles peuvent être justifiées. Les injections intrapéritonéales ne sont en général pas recommandées car la cavité péritonéale n'est pas une voie d'exposition humaine physiologiquement pertinente et cette voie n'est utilisée que si elle est justifiée scientifiquement. Le volume maximal de liquide que l'on peut administrer en une seule prise dépend de la taille de l'animal d'essai et de la voie d'administration et doit être guidé par les standards internationaux en matière de bien-être animal. Pour l'administration orale par gavage, le volume n'excède pas 1 ml/100 g de poids corporel pour les souris et les rats, sauf pour les solutions aqueuses, où un maximum de 2 ml/100 g est acceptable. Le recours à des volumes plus importants fait l'objet de justification. Sauf pour les produits chimiques testés irritants ou corrosifs, qui auront normalement des effets exacerbés à des concentrations plus élevées, la variabilité du volume d'essai est réduite au maximum par ajustement de la concentration pour garantir un volume constant par rapport au poids corporel, à tous les niveaux de doses.

6.6. Moment d'échantillonnage

6.6.1. Cellules somatiques

42. Le moment d'échantillonnage est une variable essentielle car elle dépend de la période nécessaire pour que les mutations se fixent. Cette période est différente selon les tissus et semble dépendre du temps de renouvellement de la population cellulaire (32), la moelle osseuse et les intestins répondant rapidement tandis que le foie répond beaucoup plus lentement. Le protocole recommandé pour l'évaluation des fréquences de mutants à la fois dans les tissus à prolifération rapide et dans ceux à prolifération lente après une période de 28 jours de traitements consécutifs (tel qu'indiqué au paragraphe 35) est un échantillonnage 28 jours après la dernière administration (c'est-à-dire 28j + 28j) (33). Un échantillonnage trois jours après le dernier traitement (c'est-à-dire 28j + 3j), tel qu'il était recommandé dans les versions précédentes de la Ligne directrice demeure un moment d'échantillonnage valable quand il n'est pas nécessaire d'obtenir des données sur les cellules germinales.

6.6.2. Cellules germinales

43. Les essais chez les RTG conviennent bien à l'étude de l'induction de mutations génétiques chez les cellules germinales mâles (5) (34) (35) (36) (37), pour lesquelles la chronologie et la cinétique de la spermatogenèse ont été bien définies (38) (39) (40). Du fait du nombre limité d'ovules disponibles pour l'analyse, même après une super-ovulation, et parce qu'il n'y a pas de synthèse de l'ADN dans l'ovocyte,

© OCDE (2025)

les cellules germinales femelles ne peuvent pas être utilisées pour mesurer des mutations via les essais transgéniques (28). Les données disponibles sur la mutagénicité des cellules germinales obtenues via les essais chez les RTG (5) ainsi que la modélisation de la spermatogenèse chez la souris et le rat (40) ont récemment été examinées, l'objectif étant de choisir en connaissance de cause une méthode expérimentale adaptée pour l'évaluation de la mutagénicité des cellules germinales. En ce qui concerne la modélisation, on a considéré que la phase mitotique de la spermatogenèse (cellules souches, spermatogonies proliférantes et différenciantes) est la seule phase de la spermatogenèse où se produisent la réplication de l'ADN et la prolifération cellulaire, nécessaires pour fixer les mutations dans le transgène (41).

44. Les cellules germinales mâles peuvent être prélevées au stade de spermatozoïdes matures de la queue de l'épididyme ou de cellules germinales en développement des tubes séminifères. Il est possible de prélever les cellules germinales en développement des tubes séminifères simplement en les retirant de la tunique albuginée qui entoure les testicules, ou en les expulsant des tubes séminifères par séparation enzymatique ou physique (42). On privilégiera cette dernière approche, car elle enrichit la population collectée de cellules germinales étant donné que les cellules somatiques (cellules de Leydig et de Sertoli, par exemple) présentes dans les testicules ne peuvent pas être séparées facilement des tubes.

45. La chronologie de la spermatogenèse est bien établie chez la souris et le rat (39). La progression des cellules germinales en développement du stade de cellules souches spermatogoniales exposées au stade de sperme mature rejoignant la queue de l'épididyme dure 49 jours environ chez la souris (37) (39) et 70 jours environ chez le rat (39) (40). De ce fait, chez la souris comme chez le rat, le sperme présent dans la queue de l'épididyme au bout d'un régime de 28+3 jours représente une population de cellules germinales dont l'ADN ne s'est pas répliqué pendant l'exposition : son échantillonnage ne fournit pas de données significatives sur la mutagénicité et, par conséquent, ne devrait pas être effectué. En ce qui concerne la souris, il existe également des données expérimentales selon lesquelles ce régime de 28 + 3 jours ne permet pas de détecter les mutagènes forts des cellules germinales N-nitroso-N-éthylurée (5) et benzo[a]pyrène (43). L'échantillonnage du sperme caudal devrait être effectué 49 jours (souris) ou 70 jours (rat) minimum après la fin de la période d'administration de 28 jours si l'objectif est d'évaluer les mutations des cellules souches spermatogoniales (5) (40).

46. Les cellules germinales expulsées des tubes séminifères constituent une population mixte de spermatogonies, de spermatocytes et de spermatides (34) (35) (40). La composition de la population de cellules germinales prélevée dans les tubes séminifères de souris et de rats, en fonction du nombre de traitements quotidiens administrés pendant la phase de prolifération de la spermatogenèse, a été décrite en détail pour différents moments d'échantillonnage en tenant compte de la cinétique connue de la spermatogenèse (40). Si des résultats positifs obtenus au niveau des cellules germinales des tubes séminifères après un traitement de 28 + 3 jours sont instructifs, un résultat négatif après un traitement de 28 + 3 jours est insuffisant pour écarter la possibilité qu'un produit chimique ait des effets mutagènes sur les cellules germinales, car seule une fraction limitée des cellules germinales collectées à partir des tubules a reçu un traitement en continu pendant toute la période d'administration de 28 jours durant la phase de prolifération de la spermatogenèse (5) (40).

47. *Modélisation de la souris* - D'après notamment la modélisation détaillée de la spermatogenèse (41) et des données expérimentales limitées (5), un délai supérieur à 3 jours est préférable avant l'échantillonnage des cellules germinales des tubes séminifères pour l'évaluation de la mutagénicité des cellules germinales. La modélisation montre que le régime de 28 + 28 jours permet d'évaluer les mutations qui se produisent au sein d'une population de cellules germinales de souris auxquelles on a administré 99.6 % des 28 jours de traitement pendant la phase de prolifération de la spermatogenèse, contre 42.2 % seulement avec le régime de 28 + 3 jours (40). Le modèle de spermatogenèse présuppose que l'exposition n'induit pas d'apoptose des cellules germinales de manière significative, pas plus que de retard

© OCDE (2025)

dans la progression de la spermatogénèse. Cependant, si de tels effets devaient avoir lieu, un moment d'échantillonnage plus long, tel que celui obtenu par le régime de 28 + 28 jours, permettrait une reprise de la spermatogénèse en permettant aux testicules d'être repeuplés par les cellules souches survivantes et par des spermatogonies en cours de différentiation qui ont reçu la totalité des 28 jours d'administration du produit chimique testé durant la phase de prolifération de la spermatogénèse. Pour ces raisons, les résultats positifs et négatifs obtenus au niveau des cellules germinales de souris avec ce régime de 28 + 28 jours sont considérés concluants, et le régime de 28 + 28 jours est recommandé.

48. *Modélisation du rat* - D'après une modélisation poussée de la spermatogénèse (40) et considérant le fait que la spermatogenèse est plus longue chez le rat que chez la souris, le régime de 28 + 28 jours n'induit pas le même degré d'exposition des cellules en phase de prolifération chez des rats et des souris soumis au même régime (voir paragraphe 47). La modélisation de la spermatogenèse chez le rat indique que le régime de 28 + 28 jours permet d'évaluer les mutations qui se produisent au sein d'une population de cellules auxquelles on a administré 80.3 % des 28 jours de traitement pendant la phase de prolifération de la spermatogenèse (contre 99,6 % chez la souris), alors que seulement 21.6 % sont obtenus avec le régime de 28 + 3 jours (40). Bien qu'il ne soit pas optimal, le régime de 28 + 28 jours est néanmoins considéré adéquat pour l'évaluation de la mutagénicité des cellules germinales ; il permet en outre d'évaluer chez le même animal, à la fois les mutations des tissus somatiques et celles des cellules germinales présentes dans les tubes séminifères. Lors de l'évaluation des résultats négatifs obtenus avec ce régime, il faudra considérer le fait que les cellules germinales proliférantes ont reçu moins que l'exposition complète potentielle et en évaluer l'impact.

49. Des moments d'échantillonnage autres que 28 jours peuvent également être acceptables ; cependant, il faudra considérer l'impact de l'utilisation d'un moment d'échantillonnage plus court que 28 jours, et justifier scientifiquement son choix dans la mesure où cela réduit le degré d'exposition des cellules germinales durant leur phase de prolifération aussi bien chez la souris que chez le rat (40). Lorsqu'on disposera d'un nombre suffisant d'études pour déterminer les avantages de tout autre régime pour les cellules germinales, la présente Ligne directrice sera revue et, au besoin, corrigée à la lumière de l'expérience acquise.

50. En résumé, lorsque, sur la base des exigences réglementaires ou des informations toxicologiques, il faut collecter et/ou tester aussi bien les cellules somatiques que les cellules germinales, le régime de 28 + 28 jours permet de tester les mutations dans les tissus somatiques et dans les cellules germinales des tubes séminifères à partir du même animal.

6.7. Observations

51. Les animaux font l'objet d'un examen clinique général au moins une fois par jour, de préférence aux mêmes heures, en prenant en considération la période où les effets prévus devraient être les plus marqués après l'administration. L'état de santé des animaux est consigné. Au moins deux fois par jour, l'ensemble des animaux fait l'objet d'un constat de morbidité et de mortalité. Chaque animal est pesé au début de l'étude puis au moins une fois par semaine, et juste après avoir été mis à mort humainement. La consommation alimentaire est mesurée au moins une fois par semaine. Si le produit chimique testé est dilué dans de l'eau avant d'être administré, la consommation d'eau est mesurée à chaque changement d'eau et au moins une fois par semaine. Les animaux montrant des signes non-létaux de toxicité excessive sont euthanasiés avant la fin de l'essai (29).

© OCDE (2025)

6.8. Collecte des tissus

52. Les arguments invoqués pour justifier la collecte des tissus sont clairement définis. Étant donné qu'il est possible d'étudier l'induction de mutations quel que soit le tissu, ou presque, les tissus à collecter sont choisis en fonction des raisons qui ont motivé la conduite de l'étude et des données existantes relatives à la génotoxicité, à la cancérogénicité ou à la toxicité du produit chimique testé soumis à l'étude. La voie d'administration [basée sur la/les voie(s) d'exposition humaine potentielle(s)], l'exposition tissulaire prévue et l'organe cible possible sont des facteurs importants à prendre en considération. En l'absence de documentation générale, plusieurs tissus somatiques sont collectés, selon l'intérêt qu'ils représentent. Il s'agira de tissus à prolifération rapide, de tissus à prolifération lente et de tissus appartenant au site de contact. En outre, les cellules germinales en développement des tubes séminifères (tel que décrites aux paragraphes 44 et 46) sont collectées et stockées au cas où l'analyse ultérieure de la mutagénicité des cellules germinales est nécessaire et qu'un moment d'échantillonnage adapté a été utilisé. Les organes pertinents sont pesés, et pour les organes les plus gros, la même zone est collectée sur tous les animaux.

6.9. Stockage des tissus et de l'ADN

53. Les tissus (ou homogénats de tissus) sont stockés à une température inférieure ou égale à −70 ºC et utilisés pour l'isolation de l'ADN avant expiration d'un délai de 5 ans. L'ADN isolé, réfrigéré à une température de 4 ºC dans un tampon approprié est idéalement utilisé pour l'analyse des mutations avant expiration d'un délai d'un an.

6.10. Sélection des tissus pour l'analyse des mutants

54. Les tissus sont sélectionnés en fonction des critères suivants : (i) la voie d'administration ou le site de premier contact (estomac glandulaire ou duodénum en cas d'administration par voie orale, poumon ou épithélium nasal en cas d'exposition par inhalation ou peau en cas d'administration par voie topique, par exemple) ; (ii) les paramètres pharmacocinétiques ADME (absorption, distribution, métabolisme et excrétion) observés dans les études de toxicité générale, qui indiquent la distribution, la rétention ou l'accumulation dans les tissus, ou les organes cibles pour la toxicité ; et (iii) la nécessité ou non d'obtenir des données des cellules germinales. Si des études sont menées pour faire suite à des études de cancérogénicité, les tissus cibles pour la cancérogénicité sont étudiés. La sélection des tissus à analyser doit optimiser la détection de produits chimiques qui sont des mutagènes à action directe, rapidement métabolisés, hautement réactifs ou faiblement absorbés, ou pour lesquels le tissu cible est déterminé par la voie d'administration (44).

55. En l'absence de documentation générale, et si l'on tient compte du site de contact en fonction de la voie d'administration, la mutagénicité du foie et d'au moins un tissu à division rapide (estomac glandulaire ou duodénum ou moelle osseuse, par exemple) sont évaluées. Dans la plupart des cas, ces exigences peuvent être respectées par des analyses de deux tissus soigneusement sélectionnés, mais, dans certains cas, trois tissus voire plus sont nécessaires. Les cellules germinales peuvent aussi être évaluées (voir paragraphe 52).

© OCDE (2025)

6.11. Méthodes de mesure

56. Des méthodes standard de laboratoire ou publiées conçues pour la détection des mutants sont à disposition des modèles transgéniques recommandés : bactériophage lambda *lacZ* (18) ; souris *lacI* (45) (46) ; souris *gpt* delta (14) ; rat *gpt* delta (15) (47) ; *cII* (19). Les modifications font l'objet d'une justification et sont accompagnées d'une documentation adéquate. Les données issues de multiples encapsidations peuvent être agrégées et utilisées pour atteindre un nombre de plages ou de colonies adéquat. Toutefois, le fait qu'un nombre important de réactions d'encapsidation soit nécessaire pour atteindre le nombre de plages approprié peut être un indicateur d'un ADN de faible qualité. Dans ces cas-là, la prudence s'impose car les données peuvent être dépourvues de fiabilité. Le nombre optimal de plages ou de colonies par échantillon d'ADN dépend de la probabilité statistique de détecter des mutants en nombre suffisant à une fréquence de mutants spontanés donnée. En général, un minimum de 125 000 - 300 000 plages est nécessaire pour les modèles RTG avec une fréquence de fond de mutants de l'ordre de 3×10^{-5}. Les modèles tels que le *gpt* delta avec une fréquence de fond de mutants plus basse nécessitent proportionnellement l'observation de davantage d'unités formant colonie pour permettre une puissance statistique adéquate. Les tissus et les échantillons qui en découlent sont traités et analysés selon un schéma par bloc où un nombre, de préférence équivalent, d'échantillons du groupe témoin véhicule/solvant, groupe témoin positif (s'il y en a) ou ADN du témoin positif (s'il y a lieu) et chaque groupe de traitement sont traités ensemble.

6.12. Séquençage des mutants

57. L'amplification clonale de mutants d'apparition précoce et spontanée peut survenir chez n'importe quel animal, entrainant des augmentations plus ou moins importantes des fréquences de mutations dans les tissus individuels. Les tissus des animaux présentant des fréquences de mutants élevées, en dehors de la gamme de distribution historique et différentes de celles des autres animaux du groupe peut représenter une mutation jackpot ou événement clonal. Dans la mesure où de tels événements sont bien souvent localisés au sein d'un tissu, la ré-analyse d'une portion différente du même tissu peut être une approche pour évaluer ce type d'anomalies. Des animaux supplémentaires de remplacement sont souvent ajoutés à l'étude pour palier à la perte de certains animaux du fait d'une mort prématurée ou bien de mutations jackpot. L'analyse de ces animaux complémentaires peut être appropriée dans ces cas-là.

58. Pour les applications réglementaires, le séquençage de l'ADN des mutants n'est pas obligatoire, notamment en cas de résultat clairement positif ou négatif (voir paragraphes 62 et 63). Néanmoins, les données de séquençage peuvent être utiles si l'on observe de grandes variations entre les individus. Dans ces cas-là, le séquençage peut servir à exclure l'hypothèse de jackpots ou de clonages en permettant d'identifier la proportion de mutants uniques dans un tissu particulier. Le séquençage de 10 mutants environ par tissu et par animal suffit à déterminer si les mutants clonaux contribuent à la fréquence des mutants ; séquencer jusqu'à 25 mutants peut se révéler nécessaire pour corriger mathématiquement les aspects de clonalité dans la fréquence des mutants. Le séquençage des mutants peut également être envisagé si l'on constate que la fréquence des mutants est en légère augmentation (la fréquence est alors légèrement supérieure aux valeurs des témoins non traités). Des différences en matière de spectre de mutants entre les colonies de mutants des animaux traités et des animaux non traités peuvent appuyer l'hypothèse d'un effet mutagène (7). De plus, les spectres de mutation peuvent servir à développer des hypothèses mécanistiques. Si le séquençage fait partie du protocole d'essai, la conception de l'essai fait l'objet d'une attention particulière, notamment en ce qui concerne le nombre de mutants séquencés par échantillon pour obtenir l'efficacité adéquate selon le modèle statistique utilisé (voir paragraphe 67). À cet

© OCDE (2025)

égard, des méthodes de séquençage de nouvelle génération sont disponibles pour les gènes *cII* (48) et *lacZ* (49), ce qui facilite considérablement le séquençage d'un nombre important de mutants.

7. RÉSULTATS ET RAPPORT

7.1. Présentation des résultats

59. Les résultats individuels pour chaque animal sont présentés sous forme de tableaux. L'unité expérimentale est l'animal. Le rapport comporte le nombre total d'unités formant plage ou d'unités formant colonie, le nombre de mutants et la fréquence des mutants pour chaque tissu de chaque animal. Le rapport doit aussi indiquer le nombre de réactions d'encapsidation/de récupération et le nombre de réactions par échantillon d'ADN est consigné dans le rapport. S'il convient de conserver les données relatives à chaque réaction individuelle, seul le nombre total de mutants et d'unités formant plage/unités formant colonie est consigné par écrit. Les données de toxicité et les signes cliniques tels que décrits dans le paragraphe 51 sont consignés dans le rapport. Les résultats du séquençage sont présentés pour chaque mutant analysé, et les calculs de la fréquence de mutation qui en résultent pour chaque animal et chaque tissu sont montrés.

7.2. Évaluation statistique et interprétation des résultats

7.2.1. Critères d'acceptabilité

60. Les critères suivants déterminent l'acceptabilité de l'essai :

a. Les témoins négatifs concomitants sont considérés acceptables pour être inclus dans la base de données de témoins négatifs historiques du laboratoire (voir paragraphes 28-32; annexe 2).

b. Les témoins positifs concomitants, ou témoins d'évaluation doivent induire des réponses compatibles avec celles générées dans la base de données de témoins positifs historiques du laboratoire et produire une augmentation statistiquement significative comparée au témoin négatif concomitant (voir paragraphes 24 et 25).

c. Le nombre idoine de doses, animaux par dose et unités formant plage ou unités formant colonie est analysé (voir paragraphes 16, 33 et 37).

d. Les critères de sélection de la dose la plus élevée et la voie d'administration sont cohérents avec ceux décrits aux paragraphes 36-39.

7.2.2. Évaluation et interprétation des résultats

61. Les méthodes statistiques employées considèrent l'animal comme unité expérimentale. Des méthodes statistiques appropriées pouvant servir d'aide à l'évaluation des résultats d'essai sont disponibles dans les références suivantes (6) (50) (51) (52) (53) et en annexe 2. L'évaluation des réponses doit prendre en compte toutes les données et faire appel, dans tous les cas, à un jugement scientifique solide. Quand il y a des données disponibles pour au moins trois doses et un contrôle négatif (solvant/véhicule), une analyse de la relation dose–réponse est conduite en utilisant un test de tendance approprié.

© OCDE (2025)

62. Pour un tissu donné, à condition que tous les critères d'acceptabilité ci-dessus soient remplis, un produit chimique d'essai est considéré comme positif si les deux critères a et b ci-dessous sont réunis, dans l'une des conditions expérimentales examinées :

a) Dans l'un au moins des groupes traités, la fréquence de mutants présente une augmentation statistiquement significative par rapport au témoin solvant/véhicule concomitant;
b) Les réponses, en termes de fréquence des mutants, sont liées à la dose, par exemple lorsqu'elles sont évaluées au moyen d'un test de tendance approprié (annexe 2) (ne s'applique pas à l'essai limite) ;

Les analyses des critères de type a et b, qui reposent toutes deux sur des données relatives aux témoins négatifs concomitants, sont considérées comme les analyses statistiques de première importance pour l'interprétation des données de l'étude. Les données relatives aux témoins négatifs historiques et leur distribution peuvent également être utiles pour contextualiser les résultats de la présente étude. Toutefois, cette comparaison a moins de poids que les critères a ou b. En outre, l'utilité des données des témoins historiques à cette fin dépend fortement de la quantité et de la qualité des données sous-jacentes. Par exemple, les données des témoins historiques ne peuvent pas servir de comparateur utile (par exemple, comme mesure indirecte de la variation biologique normale), lorsque la variation inter-études représente la principale source de variation. Ce concept et d'autres relatifs à l'utilisation appropriée des données des témoins historiques sont présentés à l'annexe 2. Ainsi, ce n'est que dans la mesure où des données relatives aux témoins négatifs historiques ont été obtenues en quantité suffisante, et que ces données sont jugées de qualité suffisante, qu'outre les critères a et b, un troisième critère, c, peut être appliqué comme suit :

c) Au moins un des groupes traités présente une fréquence de mutants supérieure à la limite supérieure dérivée de la distribution des données des témoins négatifs historiques.

La limite supérieure dérivée de la distribution des témoins négatifs historiques est déterminée à l'aide d'une méthode adaptée à la quantité et à la qualité des données du laboratoire (par exemple, graphiques témoins, intervalles de tolérance ou quantiles), une justification étant fournie dans le rapport d'étude. La variabilité inter-laboratoires pouvant influer sur la distribution des témoins, aucune valeur numérique fixe ne peut être prescrite, et les laboratoires doivent définir leurs limites supérieures de manière scientifiquement fondée et transparente, comme décrit à l'annexe 2.

Dans de nombreux cas, les fréquences considérées seront les valeurs moyennes de l'étude en cours comparées à une distribution dérivée des valeurs moyennes des témoins négatifs historiques. Cependant, il existe des situations dans lesquelles les données relatives aux témoins négatifs historiques et les distributions associées basées sur des données individuelles relatives à chaque animal sont utiles ; par exemple, lorsque les valeurs des témoins historiques sont peu nombreuses, ou lorsqu'il est intéressant d'estimer les sources de variation. Comme expliqué plus en détail à l'annexe 4, il est important dans ces cas-là, de procéder à des comparaisons appropriées. En particulier, une distribution basée sur des données individuelles d'animaux ne doit pas être utilisée comme comparateur pour les moyennes des groupes de traitement de l'étude en cours. Une distribution basée sur des données individuelles relatives à chaque animal ne doit être utilisée qu'à titre de comparaison pour les données individuelles relatives à chaque animal de l'étude en cours. Cette exigence est appelée "comparaison d'éléments comparables" à l'annexe 2.

© OCDE (2025)

Un résultat positif signifie que, dans les conditions de l'essai, le produit chimique d'essai induit des mutations génétiques dans le tissu analysé.

63. Pour un tissu donné, à condition que tous les critères d'acceptabilité soient remplis, et que l'exposition au produit chimique d'essai et/ou à ses métabolites ait eu lieu, un produit chimique d'essai est considéré comme négatif si les deux critères a et b, ci-dessous, sont remplis dans toutes les conditions expérimentales examinées :

a. Aucun des groupes traités ne présente une augmentation statistiquement significative la fréquence des mutants par rapport au témoin solvant/véhicule concomitant ;
b. Aucune des réponses en termes de fréquence des mutants n'est liée à la dose, par exemple lorsqu'elle est évaluée par un test de tendance approprié (voir annexe 2, section 3) (non applicable à l'essai limite);

Les analyses des critères de type a et b, qui reposent toutes deux sur des données relatives aux témoins négatifs concomitants, sont considérées comme les analyses statistiques de première importance pour l'interprétation des données de l'étude. Les données relatives aux témoins négatifs historiques et leur distribution peuvent également être utiles pour contextualiser les résultats de la présente étude. Toutefois, cette comparaison a moins de poids que les critères a ou b. En outre, l'utilité des données des témoins historiques à cette fin dépend fortement de la quantité et de la qualité des données sous-jacentes. Par exemple, les données des témoins historiques ne peuvent pas servir de comparateur utile (par exemple, comme mesure indirecte de la variation biologique normale), lorsque la variation inter-études représente la principale source de variation. Ce concept et d'autres relatifs à l'utilisation appropriée des données des témoins historiques sont présentés à l'annexe 4. Ainsi, ce n'est que dans la mesure où des données relatives aux témoins négatifs historiques ont été obtenues en quantité suffisante, et que ces données sont jugées de qualité suffisante, qu'outre les critères a et b, un troisième critère, c, peut être appliqué comme suit :

c. Aucune des fréquences de mutants de l'un quelconque des groupes traités du produit chimique testé ne dépasse la limite supérieure dérivée de la distribution des données des témoins négatifs historiques.

   La limite supérieure dérivée de la distribution des témoins négatifs historiques est déterminée à l'aide d'une méthode adaptée à la quantité et à la qualité des données du laboratoire (par exemple, graphiques témoins, intervalles de tolérance ou quantiles), une justification étant fournie dans le rapport d'étude. La variabilité inter-laboratoires pouvant influer sur la distribution des témoins, aucune valeur numérique fixe ne peut être prescrite, et les laboratoires doivent définir leurs limites supérieures de manière scientifiquement fondée et transparente, comme décrit à l'annexe 2.

   Dans de nombreux cas, les fréquences considérées seront les valeurs moyennes de l'étude en cours comparées à une distribution dérivée des valeurs moyennes des témoins négatifs historiques. Cependant, il existe des situations dans lesquelles les données relatives aux témoins négatifs historiques et les distributions associées basées sur des données individuelles relatives à chaque animal sont utiles ; par exemple, lorsque les valeurs des témoins historiques sont peu nombreuses, ou lorsqu'il est intéressant d'estimer les sources de variation. Comme expliqué plus en détail à l'annexe 2, il est important dans ces cas-là, de procéder à des comparaisons appropriées. En particulier, une distribution basée sur des données individuelles d'animaux ne doit pas être utilisée comme comparateur pour les moyennes des groupes de traitement de l'étude en cours. Une distribution basée sur des données individuelles relatives à chaque animal ne doit être utilisée qu'à titre de comparaison pour les données individuelles relatives à chaque

© OCDE (2025)

animal de l'étude en cours. Cette exigence est appelée "comparaison d'éléments comparables" à l'annexe 2.

Un résultat négatif signifie que, dans les conditions de l'essai, le produit chimique d'essai n'induit pas de mutation génétique dans le tissu analysé.

64. L'exposition du tissu analysé à une substance chimique d'essai ou à ses métabolites peut être démontrée par des données de toxicité générale (par exemple, diminution du ratio poids corporel / poids des organes), ou par des données morphologiques ou histopathologiques obtenues dans la même étude ou dans des études comparables, ou dans d'autres études de toxicité génétique *in vivo*. Pour démontrer l'exposition du tissu analysé, on peut également avoir recours aux données ADME ou toxicocinétiques, ou bien à des analyses plasmatiques réalisées dans la même étude ou dans le cadre d'une étude indépendante employant la même voie d'administration et la même espèce (1).

65. En ce qui concerne les approches statistiques décrites aux paragraphes 62 et 63, il faut savoir qu'il n'y a jamais une seule méthode correcte pour effectuer une analyse statistique. Une façon pragmatique de procéder ici est donc de suggérer un certain ensemble d'analyses statistiques à titre d'exemple du type de démarches applicables. C'est sous cet angle que sont présentées les trois étapes d'analyse (a, b, c) décrites ci-dessus. On trouvera à l'Annexe 4 des informations susceptibles d'aider les laboratoires d'essai et les autorités de contrôle à évaluer si une analyse statistique est appropriée et si l'interprétation des données est scientifiquement fondée. Quelle que soit l'approche statistique adoptée, il importe de la définir préalablement à la conduite de l'étude.

66. En particulier dans les cas où les critères considérés (a et b ; ou a, b et c) ne concordent pas tous pour conclure au caractère positif ou négatif du résultat final, un jugement scientifique solide devrait être appliqué pour tenter d'interpréter les résultats comme positifs ou négatifs. Il convient également de noter qu'en général, le critère c a moins de poids que les critères a et b.

67. Si l'application d'un jugement scientifique solide ne permet pas de déterminer si une réponse est positive ou négative, il peut être nécessaire d'effectuer un examen plus approfondi des expériences existantes pour établir la pertinence biologique d'un résultat. Par exemple, ces investigations peuvent inclure l'examen de plus de plages ou colonies mutantes.

68. Le séquençage des plages de mutants pour déterminer s'il y a ou non une évolution du spectre de mutations induit par la substance chimique d'essai peut aussi aider à conclure à un résultat positif ou négatif. Comme décrit au paragraphe 58, le séquençage peut aussi aider à identifier des mutations jackpot. Pour les analyses du séquençage de l'ADN, un certain nombre de méthodes statistiques aident à interpréter les résultats (54)(55) (56) (57).

69. Dans de rares cas, même après un examen plus approfondi, si les données ne permettent pas de conclure si le produit chimique d'essai a produit un résultat positif ou négatif, l'étude est jugée équivoque.

70. Dans certains cas, des études équivoques peuvent rendre nécessaire de répéter l'expérience en modifiant les conditions expérimentales ou en changeant le système d'essai.

7.3. Rapport d'essai

71. Le rapport d'essai comporte les informations suivantes :

Produit chimique :
- données d'identification et n° CAS, s'il est connu ;
- source, numéro de lot s'il est disponible ;
- état physique et pureté ;

© OCDE (2025)

- propriétés physico-chimiques pertinentes pour la conduite de l'étude ;
- stabilité du produit chimique testé, si elle est connue ;

Solvant/véhicule :
- justification du choix du véhicule ;
- solubilité et stabilité du produit chimique testé dans le solvant ou véhicule, si elles sont connues ;
- préparation des formulations pour administration par l'alimentation, par l'eau de boisson ou par inhalation;
- déterminations analytiques sur les formulations (stabilité, homogénéité – pour les substances non solubles – concentrations nominales, par exemple) ;

Animaux d'expérience :
- espèce et souche utilisées et justification du choix ;
- nombre, âge et sexe des animaux ;
- source, conditions d'encagement, régime alimentaire, etc. ;
- poids individuel des animaux au début de l'essai, y compris la gamme des poids, l'écart moyen et pondéré dans chaque groupe ;

Conditions expérimentales :
- indiquer si l'étude (ou certaines parties de l'étude) a été conduite en aveugle (oui/non)
- preuve de la compétence du laboratoire ;
- données des témoins positifs et négatifs (solvant/véhicule) ;
- justification du choix des doses employées, sur la base par exemple des données de l'étude préliminaire de détermination des concentrations;
- détails concernant la formulation du produit chimique testé ;
- détails concernant l'administration du produit chimique testé;
- justification du choix de la voie d'administration ;
- justification du choix des tissus et types cellulaires analysés
- méthodes de mesure de la toxicité animale, y compris, si elles existent, analyses histopathologiques ou hématologiques et fréquence des observations animales et des mesures du poids corporel ;
- méthodes permettant de vérifier que le produit chimique testé a atteint le tissu cible ou la circulation sanguine en cas de résultats négatifs ;
- dose réelle (mg/kg de poids corporel/jour) pour la plupart des voies d'administration ou, pour une exposition par la nourriture ou l'eau de boisson, en partie par million (ppm) ou dose réelle calculée en fonction de la concentration (ppm) du produit chimique testé et consommation moyenne de nourriture ou d'eau de boisson, s'il y a lieu;
- détails sur la qualité de la nourriture et de l'eau ;
- description détaillée des calendriers de traitement et d'échantillonnage et justification des choix ;
- méthode d'euthanasie ;
- procédures d'isolation et de préservation des tissus ;
- méthodes d'isolation de l'ADN génomique des rongeurs, avec récupération du transgène de l'ADN génomique et transfert de l'ADN génomique vers un hôte bactérien ;
- source et numéros de lot de toutes les cellules, matériels et de tous les réactifs (s'il y a lieu) ;
- méthodes d'énumération des mutants ;
- méthodes d'analyse moléculaire des mutants et utilisation comme correctifs pour la clonalité et/ou pour calculer les fréquences de mutation, s'il y a lieu ;

Résultats :
- état général de l'animal avant et pendant la période d'essai, y compris signes de toxicité ;
- poids corporels et changements de poids corporel au cours de la période d'essai ;
- consommation de nourriture et/ou d'eau au cours de la période d'essai, s'il y a lieu, pour des études pour lesquelles l'exposition se fait via l'alimentation ou l'eau de boisson ;
- poids corporel et s'il y a lieu poids des organes, après mise à mort dans des conditions humaines;
- preuve de l'exposition du tissu analysé ;
- pour chaque tissu/animal, évaluation du nombre de mutants, du nombre de plages ou de colonies, du nombre d'encapsidation et de l'efficacité d'encapsidation, de la fréquence des mutants ;

© OCDE (2025)

- pour chaque groupe de tissus/animaux, nombre total de mutants et fréquence moyenne de mutation, déviation standard;
- relation dose-réponse, si c'est possible ;
- pour chaque tissu/animal, nombre de mutants indépendants et fréquence moyenne de mutation si une analyse moléculaire des mutations a été menée ;
- données relatives aux témoins négatifs concomitants ;
- données relatives aux témoins positifs concomitants ;
- données historiques sur les témoins négatifs avec indication claire du nombre de rongeurs, description statistique de la distribution, y compris le type de calcul utilisé pour générer les limites inférieure et/ou supérieure utilisées pour faciliter l'interprétation des données et justifier son utilisation (voir annexe 2, section 4.2), ainsi que la période couverte et les preuves que l'essai est sous contrôle pendant la période couverte ; un tableau représentatif qui capture ces informations et d'autres informations utiles est fourni à l'Annexe 2, section 6 ;
- données historiques relatives aux témoins positifs avec indication claire du nombre rongeurs, description statistique de la distribution précisant le mode de calcul utilisé pour générer les limites inférieure et/ou supérieure devant aider à interpréter les données, et justification de son utilisation (voir l'Annexe 2, section 4.2), ainsi que la période de temps couverte, et preuve que l'essai a été bien maîtrisé pendant la période couverte.
- déterminations analytiques si elles existent (concentrations d'ADN utilisées dans l'encapsidation, données du séquençage ADN, par exemple) ;
- analyses et méthodes statistiques employées ;

Discussion des résultats :
- expliquer la démarche scientifique appliquée pour évaluer les résultats de l'essai ; en particulier dans les cas où certains critères, mais pas tous, sont remplis pour un résultat négatif ou positif.

Conclusions
Résumer succinctement les conclusions de l'étude, notamment :
- si le produit chimique testé a montré des signes de mutation génique dans les conditions de l'essai ;
- évaluation de la pertinence biologique et statistique des résultats ;
- limites ou incertitudes susceptibles d'avoir une incidence sur l'interprétation des résultats.

© OCDE (2025)

8. BIBLIOGRAPHIE

(1) OECD (Organisation for Economic Cooperation and Development) (2017) Overview of the set of OECD Genetic Toxicology Test Guidelines and updates performed in 2014-2015. Series on Testing and Assessment, No. 238 – 2nd edition. Available at: Overview of the set of OECD Genetic Toxicology Test Guidelines and updates performed in 2014-2015 - Second edition | OECD.

(2) Thybaud, V, E. Lorge, D.D. Levy, J. van Benthem, G.R. Douglas, F. Marchetti, M.M. Moore and R. Schoeny (2017), Main issues addressed in the 2014-2015 revisions to the OECD genetic toxicology test guidelines", *Environ. Mol. Mutagen*., 58:284-295.

(3) OECD (2009), *Detailed Review Paper on Transgenic Rodent Mutation Assays*, Series on Testing and Assessment, N° 103, ENV/JM/MONO(2009)7, OECD, Paris.

(4) OECD (2011), *Retrospective Performance Assessment of OECD Test Guideline on Transgenic Rodent Somatic and Germ Cell Gene Mutation Assays*, Series on Testing and Assessment, N° 145, OECD, Paris.

(5) Marchetti, F., M. Aardema, C. Beevers, J. van Benthem, R. Godschalk, C.L. Yauk, B. Young, A. Williams and G.R. Douglas (2018), "Identifying germ cell mutagens using OECD test guideline 488 (transgenic rodent somatic and germ cell mutation assay) and integration with somatic cell testing", Mutation Res., 832-833: 7-18. Corrigendum: Mutation Res., 2019, 844: 70-71.

(6) Heddle, J.A., S. Dean, T. Nohmi, M. Boerrigter, D. Casciano, G.R. Douglas, B.W. Glickman, N.J. Gorelick, J.C. Mirsalis, H.-J Martus, T.R. Skopek, V. Thybaud, K.R.Tindall and N. Yajima (2000), "In vivo Transgenic Mutation Assays", *Environ. Mol. Mutagen*., 35: 253-259.

(7) Thybaud, V., S. Dean, T. Nohmi, J. de Boer, G.R. Douglas, B.W. Glickman, N.J. Gorelick, J.A. Heddle, R.H. Heflich, I. Lambert, H.-J. Martus, J.C. Mirsalis, T. Suzuki and N. Yajima (2003), "*In vivo* Transgenic Mutation Assays", *Mutation Res*., 540: 141-151.

(8) Heddle, J.A., H.-J. Martus and G.R. Douglas (2003), "Treatment and Sampling Protocols for Transgenic Mutation Assays", *Environ. Mol. Mutagen.*, 41: 1-6.

(9) Maurice, C., D.S. Dertinger, C.L. Yauk and F. Marchetti (2019) "Integrated in vivo genotoxicity assessment of procarbazine hydrochloride demonstrates induction of Pig-a and *lacZ* mutations, and micronuclei, in MutaMouse hematopoietic cells", *Environ. Mol. Mutagen*., 60: 505-512.

(10) Erikson, R.P. (2003), "Somatic Gene Mutation and Human Disease other than Cancer", *Mutation Res.*, 543: 125-136.

(11) Erikson, R.P. (2010), "Somatic Gene Mutation and Human Disease other than Cancer: an Update", *Mutation Res.*, 705: 96-106.

(12) Jackson, M., L. Marks, G.H.W. May and J.B. Wilson (2018) "The genetic basis of disease", *Essays Biochem*., 62:643-723.

(13) Nohmi, T., M. Katoh, H. Suzuki, M. Matsui, M. Yamada, M. Watanabe, M. Suzuki, N. Horiya, O. Ueda, T. Shibuya, H. Ikeda and T. Sofuni (1996), "A new Transgenic Mouse Mutagenesis Test System using Spi⁻ and 6-thioguanine Selections", *Environ. Mol. Mutagen*., 28(4): 465–470.

(14) Nohmi, T., T. Suzuki and K.I. Masumura (2000), "Recent Advances in the Protocols of Transgenic Mouse Mutation Assays", *Mutation Res.*, 455(1–2): 191–215.

(15) Toyoda-Hokaiwado, N., T. Inoue, K. Masumura, H. Hayashi, Y. Kawamura, Y. Kurata, M. Takamune, M. Yamada, H. Sanada, T. Umemura, A. Nishikawa and T. Nohmi (2010), "Integration of *in vivo* Genotoxicity and Short-term Carcinogenicity Assays using F344 *gpt* delta Transgenic Rats: *in vivo* Mutagenicity of 2,4-diaminotoluene and 2,6-diaminotoluene Structural Isomers", *Toxicol. Sci.*, 114(1): 71-78.

(16) Gossen, J.A., W.J. de Leeuw, C.H. Tan, E.C. Zwarthoff, F. Berends, P.H. Lohman, D.L. Knook

© OCDE (2025)

and J. Vijg (1989), “Efficient Rescue of Integrated Shuttle Vectors from Transgenic Mice: a Model for Studying Mutations *in vivo”, Proc. Natl. Acad. Sci. USA,* 86(20): 7971–7975.

(17) Gossen, J.A. and J. Vijg (1993), “A Selective System for lacZ-Phage using a Galactose-sensitive *E. coli* Host”, *Biotechniques*, 14(3): 326, 330.

(18) Vijg, J. and G.R. Douglas (1996), “Bacteriophage λ and Plasmid *lacZ* Transgenic Mice for studying Mutations *in vivo*” in: G. Pfeifer (ed.), *Technologies for Detection of DNA Damage and Mutations, Part II,* Plenum Press, New York, NY, USA, pp. 391–410.

(19) Jakubczak, J.L., G. Merlino, J.E. French, W.J. Muller, B. Paul, S. Adhya and S. Garges (1996), “Analysis of Genetic Instability during Mammary Tumor Progression using a novel Selection-based Assay for *in vivo* Mutations in a Bacteriophage λ Transgene Target”, *Proc. Natl. Acad. Sci. USA*, 93(17): 9073–9078.

(20) Nohmi, T., M. Suzuki, K. Masumura,M. Yamada, K. Matsui, O. Ueda, H. Suzuki, M. Katoh, H. Ikeda and T. Sofuni (1999), “Spi⁻ Selection: an Efficient Method to Detect γ-ray-induced Deletions in Transgenic Mice”, *Environ. Mol. Mutagen*., 34(1): 9–15.

(21) Gad, S.C., C.D. Cassidy, N. Aubert, Spainhour B. and H. Robbe (2006) “Nonclinical vehicle use in studies by multiple routes in multiple species”, *Int. J. Toxicol*., 25(6): 499-521.

(22) OECD (2009), Part 2: *Annexes to the Detailed Review Paper on Transgenic Rodent Mutation Assays*, Series on Testing and Assessment, N° 103, ENV/JM/MONO(2009)29, OECD, Paris.

(23) Hayashi, M., K. Dearfield, P. Kasper, D. Lovell, H.-J. Martus, V. Thybaud (2011), “Compilation and Use of Genetic Toxicity Historical Control Data”, *Mutation Res.,* 732(2): 87-90.

(24) Ryan, T.P. (2000), “Statistical Methods for Quality Improvement”, 2nd ed., John Wiley and Sons, New York.

(25) Fang, Y. (2003), “C-chart, X-chart, and the Katz Family of Distributions”, *J. Quality Technology*, 35:1-15, 2003.

(26) Lovell D.P., M. Fellows, F. Marchetti, J. Christiansen, A. Elhajouji, K. Hashimoto, S. Kasamoto, Y. Li, O. Masayasu, M.M. Moore, M. Schuler, R. Smith, L.F. Stankowski Jr, J. Tanaka, J.Y. Tanir, V. Thybaud, F. Van Goethem and J. Whitwell (2018), “Analysis of negative historical control group data from the in vitro micronucleus assay using TK6 cells”, *Mutat Res Genet Toxicol Environ Mutagen*., 825:40-50.

(27) Dertinger S.D., J.A. Bhalli, D.J. Roberts, L.F. Stankowski Jr, B.B. Gollapudi, D.P. Lovell, L. Recio, T. Kimoto, D. Miura and R.H. Heflich (2021) “Recommendations for conducting the rodent erythrocyte Pig-a assay: A report from the HESI GTTC Pig-a Workgroup”, *Environ Mol Mutagen*., 62(3):227-237.

(28) Yauk, C.L., J.D. Gingerich, L. Soper, A. MacMahon, W.G. Foster and G.R. Douglas (2005), “A *lacZ* Transgenic Mouse Assay for the Detection of Mutations in Follicular Granulosa Cells”, *Mutation Res*., 578(1-2): 117-123.

(29) OECD (2000), *Guidance Document on the Recognition, Assessment and Use of Clinical Signs as Humane Endpoints for Experimental Animals Used in Safety Evaluation*, Series on Testing and Assessment, N°19, ENV/JM/MONO(2000)7, OECD, Paris.

(30) Diehl K.H., R. Hull, D. Morton, R. Pfister,Y. Rabemampianina, D. Smith, J.M. Vidal, C. van de Vorstenbosch and European Federation of Pharmaceutical Industries Association and European Centre for the Validation of Alternative Methods, (2001) “A good practice guide to the administration of substances and removal of blood, including routes and volumes”, *J. Appl. Toxicol*., 21(1):15-23.

(31) Turner P.V., T. Brabb, C. Pekow and M.A. Vasbinder, (2011) “Administration of substances to laboratory animals: routes of administration and factors to consider”, *J. Am. Assoc. Lab. Anim. Sci.*, 50(5):600-613.

(32) White, P.A., G.R. Douglas, D.H. Phillips and V.M. Arlt (2017) “Quantitative relationship between *lacZ* mutant frequency and DNA adduct frequency in Muta™Mouse tissues and

© OCDE (2025)

cultured cells exposed to 3-nitrobenzanthrone. *Mutagenesis*, 32(2): 299-312.
(33) Marchetti F., G. Zhou, D. LeBlanc, P.A. White, A. Williams, C.L. Yauk and G.R. Douglas (2021) "The 28 + 28 day design is an effective sampling time for analyzing mutant frequencies in rapidly proliferating tissues of MutaMouse animals", *Arch. Toxicol*., 95(3):1103-1116
(34) Douglas, G.R., J. Jiao, J.D. Gingerich, J.A. Gossen and L.M. Soper (1995), "Temporal and Molecular Characteristics of Mutations Induced by Ethylnitrosourea in Germ Cells Isolated from Seminiferous Tubules and in Spermatozoa of *lacZ* Transgenic Mice", *Proc. Natl. Acad. Sci. USA*, 92: 7485-7489.
(35) Douglas, G.R., J.D. Gingerich, L.M. Soper and J. Jiao (1997), "Toward an Understanding of the Use of Transgenic Mice for the Detection of Gene Mutations in Germ Cells", *Mutation Res.*, 388(2-3): 197-212.
(36) Singer, T.M., I.B. Lambert, A. Williams, G.R. Douglas and C.L. Yauk (2006), "Detection of Induced Male Germline Mutation: Correlations and Comparisons between Traditional Germline Mutation Assays, Transgenic Rodent Assays and Expanded Simple Tandem Repeat Instability Assays", *Mutation. Res*., 598: 164-193.
(37) Olsen, A.K., A., Andreassen, R., Singh, R., Wiger, N., Duale, P.B., Farmer, and G. Brunborg (2010), "Environmental exposure of the mouse germ line: DNA adducts in spermatozoa and formation of de novo mutations during spermatogenesis". *PLoS One*, 28;5(6):e11349
(38) Oakberg, E.F. (1956), "Duration of spermatogenesis in the mouse and timing of the stages of the cycle of the seminiferous epithelium", *Am. J. Anat*., 99: 507–516.
(39) Clermont, Y. (1972), "Kinetics of spermatogenesis in mammals: seminiferous epithelium cycle and spermatogonial renewal". *Physiol. Rev. 52:* 198-236.
(40) Marchetti, F., M. Aardema, C. Beevers, J. van Benthem, G.R. Douglas, R. Godschalk, C.L. Yauk, B. Young and A. Williams (2018), "Simulation of mouse and rat spermatogenesis to inform genotoxicity testing using OECD test guideline 488", *Mutation Res*., 832-833: 19-28. Corrigendum: Mutation Res., 2019, 844: 69.
(41) Bielas, J.H. and J.A. Heddle (2000) Proliferation is necessary for both repair and mutation in transgenic mouse cells. *Proc. Natl. Acad. Sci. USA*, 97: 11391-11396.
(42) O'Brien, J.M., M.A. Beal, J.D. Gingerich, L. Soper, G.R. Douglas, C.L. Yauk and F. Marchetti (2014), "Transgenic rodent assay for quantifying male germ cell mutant frequency", *J. Vis. Exp*., 90: e51576
(43) O'Brien J.M., Beal M.A., Yauk, C.L. and F. Marchetti (2016) "Benzo(a)pyrene is mutagenic in mouse spermatogonial stem cells and dividing spermatogonia. *Toxicol. Sci*., 152: 363-371.
(44) Dean, S.W., T.M. Brooks, B. Burlinson, J. Mirsalis, B. Myhr, L. Recio and V. Thybaud (1999), "Transgenic Mouse Mutation Assay Systems can Play an important Role in Regulatory Mutagenicity Testing *in vivo* for the Detection of Site-of-contact Mutagens", *Mutagenesis*, 14(1): 141-151.
(45) Bielas, J.H. (2002), "A more Efficient Big Blue® Protocol Improves Transgene Rescue and Accuracy in an Adduct and Mutation Measurement", *Mutation Res.*, 518: 107-112.
(46) Kohler, S.W., G.S. Provost, P.L. Kretz, A. Fieck, J.A. Sorge and J.M. Short (1990), "The Use of Transgenic Mice for Short-term, *in vivo* Mutagenicity Testing", *Genet. Anal. Tech. Appl.*, 7(8): 212–218.
(47) Nohmi, T., K. Masumura and N. Toyoda-Hokaiwado (2017), "Transgenic rat models for mutagenesis and carcinogenesis", *Genes and Environ.,* 39:11.
(48) Besaratinia, A., H. Li, J. Yoon, A. Zheng, H. Gao and S. Tommasi (2012) "A high-throughput next-generation sequencing-based method for detecting the mutational fingerprint of carcinogens", *Nucleic Acids Res*., 40(15):e116-6.
(49) Beal, M.A., R. Gagne, A. Williams, Marchetti F. and C.L. Yauk (2015) " Characterizing benzo(a)pyrene-induced lacZ mutation spectrum in transgenic mice using Next Generation

© OCDE (2025)

Sequencing, *BMC Genomics*, 16: 812.

(50) Carr, G.J. and N.J. Gorelick (1995), "Statistical Design and Analysis of Mutation Studies in Transgenic Mice*", Environ. Mol. Mutagen*, 25(3): 246–255.

(51) Fung, K.Y., G.R. Douglas and D. Krewski (1998), "Statistical Analysis of *lacZ* Mutant Frequency Data from Muta™Mouse Mutagenicity Assays", *Mutagenesis*, 13(3): 249–255.

(52) Piegorsch, W.W., B.H. Margolin, M.D. Shelby, A. Johnson, J.E. French, R.W. Tennant and K.R. Tindall (1995), "Study Design and Sample Sizes for a *lacI* Transgenic Mouse Mutation Assay", *Environ. Mol. Mutagen*., 25(3): 231–245.

(53) Piegorsch, W.W., A.C. Lockhart, G.J. Carr, B.H. Margolin, T. Brooks, G.R. Douglas, U.M. Liegibel, T. Suzuki, V. Thybaud, J.H. van Delft and N.J. Gorelick (1997), "Sources of Variability in Data from a Positive Selection *lacZ* Transgenic Mouse Mutation Assay: an Interlaboratory Study", *Mutation. Res*., 388(2–3): 249–289.

(54) Adams, W.T. and T.R. Skopek (1987), "Statistical Test for the Comparison of Samples from Mutational Spectra", *J. Mol. Biol*., 194: 391-396.

(55) Carr, G.J. and N.J. Gorelick (1996), "Mutational Spectra in Transgenic Animal Research: Data Analysis and Study Design Based upon the Mutant or Mutation Frequency", *Environ. Mol. Mutagen,* 28: 405–413.

(56) Dunson, D.B. and K.R. Tindall (2000), "Bayesian Analysis of Mutational Spectra", *Genetics*, 156: 1411–1418.

(57) Lewis P.D., B. Manshian, M.N. Routledge, G.B. Scott and P.A. Burns (2008), "Comparison of Induced and Cancer-associated Mutational Spectra using Multivariate Data Analysis", *Carcinogenesis,* 29(4): 772-778.

(58) Dertinger SD, D Li, C Beevers, GR Douglas, RH Heflich, DP Lovell, DJ Roberts, R Smith, Y Uno, A Williams, KL Witt, A Zeller, C Zhou (2023) Assessing the quality and making appropriate use of historical negative control data: A report of the International Workshop on Genotoxicity Testing (IWGT). Environ Mol Mutagen, in press, doi: 10.1002/em.22541

© OCDE (2025)

Annexe 1: définitions

Capside : structure protéinique qui entoure une particule virale.

Décalage du cadre de lecture : mutation génétique provoquée par des insertions ou des délétions de nucléotides dont le nombre n'est pas un multiple de trois dans la séquence d'ADN qui code une protéine ou un peptide.

Délétion : mutation induisant une perte d'une partie du génome allant d'un à plusieurs nucléotide(s) (séquentiel(s)).

Délétions importantes : délétions de l'ADN dont le nombre est supérieur à plusieurs kilobases (l'essai de sélection du gène *Spi*$^-$ permet de les détecter efficacement).

Efficacité de l'encapsidation : efficacité avec laquelle les bactériophages encapsidés sont retrouvés dans la bactérie hôte.

Électroporation : envoi d'impulsions électriques qui améliorent la perméabilité membranaire des cellules.

Encapsidation : synthèse de particules phagiques infectieuses issues d'une préparation à base de *capsides* et de queues protéiques de phages et d'un concatémère de molécules d'ADN de phages. Technique fréquemment utilisée pour introduire l'ADN cloné et véhiculé par un vecteur lambda (séparé par des *sites cos*) à l'intérieur des particules lambda infectieuses.

Expansion clonale : production de nombreuses cellules à partir d'une seule cellule (mutante).

Gène endogène : gène qui prend sa source dans le génome.

Gène neutre : gène non soumis aux pressions de la sélection positive ou négative.

Gène rapporteur : gène dont le produit (gène mutant) est facilement détecté.

Insertion : ajout d'une ou de plusieurs paire(s) de bases nucléotidiques dans une séquence d'ADN.

Intervalle de prédiction: intervalle estimé susceptible de contenir une observation future avec une certaine probabilité, compte tenu d'observations antérieures.

Jackpot : apparition d'un nombre important de mutants suite à une expansion clonale à partir d'une seule mutation.

Ligature : formation d'une liaison covalente entre deux extrémités de molécules d'ADN par l'ADN ligase.

Limite de contrôle : ligne(s) horizontale(s) tracée(s) sur un graphique de contrôle statistique ; ces valeurs sont définies par les chercheurs au cas par cas ; dans le domaine du contrôle de la qualité, elles correspondent généralement à la moyenne de l'échantillon ± 3 fois l'écart type, mais d'autres multiples de l'écart type peuvent être utilisés, par exemple 2× ou 1.96×

Mitogène : substance qui aide une cellule à entamer sa division cellulaire, provoquant ainsi la mitose (ou division cellulaire).

Moment d'échantillonnage : moment qui marque la fin de la période antérieure à la mise à mort de l'animal dans des conditions humaines, au cours de laquelle le produit chimique testé n'est pas administré et au

© OCDE (2025)

cours de laquelle les lésions d'ADN non réparées se fixent sous forme de mutations stables.

Mutation ponctuelle : terme générique désignant une mutation qui affecte seulement une petite séquence d'ADN (petites insertions, délétions et substitutions d'une base par une autre).

Période d'administration : période totale au cours de laquelle un animal se voit administrer le produit chimique testé.

Produit chimique testé : Le terme produit chimique testé désigne ce qui est testé.

Sélection positive : méthode permettant uniquement aux mutants de survivre.

Site cos : segment d'ADN simple brin possédant 12 nucléotides, présent aux deux extrémités du génome double brin du bactériophage lambda.

Substitution d'une paire de bases : type de mutation qui engendre le remplacement d'un nucléotide à une seule base par un autre nucléotide d'ADN.

Transgénique : qualifie un être vivant issu d'un organisme ou un organisme dont le génome a été modifié par l'introduction d'un ou plusieurs gène(s) d'une autre espèce.

Unité Formant Colonie (UFC) : unité utilisée pour dénombrer les bactéries viables.

Unité Formant Plage (UFP) : unité utilisée pour dénombrer les bactériophages viables.

Variation extra-binomiale : variabilité des estimations répétées d'une proportion de population plus importante que ce qui pourrait être escompté si la population avait une distribution binomiale.

Vecteur navette : vecteur dont la structure lui permet de se propager dans deux espèces hôtes différentes ; ainsi, l'ADN introduit dans un vecteur navette peut être testé ou manipulé dans deux types de cellules différents ou dans deux organismes différents.

© OCDE (2025)

Annexe 2 : Analyse statistique et interprétation des données

1. Préambule

1. Cette annexe s'inspire en partie d'un rapport de l'International Workshop on Genotoxicity Testing (IWGT) [Dertinger et al., 2023]. Les lecteurs sont invités à consulter cette publication pour y trouver des exemples de séries de données illustratives et des descriptions détaillées de méthodes.

2. Les méthodes d'analyse statistique utilisées dans cet essai peuvent varier, et cette Ligne directrice pour les essais fournit trois exemples aux paragraphes 62-63, qui constituent des approches pratiques et non prescriptives. D'autres méthodes sont acceptables si elles sont justifiées par un raisonnement statistique solide. Les méthodes choisies seront documentées dans un plan d'étude ou de validation écrit avant de commencer l'étude et l'approche choisie devra pouvoir être défendue.

3. Les approches statistiques ci-dessous sont des outils représentatifs pour l'évaluation de la génotoxicité, mais un jugement scientifique solide est essentiel, en particulier lorsque les analyses statistiques donnent des résultats contradictoires. L'application d'un jugement scientifique solide est conforme aux recommandations formulées par un groupe de travail de l'OCDE sur la génotoxicité qui s'est réuni à Ottawa, au Canada, en 2013 [OCDE, 2017].

4. Compte tenu de la formulation des paragraphes 62-63 de la Ligne directrice et des outils d'évaluation des données décrits dans la présente annexe, il convient de garder à l'esprit les principes généraux suivants :

i. Lorsque les données relatives aux témoins négatifs concomitants remplissent les critères d'acceptabilité de l'étude, elles représentent le comparateur le plus important pour juger si un produit chimique testé particulier a induit un effet génotoxique.

ii. Les données des témoins historiques peuvent fournir un contexte utile pour interpréter les résultats de l'étude, mais cela nécessite des preuves à l'appui, notamment : a) des données des témoins historiques pertinentes et générées en quantité suffisante, et ; b) des données qui décrivent avec précision la variabilité inter-animale.

iii. Les données relatives aux témoins historiques doivent être visualisées avant toute comparaison des études ; un ou plusieurs graphiques montrant la stabilité des données dans le temps sont particulièrement utiles.

iv. Les évaluations qualitatives et semi-quantitatives des données des témoins historiques doivent être complétées par des évaluations quantitatives portant notamment sur : a) la stabilité des données dans le temps, et b) une information indiquant dans quelle mesure la variation entre les études explique la variabilité totale observée.

v. Lorsque la variation inter-animaux est la principale source de variabilité, la relation entre les réponses observées dans l'étude et un intervalle ou une valeur limite supérieure dérivés d'un témoin historique (*c'est-à-dire le* critère c) peut être utilisée avec un degré de confiance élevé pour contextualiser les résultats.

vi. Lorsque la variation inter-études est la principale source de variabilité, les comparaisons entre les données de l'étude et les limites dérivées des témoins historiques sont moins utiles et, par conséquent, il convient de mettre moins l'accent sur l'utilisation

© OCDE (2025)

des données des témoins historiques pour contextualiser les résultats d'une étude particulière.

vii. La plupart des tests statistiques, y compris ceux de la présente annexe, reposent sur l'hypothèse d'un échantillonnage ou d'un blocage aléatoire afin d'atténuer les facteurs non contrôlés affectant les résultats de l'étude.

2. Comparaisons par paires

5. Un type d'essai statistique qui peut être mis en œuvre consiste à effectuer des comparaisons par paires afin de déterminer si le groupe témoin véhicule/solvant (négatif) concomitant est significativement différent des groupes exposés au produit chimique d'essai. Les analyses paramétriques qui utilisent l'analyse de variance (ANOVA) avec un test de comparaisons multiples approprié sont couramment utilisées, mais d'autres méthodologies sont également acceptables tant que les hypothèses sous-jacentes aux analyses sont satisfaites. Par exemple, si la transformation des données n'entraîne pas d'homoscédasticité, on peut appliquer une ANOVA pondérée (corrigée de la variance) et/ou des tests t, ou des méthodes non paramétriques appropriées.

2.1 Plan factoriel

6. Si plusieurs sexes sont utilisés dans une étude, les approches de plan factoriel sont généralement avantageuses car elles évaluent les effets du sexe et de la dose du produit chimique d'essai simultanément et examinent l'interaction de ces facteurs. Un logiciel statistique standard peut être utilisé pour analyser les données générées par ce plan.

7. L'analyse comprend tout d'abord l'examen du terme d'interaction sexe-dose dans le tableau ANOVA. (Notons que les statisticiens qui optent pour une démarche de modélisation avec par exemple des modèles linéaires généralisés peuvent aborder l'analyse d'une manière comparable, mais différente puisqu'ils n'établiront pas nécessairement un tableau ANOVA classique.) En l'absence de terme d'interaction significatif, la combinaison des valeurs obtenues pour les deux sexes ou pour tous les niveaux de dose permet de réaliser des tests statistiques valides entre les niveaux, en se basant sur le terme de variabilité intragroupe combinée fourni par l'ANOVA.

8. L'étape suivante de l'analyse consiste à partitionner l'estimation de la variabilité entre niveaux de dose en contrastes, ce qui permet de pratiquer un test des contrastes linéaires et quadratiques des réponses entre le témoin véhicule/solvant concomitant et les niveaux de doses du produit chimique d'essai. Lorsqu'il y a une interaction significative sexe × niveau de dose, ce terme peut à son tour être partitionné en contrastes d'interaction linéaire × sexe et quadratique × sexe. Ces termes permettent de vérifier si les réponses aux niveaux de dose sont parallèles pour les deux sexes ou si elles diffèrent selon le sexe.

9. L'estimation de la variabilité intragroupe combinée permet de procéder à des tests par paires entre les moyennes, par exemple entre sexes ou niveaux de dose (par exemple contre les niveaux de témoins négatifs (solvant/véhicule)). En cas d'interaction significative, des comparaisons peuvent être faites entre les moyennes des différents niveaux de dose pour un même sexe, ou de plusieurs sexes à l'intérieur d'un même niveau de dose.

© OCDE (2025)

3. Test de tendance

10. Le second type d'analyse qui peut être utilisé est le test de tendance visant à identifier une relation dose-réponse. Pour réaliser ce test, les données doivent être disponibles pour le groupe témoin négatif et pour tous les groupes exposés. Toutefois, les tests de tendance ne sont pas applicables, en principe, aux essais limites (qui n'utilisent qu'un seul niveau de dose du produit chimique d'essai, tels que décrits au paragraphe 40). Si l'on recourt à ce type d'analyse, la prudence est de mise dans l'interprétation des résultats. Ainsi, un test de tendance linéaire simple peut échouer à détecter une tendance si la relation dose-réponse est non monotone. Dans de tels cas, des tests plus avancés, comme le test proposé par Bretz et Hothorn [2003] permettent de corriger ce problème et peuvent être utiles. La création d'un graphique de la relation dose-réponse peut aider à choisir le test approprié ou à déterminer si une transformation est nécessaire.

4. Distribution des données relatives aux témoins négatifs historiques

11. Les distributions des données relatives aux témoins négatifs historiques sont importantes pour déterminer le nombre minimal d'observations qui doit être évalué, et pour apprécier l'acceptabilité d'une étude particulière, ainsi que pour déterminer si les résultats d'un groupe traité exposé au produit chimique testé dépasse la limite supérieure de la distribution des témoins négatifs historiques (critère c). Les ensembles de données relatives aux témoins négatifs historiques étant très importants, on trouvera ci-après des conseils pour les constituer.

4.1 Constitution des ensembles de données relatives aux témoins négatifs historiques : moyennes du groupe par rapport données individuelles relatives à chaque animal

12. Il est utile de compiler les données relatives aux témoins négatifs historiques à partir des données individuelles relatives aux animaux et des moyennes des groupes d'étude. Premièrement, les données individuelles relatives à chaque animal facilitent certaines analyses statistiques qui ne peuvent pas être réalisées avec des moyennes de groupe. Un exemple en est l'Analyse des Composantes de la Variance, qui vise à estimer dans quelle mesure les différences entre animaux par rapport aux facteurs de nuisance dépendant de l'étude expliquent la variation observée dans un essai. Cette considération importante ne peut être prise en compte lorsque les analyses sont fondées sur des données moyennes de groupe. Deuxièmement, privilégier dans un premier temps les données individuelles relatives à chaque animal peut être considérée pour des raisons pratiques car il est parfois difficile, pour certains systèmes d'essai *in vivo*, d'obtenir des bases de données de taille adéquate avec des moyennes de groupe en raison des limites de l'expérimentation animale.

13. À mesure qu'un laboratoire acquiert de l'expérience et réalise davantage d'études, les bases de données des témoins historiques peuvent être élargies pour inclure à la fois des données individuelles relatives à chaque animal et des données moyennes sur les groupes. Les données individuelles relatives à chaque animal resteront essentielles pour évaluer la qualité et la stabilité des données relatives aux témoins négatifs historiques. Cependant, la distribution moyenne des données relatives aux témoins négatifs historiques du groupe peut remplacer la distribution des données relatives à chaque animal dans le contexte des évaluations de type c du critère.

14. Bien que les bases de données fondées sur les moyennes individuelles relatives à chaque animal et pour chaque groupe soient interdépendantes, les intervalles des données relatives aux témoins négatifs historiques basées sur chaque animal sont plus larges que ceux basés sur les moyennes de groupe. Par conséquent, le principe de « comparaison d'éléments comparables » est primordial, à savoir que les résultats de chaque animal doivent être comparés à des intervalles basés sur des données de chaque

© OCDE (2025)

animal, tandis que les résultats moyens d'un groupe sont comparés de manière appropriée à des intervalles basés sur des moyennes de groupe.

4.2 Taille de la base de données

15. Une base de données des témoins négatifs historiques (HCD) doit contenir des données provenant d'au moins ≥ 20 études indépendantes et ≥ 100 animaux. Ces valeurs sont conformes aux recommandations formulées par Hayashi et al. [2011] et Igl et ses collègues [2019]. Une base de données comportant < 20 études indépendantes et < 100 individus peut néanmoins être utile, mais devrait avoir moins de poids que les versions ultérieures dérivées de ≥ 20 études indépendantes et de ≥ 100 individus.

16. L'objectif de réaliser 20 études indépendantes, et d'obtenir plus de 100 animaux est compliqué par le fait que certains effets peuvent être affectés par certains facteurs expérimentaux tels que l'espèce, la souche, l'âge, le sexe, la composition du véhicule, etc. Pour l'essai de mutation chez les RTG, il convient également de tenir compte du tissu analysé pour détecter les mutations ou les cassures de brins d'ADN. Lorsqu'il est établi que le ou les facteurs influent de façon appréciable sur l'effet étudié, il peut s'avérer nécessaire de constituer des bases de données de témoins historiques séparées, ce qui représente une charge de travail trop lourde pour être réaliste et risque de se traduire par de nombreuses petites bases de données qui n'atteindront jamais l'objectif de 100 observations. Dans ces cas, une solution pratique peut s'avérer nécessaire, par exemple en réalisant un nombre relativement faible d'expériences indépendantes, chacune avec 10 animaux, et chacune avec un véhicule différent, couramment utilisé. Une fois ces études pilotes terminées, et tant que les données sont jugées « sous contrôle » (voir la section 4.4 « Évaluation de la qualité des données des témoins historiques »), un intervalle de contrôle historique calculé de manière appropriée sur la base des données individuelles relatives à chaque animal peut être utilisé pour évaluer l'acceptabilité de l'essai et répondre au critère c. Toutefois, comme expliqué précédemment, les analyses réalisées à partir de ces données devraient avoir moins de poids que les bases de données ultérieures, dont la taille sera plus appropriée.

17. Dans la plupart des cas, les bases de données des témoins négatifs historiques seront générées à partir des données relatives aux animaux ayant reçu le solvant/véhicule seul. Pour certains effets, il peut être utile d'inclure des données d'animaux naïfs dans les bases de données historiques relatives aux témoins négatifs. Cela suppose que le solvant/véhicule est sans effet. Cet exemple souligne l'importance de compléter les données de génotoxicité des témoins historiques par la communication de métadonnées qui incluent des facteurs expérimentaux tels que le modèle animal, la souche, le sexe, le solvant/véhicule, l'âge au moment du prélèvement des tissus et la configuration du matériel d'analyse.

18. Dans le même ordre d'idées, en supposant que des publications à l'appui ou d'autres données soient disponibles, il peut être approprié de considérer initialement un ou plusieurs facteurs tels que le sexe et/ou le solvant/véhicule comme n'ayant aucun effet génotoxique. Cela permet de construire plus efficacement une base de données de taille adéquate. Toutefois, à mesure que les données s'accumulent, ces hypothèses doivent être réévaluées, et de cette manière, les décisions relatives à la mise en commun ou à l'analyse des bases de données peuvent être prises sur la base d'un raisonnement scientifique solide.

4.3 Gestion des bases de données des témoins historiques

19. Les données relatives aux témoins historiques peuvent provenir d'études réalisées conformément aux bonnes pratiques de laboratoire (BPL), d'études non conformes aux BPL ou des deux, à condition que les conditions expérimentales soient similaires et que les mesures de contrôle de la qualité confirment l'exactitude des données non couvertes par une déclaration de conformité aux BPL. Une base de données de témoins historiques établie doit être tenue à jour et publiée à intervalles réguliers, ou bien par suite d'un autre événement déclencheur défini, sauf si aucune nouvelle donnée n'a été générée au cours des 12 mois précédents. Si les bases de données des témoins historiques peuvent être mises à jour selon un

© OCDE (2025)

calendrier prédéterminé, les critères d'acceptabilité de l'essai doivent être évalués pour chaque étude, ce qui permet un retour d'information en temps réel sur la performance de l'essai, afin d'être utiles pour modifier un calendrier de mise à jour de base de données fixe.

20. Par exemple, si une étude ne remplit pas les critères d'acceptation de l'essai, des travaux visant à *identifier* la ou les causes sous-jacentes doivent être entrepris. Si une cause sous-jacente est identifiée (par exemple, une erreur technique explicable, un problème de réactif ou une défaillance de l'équipement), les données relatives aux témoins de l'essai *ne* doivent *pas* être utilisées dans l'itération suivante de la base de données des témoins historiques. Dans un souci de transparence, ces données doivent être ajoutées à la base de données, mais omises du calcul des intervalles qui constituent la base des critères d'acceptabilité de l'essai ou du critère c.

21. Si un essai ne remplit pas les critères d'acceptation et que la cause sous-jacente n'*est pas identifiée*, il convient de décider si les données relatives aux témoins de l'essai doivent être incluses dans une future base de données de témoins historiques. Pour une transparence maximale, le processus décisionnel relatif à l'exclusion des données de contrôle d'une base de données devrait idéalement être articulé *a priori*. Si les points de données de contrôle inhabituels sont jugés acceptables pour inclusion dans une future base de données de contrôles historiques parce qu'aucune erreur technique, etc., n'a été identifiée, cela devrait, dans l'idéal, entraîner la réédition de la base de données et des analyses associées (*par exemple,* génération d'un graphique de contrôle, nouveau calcul d'intervalle, etc.) avant la date prévue, de sorte que la base de données de contrôles historiques reste représentative des études à venir.

22. Enfin, il convient de choisir un calendrier pour l'analyse des données relatives aux témoins historiques en ce qui concerne la stabilité, par exemple une limite supérieure de 5 ans. Par conséquent, il conviendra souvent de prendre en compte des facteurs spécifiques à l'essai. Par exemple, si l'essai est peu fréquent ou si un tissu particulier est rarement étudié, on peut utiliser des données plus anciennes pour obtenir environ 20 études et 100 observations indépendantes, à condition qu'il n'y ait pas eu de changement de méthodologie au cours de la période considérée.

4.4 Évaluation de la qualité des données relatives aux témoins négatifs historiques

23. Pour évaluer la qualité des DCH, il convient d'adopter une approche qualitative et semi quantitative [OCDE, 2017]. Les évaluations initiales doivent évaluer si (i) les données de génotoxicité sont cohérentes avec les résultats publiés par les laboratoires compétents (dans ce cas, il peut être utile de rechercher des articles centrés sur la validation/transférabilité interlaboratoires) ; (ii) le niveau de variation entre les échantillons au sein d'une étude et entre les études par rapport aux résultats publiés par les laboratoires compétents ; et (iii) il existe une dérive évidente vers des fréquences plus élevées ou plus basses en fonction du temps. Des analyses quantitatives doivent également être effectuées et commencent généralement par caractériser la distribution des données, ce qui peut être utile car les données résiduelles qui ne se rapprochent pas d'une distribution normale ne doivent pas étayer certaines des analyses décrites ci-dessous.

24. Lorsque la transformation des données est jugée nécessaire, des outils informatiques peuvent être utilisés pour sélectionner une transformation appropriée. Par exemple, l'approche Box-Cox disponible dans de nombreux logiciels statistiques [Box et Cox, 1964] calcule une estimation du paramètre de puissance lambda, qui peut être utilisée pour identifier une transformation optimale. Toutefois, le simple arrondi de lambda à la valeur 0.5 la plus proche permet de faire un choix parmi les transformations courantes. Par exemple, un lambda de -1 = réciproque, 0 = logarithmique, 0.5 = racine carrée, 1 = non transformée et 2 = au carré.

25. La prochaine analyse devrait utiliser des outils statistiques largement utilisés dans les domaines de la fabrication et du contrôle des procédés. Les graphiques de contrôle sont utiles pour étudier les DCH

© OCDE (2025)

[Hayashi et al., Mutat. 2011] et sont recommandées dans plusieurs LD de l'OCDE sur la toxicologie génétique [OCDE, 2017]. Un graphique de contrôle est un graphique utilisé pour étudier l'évolution d'un processus dans le temps. Les données sont représentées dans un ordre séquentiel qui reflète la date à laquelle elles ont été collectées. Un graphique de contrôle comporte une ligne centrale tracée parallèlement à l'axe des abscisses décrivant la moyenne, une ligne supérieure pour la limite de contrôle supérieure (LCL ; moyenne plus 3 écarts-types, *c'est-à*-dire sigma +3), et une ligne inférieure pour la limite de contrôle inférieure (LCL ; sigma -3). Les graphiques peuvent montrer des lignes sigma ±2, appelées « limites d'avertissement » (à noter que les valeurs d'écart-type figurant sur les graphiques témoins ne sont pas les mêmes que celles des statistiques synthétiques de l'échantillon ; les valeurs d'écart-type des graphiques témoins sont plutôt estimées par la plage moyenne divisée par une constante de normalisation, d2, voir Dertinger et al., 2023). En comparant les données actuelles à ces courbes, on peut tirer des conclusions indiquant si un processus est cohérent (« sous contrôle ») ou variable (« hors contrôle »).

26. Différents types de graphiques témoins ayant des objectifs différents peuvent être utiles pour les données de toxicologie génétique et sont brièvement décrits ici. Les diagrammes en I peuvent être utilisés avec des données continues ou discrètes (comptage). Il s'agit de diagrammes séquentiels ordonnés dans le temps qui représentent les valeurs individuelles sur l'axe des ordonnées et l'ordre des individus sur l'axe des abscisses. Les graphiques en C représentent également les valeurs individuelles pour lesquelles on n'a pas essayé de prendre en considération les groupements qui pourraient exister dans les données. La principale différence entre les graphiques en I et en C est que ces derniers se limitent aux comptages (nombre d'événements) et supposent une distribution de Poisson sous-jacente. Les diagrammes en X-barres représentent les moyennes obtenues à partir d'une série d'expériences, chacune d'elles étant réalisée sur plusieurs individus (*par exemple,* la moyenne de 5 animaux d'un groupe témoin négatif).

27. Le diagramme à barres en X est généralement plus instructif lorsqu'il existe un nombre raisonnable d'études (*par exemple,* 20 ou plus), tandis que le diagramme en I est utile pour un plus petit nombre d'études. Toutefois, il convient d'être prudent dans l'utilisation des graphiques en I, car les données peuvent ne pas être vraiment indépendantes, par exemple parce que les observations issues de la même étude peuvent avoir des réponses corrélées. Si un groupe se situe en dehors des limites d'un graphique à barres en X, il est probable que plusieurs valeurs individuelles du groupe se situent également en dehors des limites du graphique en I. En fin de compte, ces graphiques de contrôle doivent être considérés comme des outils complémentaires plutôt que comme des outils supérieurs les uns aux autres.

28. Les règles de décision peuvent être utilisées avec des graphiques de contrôle pour mettre en évidence lorsqu'un processus est hors de contrôle. Des règles codifiées sont utilisées pour garantir que les divers graphiques de contrôle sont interprétés de manière cohérente. Les règles de Western Electric et les règles de Nelson sont deux exemples largement utilisés qui peuvent mettre en évidence un processus incontrôlable ; elles sont fondées sur des caractéristiques non aléatoires des données [Nelson, 1984]. Par exemple, la règle de Nelson numéro 3 est « violée » lorsque six points de données (ou plus) d'une ligne augmentent ou diminuent continuellement.

29. Certains logiciels statistiques complètent les graphiques de contrôle par des indicateurs mis au point pour surveiller les processus de fabrication, qui quantifient la variabilité d'un processus. L'indice de stabilité (SI) [Jensen et al., 2019] est un exemple de données non groupées qui est calculé en divisant l'écart type à long terme (*c'est-à-dire* l'écart type total de la base de données) par l'écart type à court terme (*c'est-à-dire* la moyenne de la plage de variation divisée par d2). Des valeurs d'IS proches de 1.0 sont considérées comme stables avec une faible variabilité. Par conséquent, l'IS, ainsi que les règles (ou sous-ensembles) de Nelson ou Western Electric, fournissent des signaux d'alerte indiquant que certains aspects de l'essai sont passés ou sont en passe de basculer dans un état incontrôlable.

30. L'Analyse des Composantes de la Variance (ACV) est une autre catégorie de tests qui peut fournir des indications sur la qualité des DSH. Il existe diverses méthodes connexes, notamment l'ANOVA

© OCDE (2025)

hiérarchique ou imbriquée, les approches REML (Restricted Maximum Likability) et les méthodes de modélisation bayésiennes [Gelman et Hill, 2006]. Une autre stratégie, intitulée Evaluating the Measurement Process, a été décrite par Wheeler [2006]. Ces méthodes répartissent et estiment la proportion de la variation totale imputable à divers facteurs expérimentaux, ce qui donne une estimation de l'écart-type pour la source.

4.5 Calcul des intervalles décrivant les distributions des témoins négatifs historiques

31. Plusieurs approches valides permettent de caractériser la distribution des données relatives aux témoins négatifs historiques, et chaque laboratoire doit utiliser une méthode appropriée pour décrire ses données. Il s'agit généralement de calculer une limite supérieure et/ou inférieure afin de définir la fourchette dans laquelle la plupart des valeurs d'animaux témoins négatifs devraient se situer. Des considérations essentielles sont notamment la taille de l'échantillon et le fait que les données soient normalement distribuées. On trouvera ci-dessous un bref aperçu des méthodes de calcul de la limite supérieure et/ou inférieure, ainsi que des informations plus détaillées dans la littérature [Vardeman, 1992 ; InfinityQS International, 2014 ; et Kluxen et al., 2021].

4.5.1 Méthodes inappropriées

- Plage : la plage est généralement basée sur toutes les valeurs des témoins négatifs et correspond à la différence entre les valeurs minimale et maximale observées. Elle ne décrit *pas de manière adéquate* la distribution des témoins négatifs historiques en vue de l'établissement de limites inférieure et/ou supérieure pertinentes. En effet, la plage s'élargit à mesure que le nombre d'échantillons augmente, et peut dépendre de deux valeurs extrêmes (inhabituelles/aberrantes). Une large plage « récompense » parfois un laboratoire peu performant en masquant les résultats potentiellement néfastes de l'étude.
- Intervalle de confiance : un intervalle de confiance est une plage de valeurs estimées susceptible d'inclure une valeur (un paramètre) de la population, telle que la moyenne, avec un certain degré de confiance défini (95 %, par exemple). Bien que les définitions formelles impliquent la prise en compte d'échantillons répétés à partir de populations hypothétiques, celles-ci sont souvent difficiles à appliquer dans des contextes réels. Il est important de noter que les intervalles de confiance *ne permettent pas* de décrire de manière adéquate la distribution des témoins négatifs historiques lors de l'établissement d'une limite supérieure pour le critère c. En effet, les intervalles de confiance se resserrent à mesure que la taille de l'échantillon augmente, ce qui reflète la précision accrue de l'estimation de la valeur (du paramètre) de la population, et ne se prête donc *pas à l'*interprétation des données du critère c de l'OCDE.

4.5.2 Méthodes appropriées

32. Les méthodes de calcul d'un intervalle décrivant la distribution des données relatives aux témoins négatifs historiques sont énumérées ci-dessous. Une base de données de ≥ 100 rongeurs/points de données issus de ≥ 20 expériences indépendantes est préférable. Toutefois, des intervalles plus petits peuvent suffire, en particulier si les données sont normalement distribuées ou peuvent être normalisées, et si l'essai est bien contrôlé. On donnera moins de poids à des intervalles issus de petits ensembles de données, qu'à ceux issus des bases de données plus grandes, car ces premiers sont moins susceptibles de couvrir la proportion souhaitée.

33. Bien que les données des témoins historiques puissent se situer dans ces intervalles, cela ne signifie pas que les données sont appropriées pour les évaluations de type c. Une variabilité excessive peut entraîner des valeurs limites/d'intervalle trop larges. Cela peut indiquer une performance médiocre du

© OCDE (2025)

laboratoire et ne doit pas être confondu avec la « variabilité biologique normale » de l'essai. Cet avertissement s'applique à tous les intervalles décrits ci-dessous et souligne la recommandation d'utiliser des méthodes telles que les règles de Nelsen ou de Western Electric, SI et/ou ACV pour évaluer la qualité des données des témoins historiques avant de calculer les intervalles de distribution.

- Limite de contrôle : dans le domaine du contrôle de la qualité, on utilise comme limites de contrôle des valeurs situées, de part et d'autre de la moyenne, à un intervalle correspondant à un multiple de l'écart type (habituellement 3 fois l'écart type). Représentées par des lignes tracées sur un graphique de contrôle, elles peuvent être accompagnées d'autres valeurs multiples de l'écart-type, égales à ± 2 fois l'écart-type par exemple, appelées limites d'alarme. En conjonction avec les graphiques de contrôle, les limites de contrôle et d'alarme sont des outils précieux pour évaluer le degré de maîtrise d'un processus ou d'un essai réalisé de façon répétitive. Lors de la configuration des graphiques de contrôle, il convient de tenir compte du fait que l'écart type utilisé n'est pas le même que l'écart type par rapport aux statistiques d'échantillon récapitulatives. En multipliant cet écart-type par 2 ou 3, on obtient des valeurs différentes de celles utilisées pour créer les limites d'avertissement et de contrôle. Cet écart s'explique par le fait que les logiciels statistiques estiment souvent l'écart-type à l'aide de la « plage de variation moyenne », car l'écart-type réel de la population est inconnu et doit être estimé à partir des données. L'estimation de l'écart-type par échantillonnage est ajustée en conséquence afin de minimiser les biais ou les erreurs. Dans l'hypothèse d'une distribution normale et d'un processus « sous contrôle », 99.73 % des données des témoins négatifs doivent se situer dans les limites de trois écarts types de la moyenne, et près de 95 % dans les limites de deux écarts types. Les limites de contrôle et/ou d'alarme sont utiles pour évaluer les données relatives aux témoins négatifs historiques, et pour fournir des limites supérieure et/ou inférieure qui aident à interpréter les données. Lors de l'examen des données d'étude dans le contexte des évaluations de type c du critère, les limites d'avertissement qui décrivent où se situent 95 % environ des données en cas de distribution normale (± 2 écarts types) sont généralement plus appropriées que les limites fondées sur 3 écarts types. Les premières sont cohérentes avec de nombreuses Lignes directrices (LD 474 de l'OCDE, par exemple), tandis que les secondes caractérisent des points de données extrêmement rares/inhabituels, générant des intervalles trop larges pour le cas d'utilisation du critère c. Cependant, des limites de contrôle basées sur 3 écarts-types peuvent être utiles pour les critères d'acceptabilité des données.
- Intervalles de prédiction : les intervalles de prédiction sont conçus pour permettre de prédire une ou plusieurs observation(s) future(s) à partir de données existantes. Ainsi, avec un intervalle de prédiction à 95 %, la probabilité qu'un nouveau résultat (ou un nombre spécifié d'observations futures) d'un produit chimique négatif, se situe dans cette plage est de 95 %. Les données non normales doivent être transformées si nécessaire avant de calculer l'intervalle, qui peut être à nouveau transformé en unités d'origine pour utilisation et notification. Alternativement, certains logiciels informatiques permettent un calcul non paramétrique qui ne suppose pas une distribution normale. Les intervalles de prédiction correspondent généralement à un *nombre* spécifié d'observations futures, généralement une ou plusieurs, plutôt qu'à un pourcentage des observations futures. Il peut donc être difficile d'utiliser un intervalle prédictif pour juger de la validité de l'étude aux fins du critère c, à moins d'accompagner le calcul d'une justification du choix d'un nombre particulier d'observations futures.
- Intervalles de tolérance : les intervalles de tolérance, comme les intervalles de prédiction, sont prospectifs mais sont conçus pour prédire des observations futures avec des niveaux de couverture et de confiance spécifiés. La couverture fait

© OCDE (2025)

référence au pourcentage d'observations futures définies par les limites inférieure et supérieure calculées (*p. ex.,* 95 % ou 99 %), modélisées avec un niveau de confiance spécifié. Si les intervalles de tolérance sont souvent calculés avec un degré de confiance de 95 %, ce paramètre peut être défini par l'utilisateur. Les considérations relatives à la normalité et les approches non paramétriques facultatives examinées pour les intervalles de prédiction s'appliquent également aux intervalles de tolérance. En général, les intervalles de tolérance sont plus larges que les intervalles de prédiction car ils sont conçus pour prédire un pourcentage élevé de valeurs futures, tandis que les intervalles de prédiction se concentrent généralement sur l'estimation d'une ou de quelques nouvelles observations.

- Quantiles : les quantiles fournissent un aperçu du rang des points de données selon leur valeur, indépendamment de toute hypothèse de distribution des probabilités. Les quantiles sont largement utilisés dans les applications biomédicales, où la non-normalité due aux valeurs aberrantes et/ou à l'asymétrie est courante. Par exemple, les intervalles basés sur les percentiles sont souvent utilisés pour aider à interpréter des résultats d'essai. Il est possible de calculer les intervalles de confiance des quantiles pour estimer l'incertitude autour de leur mesure. Cela peut aider à évaluer la qualité de l'ensemble de données sous-jacent. Les intervalles de confiance des quantiles seront particulièrement larges en queue de distribution, à moins que la taille de l'échantillon soit importante.

4.6 Utilisation des données des témoins historiques pour interpréter les données de l'étude

34. Les données de contrôle historiques de faible qualité ne sont pas adaptées aux évaluations de type c. Dans cette optique, la présente section décrit l'utilisation des données des témoins historiques pour interpréter les données de l'étude et repose sur l'hypothèse que la base de données est suffisamment grande et de qualité acceptable à cette fin.

35. Une approche courante pour remplir le critère c consiste à mener un grand nombre d'études, puis à utiliser les valeurs moyennes de l'étude pour calculer un intervalle de distribution des témoins historiques (ou bien simplement la limite supérieure de la valeur). Dans ce cas, la « comparaison d'éléments comparables » mentionnée ci-dessus se traduit par une évaluation simple. Autrement dit, pour toute étude particulière, les valeurs moyennes des groupes de traitement sont comparées à la valeur limite supérieure (dérivée des moyennes historiques des témoins négatifs) qui sépare les valeurs « attendues » des valeurs « élevées ». Il s'agit là d'une approche valable et utile, mais il faut savoir qu'une telle stratégie nécessite de nombreuses études distinctes et, par conséquent, un grand nombre de rongeurs, et que l'utilisation d'animaux de laboratoire à cette fin et à d'autres fins devient plus difficile dans de nombreuses régions du monde.

36. Comme on l'a vu précédemment, il est également possible d'utiliser des données individuelles relatives à chaque animal pour évaluer la qualité des données des témoins historiques, et de calculer un intervalle utile (ou bien simplement une valeur limite supérieure) qui peut être utilisé aux fins du critère c (du moins jusqu'à ce qu'un nombre suffisant d'études ait été réalisé pour justifier un intervalle basé sur la moyenne de l'étude). Cependant, une fois que l'on est parvenu à une limite supérieure séparant les valeurs attendues des valeurs élevées, il peut être moins évident de déterminer comment évaluer les résultats de l'étude sur une base individuelle. Dans cette situation, faire appel à un raisonnement scientifique solide et bien documenté.

37. Lorsqu'un raisonnement scientifique rationnel est utilisé pour évaluer des données individuelles relatives à chaque animal par rapport à une valeur limite supérieure convenablement dérivée, au moins quatre facteurs doivent être pris en compte. Premièrement, combien d'animaux et combien d'études ont été utilisés pour calculer l'intervalle ou la limite supérieure? Par exemple, une base de données et une

© OCDE (2025)

limite supérieure fondées sur 10 études et 50 animaux devraient avoir un poids plus élevé que 5 études et 25 animaux. Deuxièmement, combien de réponses des animaux traités avec le produit chimique d'essai ont été supérieures à la limite supérieure ? En d'autres termes, plus le nombre est élevé, plus la préoccupation est grande. Troisièmement, quel(s) groupe(s) de traitement a/ont présenté des valeurs élevées ? Par exemple, toutes les valeurs élevées apparaissant dans le groupe à faible dose peuvent être le signe d'un problème technique par opposition à un effet lié au traitement. Quatrièmement, quelle a été l'ampleur de la ou des réponses élevées? Autrement dit, des valeurs très élevées seraient plus préoccupantes que des valeurs qui dépassent à peine la limite supérieure.

38. Il ne s'agit pas d'une liste exhaustive de facteurs à prendre en compte. Cette section se concentre plutôt sur l'interprétation des réponses individuelles élevées des animaux par rapport à une limite supérieure, et sur la façon dont elles peuvent contribuer à la contextualisation des analyses statistiques (*c'est-à-dire les* critères a et b). D'autres aspects importants, tels que la distribution et le métabolisme d'une substance d'essai, la cinétique d'apparition et de disparition de l'effet mesuré, etc., sont aussi à prendre en compte lors de l'évaluation de la plausibilité biologique.

5. Interprétation des résultats de l'étude

39. Les approches statistiques décrites dans les sections précédentes sont considérées comme des outils représentatifs pour évaluer si le produit chimique d'essai induit une génotoxicité. Cela étant dit, un jugement scientifique solide doit jouer un rôle de premier plan dans l'examen de tout ensemble de données. Cela est particulièrement évident lorsque les analyses statistiques présentent des résultats qui, à première vue, peuvent paraître contradictoires. Ce paradigme a été renforcé par un groupe de travail de l'OCDE sur la génotoxicité qui s'est réuni à Ottawa, Canada, 2013 [OCDE, 2017].

40. Dans certains cas, même après avoir appliqué un jugement scientifique, avoir utilisé des outils statistiques appropriés et éventuellement avoir évalué des données supplémentaires, il ne sera pas possible de déterminer une réponse comme positive ou négative. Dans ce cas, la réponse est équivoque et des essais supplémentaires peuvent être nécessaires pour déterminer la génotoxicité du produit chimique testé.

6. Tableau représentatif des témoins négatifs historiques

41. Comme indiqué dans la section du rapport d'essai, les rapports d'étude doivent fournir des informations sur la manière dont les données relatives aux témoins négatifs historiques ont été collectées, évaluées et utilisées. Le tableau suivant est présenté à titre indicatif.

© OCDE (2025)

	Information	Description and Commentaires
1. Description générale	Période	Période au cours de laquelle les DCH ont été collectées.
	Système d'essai	Critère d'évaluation étudié et modèle utilisé. Dans le cas d'une étude in vivo, il peut s'agir de l'espèce, de la souche, du sexe et de la tranche d'âge des animaux au moment de l'initiation de la dose et du prélèvement des tissus, ainsi que des procédures de randomisation ou de blocage utilisées.
	Système analytique	Méthode de collecte des données.
2. Indicateurs de base	Nombre d'études et d'observations	Nombre d'études à partir desquelles les données ont été recueillies, et nombre total d'observations utilisées dans les DCH, y compris une mention précisant si les observations sont des moyennes individuelles ou collectives.
	Médiane, moyenne** et Std. Dev.	Médiane et moyenne des DCH, avec son écart-type. **D'autres mesures de la tendance centrale peuvent être utilisées.
	Plage observée	Valeurs minimale et maximale observées dans les DCH.
	Structures du groupe	Décrire les raisons justifiant la combinaison ou la séparation des bases de données des témoins historiques en fonction de facteurs expérimentaux tels que le sexe, véhicules, voies d'administration, groupes d'âge, quantité de données disponibles, etc.
3. Indicateurs de qualité	Transformation	Une description de l'analyse effectuée pour sélectionner la transformation et justification de la transformation finale sélectionnée. Visualisation des DCH transformées est recommandée.
	Graphique de contrôle	Un affichage graphique longitudinal des DCH, qui inclut le type de graphique de contrôle présenté et affiche les limites d'avertissement et de contrôle.
	Variabilité au cours du temps	Une évaluation de la stabilité des DCH sur le temps, y compris une brève description des méthodes utilisées.
	Sources de variation	Une évaluation de la principale source de variabilité de l'indice DCH, y compris une brève description des méthodes utilisées.
	Intervalle de distribution	Intervalle décrivant la distribution des DCH, ainsi qu'une brève description de la méthode utilisée pour calculer l'intervalle et une justification du choix de l'intervalle spécifique.

*Compte tenu de la quantité et du type d'informations demandées ci-dessus, une suggestion pratique serait que les sponsors de l'étude fournissent ces données dans une annexe ou une annexe à un rapport d'étude réglementaire. Elle resterait stable pendant un certain temps, jusqu'à ce que la base de données des témoins négatifs historiques ait été mise à jour avec de nouvelles données/distributions.

© OCDE (2025)

© OCDE (2025)

Références de l'Annexe 2, par ordre d'apparition

Dertinger SD, D Li, C Beevers, GR Douglas, RH Heflich, DP Lovell, DJ Roberts, R Smith, Y Uno, A Williams, KL Witt, A Zeller, C Zhou (2023) Assessing the quality and making appropriate use of historical negative control data: A report of the International Workshop on Genotoxicity Testing (IWGT). Environ Mol Mutagen, in press, doi: 10.1002/em.22541.

OECD (Organisation for Economic Cooperation and Development). (2017) Overview of the set of OECD Genetic Toxicology Test Guidelines and updates performed in 2014-2015. Series on Testing and Assessment, No. 238 – 2nd edition. Available at: https://www.oecd.org/en/publications/overview-of-the-set-of-oecd-genetic-toxicology-test-guidelines-and-updates-performed-in-2014-2015-second-edition_61eca5cd-en.html.

Bretz F, LA Hothorn (2003) Statistical analysis of monotone or non-monotone dose-response data from in vitro toxicological assays. ATLA-Altern Lab Animal 31(Suppl 1):81-96.

Hayashi M, K Dearfield, P Kasper, D Lovell, H-J Martus, V Thybaud (2011) Compilation and use of genetic toxicity historical control data. Mutat Res 723:87-90.

Igl B-W, A Bitsch, F Bringezu et al. (2019) The rat bone marrow micronucleus test: Statistical considerations on historical negative control data. Regul Toxicol Pharmacol 102:13-22.

Box GEP, DR Cox (1964) An Analysis of Transformations, *Journal of the Royal Statistical Society: Series B (Methodological)*, 26(2):211-243.

LS Nelson (1984) The Shewhart Control Chart—Tests for Special Causes. Journal of Quality Technology, 16(4):237-239.

WA Jensen, J Szarka III, J., K White (2019). Stability assessment with the stability index. Quality Engineering, 31(2):289-301.

A Gelman, J Hill (2006) Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge: Cambridge University Press.

DJ Wheeler (2006) EMP III (Evaluating the Measurement Process) Using Imperfect Data. SPC Press, Knoxville, TN.

Vardeman SB (1992) What about the other intervals? The American Statistician 46:193-197.

InfinityQS International (2014) A practical guide to selecting the right control chart. pp 1-21. Available at: http://www.infinityqs.com/sites/infinityqs.com/files/files/PDFs/InfinityQS_Practical_Guide_to_Selecting_the_Right_Control_Chart_Oct2013.pdf.

Kluxen FM, K Weber, C Strupp, SM Jensen, LA Hothorn, J-C Garcin, T Hofmann (2021) Using historical control data in bioassays for regulatory toxicology. Reg Toxicol Pharmacol 125:105024.

OECD (Organisation for Economic Cooperation and Development) (2016) OECD Guideline for the testing of chemicals: Mammalian erythrocyte micronucleus test. Test Guideline 474. Available at: https://www.oecd.org/en/publications/2016/07/test-no-474-mammalian-erythrocyte-micronucleus-test_g1g6fb16.html.

© OCDE (2025)