

Les modèles animaux ont-ils une valeur prédictive pour l'homme ?

Niall Shanks¹, Ray Greek^{*2} et Jean Greek²

Adresses :

¹Wichita State University, Faculté d'histoire, 1845 N Fairmont, Fiske Hall, Wichita KS 67260, É.-U. ;

²Americans For Medical Advancement, 2251 Refugio Rd Goleta, CA 93117, É.-U.

Courriels :

niall.shanks@wichita.edu ; drraygreek@aol.com ; jeangreek@aol.com

* Auteur correspondant

Publié le 15 janvier 2009 dans *Philosophy, Ethics, and Humanities in Medicine* 2009, 4:2 doi:10.1186/1747-5341-4-2
(Reçu le : 23 juillet 2008 ; accepté le : 15 janvier 2009)

Cet article est disponible à l'adresse suivante : <http://www.peh-med.com/content/4/1/2>

© 2009 Shanks et al ; titulaires d'une licence BioMed Central Ltd.

Le présent article Open Access est distribué selon les termes de la licence Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), qui autorise son utilisation, sa distribution et sa reproduction sur tout support à condition de citer correctement le contenu d'origine.

Résumé

L'un des principaux objectifs de la philosophie des sciences est d'éclaircir la signification des termes scientifiques, et de soumettre leur emploi à un examen critique. Le présent article s'intéresse au terme scientifique *prédire*, et tente de déterminer s'il existe des preuves crédibles démontrant que les modèles animaux peuvent être utilisés pour prédire des résultats chez l'être humain, en particulier en toxicologie et pathophysiologie. La question de savoir si les animaux peuvent être utilisés pour prédire les réactions des êtres humains aux médicaments et autres produits chimiques apparaît comme un sujet très controversé. Or, l'analyse empirique des modèles animaux au moyen d'outils scientifiques démontre l'impuissance de ces modèles à prédire les réactions chez l'homme. Ce résultat n'est pas une surprise à la lumière de ce que nous ont appris des disciplines telles que la biologie de l'évolution et la biologie développementale, la régulation et l'expression géniques, l'épigénétique, la théorie de la complexité et la génomique comparative.

Revue

« Quand, moi, j'emploie un mot, déclara le Gros Coco d'un ton assez dédaigneux, il veut dire exactement ce qu'il me plaît qu'il veuille dire... ni plus, ni moins.

- La question est de savoir, répondit Alice, si vous pouvez obliger les mots à vouloir dire des choses différentes. »

Lewis Carroll. *De l'autre côté du miroir*. 1871

(trad.Jacques Papy pour les éditions Jean-Jacques Pauvert)

La controverse scientifique fait rage au sujet de la valeur prédictive des modèles animaux. Dans cet article, nous utiliserons l'expression *modèle animal* (et, dans ce contexte, le terme *animal*) pour signifier l'utilisation d'un animal non humain, le plus souvent un mammifère ou un vertébré, pour prédire la réaction de l'homme aux médicaments et aux maladies. Nous reconnaissons sans retenue que les animaux peuvent être utilisés avec succès dans de nombreux domaines scientifiques, notamment en recherche fondamentale, en tant que dispositifs de remplacement pour les êtres humains, bioréacteurs, etc. Nous excluons ces utilisations de notre définition et de notre analyse critique car aucune revendication n'est faite quant à leur valeur prédictive. Le présent article s'intéresse uniquement à l'utilisation des animaux et modèles animaux dans le but de prédire les réactions humaines, à la lumière de la signification du terme *prédire* dans le domaine des sciences.

Un aperçu de la philosophie des sciences

La signification des mots revêt une très grande importance dans tous les domaines d'études, mais plus particulièrement dans les domaines scientifiques.

Les philosophes des sciences, parmi lesquels Quine, Hempel et d'autres, ont fait valoir que les mots doivent avoir une signification du point de la vue de la science, et que ces significations séparent de fait la science de la pseudoscience. Prenons l'exemple du terme *prédiction*. Une méthode de recherche n'a pas besoin d'être prédictive pour être utilisée, mais revendiquer la capacité prédictive d'un essai ou d'un projet signifie quelque chose de très spécifique.

Le présent article concerne l'utilisation du terme *prédire* appliqué aux modèles animaux. Nous pensons que la signification de ce terme a été galvaudée et que par conséquent, le concept qui le sous-tend est en danger, ainsi que l'ensemble des implications de ce concept. Au-delà d'un simple mot, *prédire* est surtout un concept étroitement lié à la notion d'*hypothèse*. On peut définir la notion d'hypothèse comme l'explication proposée d'un phénomène donné, observé ou imaginé, explication dont la validité doit être soumise à l'épreuve de l'essai. Selon Sarewitz et Pielke :

Dans la société moderne, la prédiction sert deux objectifs importants. Premièrement, la prédiction est un test de la compréhension scientifique, ce qui lui confère autorité et légitimité. Les hypothèses scientifiques sont soumises à l'épreuve de l'essai en comparant ce qui est attendu à ce qui se produit réellement. Lorsque les attentes et les faits observés coïncident, la capacité de la compréhension scientifique à expliquer le fonctionnement des choses s'en trouve justifiée. « Prédire des faits inconnus est essentiel au processus de vérification empirique des hypothèses, qui constitue la caractéristique la plus remarquable de toute démarche scientifique », observe quant à lui le biologiste Francisco Ayala (Ayala, F. 1996. The candle and the darkness *Science* 273:442.) [1] (Souligné par nous)

Dans le cas des modèles animaux, *ce qui se produit réellement* est ce que l'on observe chez l'être humain. Si l'objectif de l'essai, qu'il s'agisse d'un essai sur l'animal ou d'un essai *in silico*, est de prédire la réaction chez l'homme, alors la capacité de cet essai à se conformer à la réaction chez l'homme doit être évaluée. Ici encore, nous reconnaissons que les essais et les travaux impliquant des animaux ne sont pas tous réalisés dans un but prédictif. Toutefois, les essais qualifiés de prédictifs doivent être évalués pour juger de leur capacité à effectivement prédire la réaction chez l'homme.

Sarewitz and Pielke poursuivent ainsi :

Deuxièmement, la prédiction peut aussi servir d'aide potentielle à la prise de décision. Nous pouvons chercher à connaître l'avenir en se fondant sur la croyance que cette connaissance aura pour effets de stimuler et rendre possibles des actions bénéfiques dans le présent [1].

Revenons à la notion de *prise de décision*.

Le philosophe W.V.O. Quine remarquait :

Une prédiction peut s'avérer vraie ou fausse, mais dans les deux cas, elle comprend la notion de diction : c'est-à-dire qu'elle doit être dite ou, par extension, écrite. L'étymologie et le dictionnaire s'accordent sur ce point. Les synonymes les plus proches, « prévision », « connaissance anticipée » et « prescience » s'affranchissent de cette limite, mais sont sujettes à d'autres. La connaissance anticipée se doit d'être vraie, c'est-à-dire infaillible. D'un point de vue étymologique, la prévision est limitée au domaine visuel – au-delà elle reste vague. « Prescience » implique une clairvoyance (...) La prédiction est ancrée dans une tendance générale, chez les vertébrés supérieurs, à penser que des expériences semblables ont des conséquences similaires [[2]159] (...) (Souligné par nous.)

Les prédictions, générées par les hypothèses, ne sont pas toujours correctes. Mais une modalité, un essai ou une méthode qualifiés de prédictifs se doivent de donner la réponse correcte un pourcentage très élevé de fois dans le domaine de la biologie, et à chaque fois dans le domaine de la physique. (Nous reviendrons plus loin sur ce point et son application à la biologie).

Lorsqu'une modalité échoue régulièrement à former des prédictions exactes, alors cette modalité ne peut pas être qualifiée de prédictive simplement parce que de temps à autre, elle prévoit une réponse correcte. Ce qui précède distingue l'usage scientifique du terme *prédire* de son emploi usuel, qui est plus proche de *prévoir*, *deviner*, *conjecturer*, *projeter* et autres. Nous reviendrons un peu plus loin sur ce point.

De nombreux philosophes des sciences pensent qu'une théorie (et nous ajoutons : une modalité) pourrait être confirmée ou infirmée en vérifiant les prédictions qu'elle énonce. Contrairement à la physique, la biologie, qui étudie des systèmes complexes, doit s'appuyer sur les statistiques et les probabilités lorsqu'elle cherche à déterminer ce que sera la réaction à un stimulus ou qu'elle évalue la probabilité qu'a un phénomène de se produire. Il existe, chez les partisans des modèles animaux, un postulat qui a pris les traits d'une théorie, ou peut-être une hypothèse généralisée selon laquelle les résultats des expériences sur les animaux peuvent être appliqués directement à l'homme ; que les modèles animaux sont prédictifs. Il en a résulté une démarche méthodologique incontestée : utiliser les animaux en tant qu'humains de substitution. L'ironie

veut que cette hypothèse n'a jamais été questionnée comme devraient l'être les hypothèses en science, d'où le qualificatif choisi d'hypothèse *généralisée*. Le fait que le modèle animal en tant que tel puisse être utilisé pour prédire les réactions chez l'homme *peut* être testé, et si les résultats montrent une sensibilité, une spécificité, une valeur prédictive positive ou négative suffisamment élevées, alors l'hypothèse selon laquelle les animaux peuvent prédire les réactions chez l'homme serait vérifiée. Dans ce cas, on pourrait affirmer que les modèles animaux ont une valeur prédictive pour l'homme ; dans le cas contraire, on pourrait alors affirmer que les modèles animaux n'ont pas de valeur prédictive pour l'homme.

Les animaux et les hypothèses sont utilisés de deux façons très différentes dans le domaine des sciences. Les hypothèses engendrent des prédictions qui doivent être testées pour confirmer ou infirmer l'hypothèse de départ. Prenons le cas d'un chercheur qui utilise des animaux dans le cadre de recherches fondamentales. À la fin de la série d'expériences sur l'animal, le chercheur dispose tout au plus d'une hypothèse sur une réaction probable chez l'homme vis-à-vis du même stimulus ou de la même substance, compte tenu des différences de poids corporel, d'exposition, etc. La prédiction issue de l'hypothèse doit à son tour être testée, ce qui nécessite de rassembler des données chez l'homme. La prédiction peut être confirmée ou infirmée par ces données humaines, mais la charge de la preuve s'impose, du point de vue de la méthodologie scientifique de base. À aucun moment de cette utilisation des animaux aux fins de générer une hypothèse, ceux-ci n'ont été considérés comme prédictifs. LaFollette et Shanks ont nommé cette façon d'utiliser les animaux des « modèles animaux heuristiques ou hypothétiques » (HAM) [3,4].

Ceci contraste avec l'hypothèse de départ prise par certains scientifiques, à savoir que les animaux sont prédictifs pour les humains (voir le tableau 1). En se fondant sur cette hypothèse, ces scientifiques affirment que les médicaments et les produits chimiques qui auraient été nocifs pour les humains ont été écartés du marché suite aux résultats des essais menés sur les animaux. Cette affirmation est hypocrite sauf si nous avons *a priori* raison de considérer que les modèles animaux ont une valeur prédictive. Or dans les cas cités, cette hypothèse n'a jamais été testée. Dans de nombreux cas, il ne serait pas éthique de conduire, sur des êtres humains, les types d'études de laboratoire soigneusement contrôlées régulièrement menées sur des animaux, notamment les rongeurs. Il existe toutefois d'autres moyens éthiques d'obtenir des données chez l'homme dans le contexte de l'épidémiologie (notamment des études épidémiologiques rétrospectives et prospectives) : les études *in vitro* sur tissus humains, les études *in silico* et l'avancée technologique majeure qu'a représenté le microdosage [5]. Il ne faut en outre jamais oublier que lorsque des produits chimiques industriels se retrouvent dans l'environnement, ou que des médicaments sont vendus au grand public, des expériences sur l'homme ont déjà eu lieu. De plus, comme l'observe Altman [6], il existe de nombreux exemples récents et passés dans lesquels les chercheurs, qui doutaient de l'applicabilité ou de la pertinence des modèles animaux aux situations humaines, ont mené des expériences sur eux-mêmes – une pratique encore d'actualité comme le note Altman (Barry Marshal, récent lauréat du prix Nobel, pour n'en citer qu'un). Dans tous les cas, une réussite prouvée (en termes de valeurs prédictives positives et négatives) de l'utilisation de modèles animaux spécifiques représente de toute évidence le strict minimum pour que la société accepte le caractère prédictif, pour l'homme, d'hypothèses émanant d'essais sur des animaux.

Tableau 1 : La notion d'hypothèse

	Modèles animaux en toxicologie et recherche médicale	Utilisation correcte des modèles animaux (rech. fondamentale, par exemple)
Postulats	Postulats habituels + nature prédictive des modèles animaux	Postulats habituels (par ex. « il existe des lois universelles »)
Hypothèse	X induit Y.	Aucune, ou X induit Y.
Essai sur l'animal	Dans l'essai sur l'animal, X induit Y	X induit Y, d'après les essais sur l'animal
Hypothèse		X induit Y chez l'homme
Test		On applique X chez l'homme, puis on voit si cela induit Y ; ou bien, on étudie des populations où X s'applique, en vérifiant les résultats.
Conclusion	X induisant (ou n'induisant pas) Y chez l'animal, on en infère que X induit (ou n'induit pas) Y chez l'homme.	Résultats obtenus chez l'homme.

Nous devons donc insister sur le fait que lorsque nous nous interrogeons sur les animaux en tant que modèles prédictifs, c'est l'hypothèse généralisée selon laquelle les animaux sont prédictifs qui est concernée, et non l'utilisation des animaux pour générer des hypothèses, comme c'est le cas en recherche fondamentale. Ce qui amène à s'interroger sur la notion de preuve et sur qui repose, en science, la charge d'établir les preuves. Comme dans le domaine juridique, c'est au requérant d'établir la preuve. L'hypothèse nulle veut que des événements ne sont pas considérés comme liés tant qu'un lien de

causalité n'a pas été démontré entre eux. Ainsi, ceux qui affirment que les modèles animaux permettent de prédire la réaction chez l'homme dans le domaine de la recherche biomédicale doivent prouver que ce qu'ils avancent est vrai. Ce n'est pas à nous de prouver que les modèles animaux, utilisés par exemple en carcinogénèse ou toxicité, ne sont *pas* prédictifs. Il incombe aux partisans du caractère prédictif des modèles animaux de faire la démonstration qu'ils le sont, ce qui nécessite de considérer ce que les preuves montrent réellement.

Contrairement à la physique, qui s'occupe de systèmes simples auquel le réductionnisme peut être appliqué, la biologie n'a pas toujours ce luxe. Il existe des domaines de la biologie, l'anatomie comparative par exemple, où les principes d'échelle peuvent véritablement être appliqués. Mais la biologie n'est pas la physique, et dans d'autres domaines, par exemple certaines branches de la toxicologie prédictive, l'utilisation de ces facteurs d'échelle (par exemple le poids corporel²⁹) s'est avérée moins utile, pour des raisons que nous détaillerons plus bas. Les systèmes biologiques suivent bien entendu les lois de la physique, mais ils présentent des propriétés dues à leur organisation interne et, au final, à l'histoire de l'évolution, que l'on ne retrouve pas dans le domaine de la physique. Ceci signifie qu'un même stimulus peut donner lieu à des résultats finaux présentant des différences marquées. La réaction de différents individus à un même médicament ou une même maladie est un exemple bien connu de ce phénomène [7-13]. Il existe cependant des manières de juger de la nature prédictive d'essais, même dans les systèmes complexes biologiques. Des valeurs telles que la valeur prédictive positive, la sensibilité, la spécificité et la valeur prédictive négative (nous reviendrons brièvement sur ces valeurs) peuvent être calculées pour confirmer ou infirmer des hypothèses. Les essais à visée prédictive dont les valeurs correspondraient à une réponse aléatoire n'appartiennent bien évidemment pas à la catégorie des essais *prédictifs*

Arguments avancés par les partisans du caractère prédictif des modèles animaux

Selon Salmon, il existe au moins trois motivations pour effectuer des prédictions :

1. connaître ce qui va se passer dans l'avenir ;
2. tester une théorie ;
3. décider d'une action, le meilleur moyen pour cela étant de prédire l'avenir [14].

Dans le cas de la carcinogénèse nous voulons savoir : (1) ce qui se produira dans l'avenir (ce produit chimique causera-t-il des cancers chez l'homme ?) et (3) quelle décision prendre (autoriser la mise sur le marché du produit concerné, ou non ?), la meilleure manière de prendre cette décision étant de pouvoir prédire l'avenir. Ni (1), ni (3) ne sont subtils. Nous voulons une réponse correcte à la question : « ce produit chimique est-il cancérigène pour l'être humain ? », ou à des questions similaires telles que « quels effets aura tel médicament chez l'être humain ? » ou « ce médicament est-il tératogène ? », ou encore « ce récepteur est-il celui utilisé par le VIH pour pénétrer dans les cellules humaines ? » Mais, comme nous l'avons vu, deviner correctement ou trouver des corrélations ne sont pas la même chose que prédire une réponse. Et, comme nous le verrons plus loin, un degré de sensibilité élevé ne constitue pas à lui seul une prédiction.

Ce qui suit aidera le lecteur à mieux percevoir les contours de ce débat scientifique.

Butcher [15], Horrobin [16], Pound et al. [17] et d'autres [3,4,18-24] se sont interrogés sur la valeur de l'emploi d'animaux pour prédire les réactions chez l'homme. Dans tous les cas, la prédiction demeure un problème. Le secrétaire d'État américain à la santé, Mike Leavitt, déclarait en 2007 :

À l'heure actuelle, neuf médicaments expérimentaux sur dix échouent au stade des essais cliniques, car nous sommes incapables de prédire précisément leurs effets chez les êtres humains à partir des études en laboratoire et sur les animaux [24].

Cette déclaration est une très grosse pierre dans le jardin de ceux qui affirment que les animaux sont prédictifs. Les faits derrière cette déclaration suffiraient amplement pour répondre, sans arguments supplémentaire, à la question de la prédiction. Mais nous allons aller plus loin.

Le débat s'est récemment étendu à la revue *Philosophy, Ethics, and Humanities in Medicine*. Knight [25] a récemment remis en question l'utilisation de chimpanzés dans les recherches biomédicales en citant, entre autres raisons, leur manque de prédictibilité. Shanks et Pyles [26] se sont interrogés sur la capacité des animaux à prédire les réactions chez l'homme ; Vineis et Melnick [27] leur ont répondu que les animaux peuvent être employés pour prédire la réaction des êtres humains à des substances chimiques, en citant la carcinogénèse, et que l'épidémie de cancer aurait pu être évitée si les données animales avaient été utilisées pour réduire l'exposition des êtres humains au produit concerné, ou interdire totalement ce produit. Cette affirmation selon laquelle les animaux peuvent prédire les réactions chez l'homme n'est pas la seule [28,29]. Gad écrivait en 2007, dans *Animal Models in Toxicology* :

L'emploi, dans le domaine des sciences biomédicales, d'animaux en tant que modèles afin d'aider à comprendre et prédire les réactions chez l'homme, en particulier en toxicologie et pharmacologie, demeure à la fois le principal outil de progrès dans le domaine biomédical et une source de controverse importante (...)

Quoi qu'il en soit, bien que des cas de mauvaises pratiques soient avérés, il est indéniable que bien utilisés, les animaux fonctionnent exceptionnellement bien en tant que modèles prédictifs pour l'être humain (...)

Qu'elles permettent de disposer d'organelles, de cellules ou de tissus isolés, ou qu'elles jouent un rôle de modèle pour les maladies ou de prédiction de l'action ou de la transformation d'un médicament ou autre xénobiotique chez l'homme, les expériences sur les animaux sont à la base de l'explosion des connaissances médicales et biomédicales dans la seconde moitié du 20^e siècle et en ce début de 21^e (...)

Les animaux sont utilisés depuis des siècles comme modèles pour prédire ce que les produits chimiques et les facteurs environnementaux provoqueraient chez les êtres humains... Le recours aux animaux en tant prédicteurs des effets néfastes potentiels n'a cessé de croître depuis [l'année 1792].

Les procédures d'essai actuelles (ou même celles en vigueur à l'époque aux États-Unis, où l'utilisation de ce médicament [le thalidomide] chez l'être humain n'a jamais été autorisée) auraient identifié le risque et permis d'éviter cette tragédie [29]. (Souligné par nous.)

Fomchenko et Holland remarquent ce qui suit :

Les SGM [souris génétiquement modifiées] reproduisent étroitement la maladie chez l'homme et sont utilisées pour prédire la réaction, chez l'homme, à une thérapie, un traitement ou un programme d'irradiation [30]. (Souligné par nous.)

Hau, qui a dirigé un ouvrage influent sur les recherches sur les animaux, écrit ce qui suit :

Un troisième groupe important de modèles animaux est utilisé à titre prédictif. Ces modèles sont utilisés dans le but de découvrir et de quantifier l'impact d'un traitement, qu'il s'agisse de guérir une maladie ou d'évaluer la toxicité d'une substance chimique [31].

Hau propose clairement d'utiliser les animaux en tant que modèles prédictifs, au sens où nous l'entendons précisément ici. La revendication du caractère prédictif est également forte lorsque le terme « prédiction » n'est pas employé mais qu'il est implicite, ou lié à la causalité. Fomchenko et Holland indiquent également :

L'utilisation de systèmes *in vitro* et de modèles de tumeurs cérébrales *in vivo* par xénogreffes constitue un moyen rapide et efficace de tester de nouveaux agents et de nouvelles cibles thérapeutiques, des connaissances qui peuvent ensuite être traduites et testées sur des SGM plus sophistiquées, qui reproduisent fidèlement les tumeurs cérébrales humaines et prépareront probablement la voie à des essais clinique de qualité élevée, avec des résultats de traitement satisfaisants et une moindre toxicité des médicaments. L'utilisation complémentaire des SGM pour établir des liens de cause à effet entre la présence de diverses altérations génétiques et la survenue de tumeurs cérébrales ou déterminer leur implication dans le maintien et/ou la progression des tumeurs nous donne un aperçu d'autres aspects importants de la biologie des tumeurs cérébrales [30]. (Souligné par nous.)

Fomchenko et Holland affirment ici clairement que ce qui survient chez l'animal surviendra également chez l'être humain, c'est-à-dire que les animaux sont prédictifs. Akkina dit la même chose :

L'un des principaux avantages de ce système *in vivo* [des souris SCID génétiquement modifiées] est que toutes les données obtenues avec les souris SCID-hu sont directement applicables à une situation humaine [32].

Cette utilisation de la prédiction ne se limite pas à la littérature scientifique. Elle se répand encore plus lorsque les scientifiques s'adressent à un public non scientifique.

On pourrait trouver sans effort de nombreux exemples dans la veine de ceux qui précèdent. L'omniprésence de commentaires comme les exemples ci-dessus nous laisse penser que de nombreux membres de la communauté scientifique

emploient le terme *prédire* pour signifier que ce qui se produit au niveau des modèles animaux se transpose directement chez les êtres humains. Mais cela est-il une interprétation factuelle de la réalité ?

La notion de prédiction dans les systèmes complexes biologiques

Qu'est-ce qui constitue une prédiction dans les systèmes biologiques complexes ? Nombreux sont ceux qui justifient l'utilisation des animaux comme modèles prédictifs en déclarant que les animaux sont prédictifs mais pas forcément de façon fiable. Cette formule ressemble fort à un oxymore. *Prédictif de façon fiable* est une tautologie ; dans le domaine scientifique, une méthode ne peut être dite prédictive si elle ne l'est pas de façon fiable. Nous reconnaissons que la biologie n'est pas la physique, et qu'il convient peut-être d'être indulgent lors des discussions relatives à la notion de prédiction dans les systèmes biologiques complexes. Que penser dans ce cas de cette notion dans le contexte de la toxicologie, de la pathophysiologie et de la pharmacologie ? Le tableau 2 × 2, qui permet de calculer la sensibilité, la spécificité, la valeur prédictive positive et la valeur prédictive négative est l'outil permettant d'évaluer le caractère prédictif dans ces disciplines (voir le tableau 2).

Tableau 2 : Statistiques utilisées pour analyser les prédictions

Test	Règle d'or	
	+	-
+	VP	FP
-	FN	VN

VP = vrai positif	Sensibilité = $VP / (VP + FN)$
FP = faux positif	Spécificité = $VN / (FP + VN)$
FN = faux négatif	Valeur prédictive positive = $VP / (VP + FP)$
VN = vrai négatif	Valeur prédictive négative = $VN / (FN + VN)$

En biologie, le meilleur moyen d'évaluer des concepts consiste dans bien des cas à utiliser des méthodes statistiques simples, qui font appel aux probabilités. En médecine par exemple, nous pouvons pratiquer une analyse de sang pour savoir si une personne est atteinte de troubles hépatiques. Pour évaluer si cet essai détermine réellement correctement la santé du foie, nous calculons la sensibilité et la spécificité de l'essai, ainsi que les valeurs prédictives d'un test positif (Vp+) et d'un test négatif (Vp-). La sensibilité d'un essai est la probabilité, sur une échelle de 0,0 à 1,0, d'obtenir un résultat positif chez les personnes devant donner lieu à un résultat positif – c'est-à-dire, les personnes réellement atteintes de troubles hépatiques. La spécificité est la probabilité d'obtenir un résultat négatif chez les personnes devant donner lieu à un résultat négatif – c'est-à-dire les personnes ne souffrant pas de maladie hépatique. La valeur prédictive positive d'un essai est la proportion de personnes réellement atteintes parmi toutes les personnes ayant donné lieu à un résultat positif. La valeur prédictive négative est la proportion de personnes réellement non atteintes parmi toutes les personnes ayant donné lieu à un résultat négatif. Tout ceci est donc assez simple. Très peu d'essais ont une sensibilité ou une spécificité égale à 1,0, ou bien une Vp+ et une Vp- de 1,0, mais pour qu'un essai soit considéré comme utile au regard des normes exigeantes auxquelles est soumise la pratique médicale (c'est-à-dire, dans notre exemple, pour qu'il nous indique si le patient est réellement atteint de troubles hépatiques), la Vp+ et la Vp- doivent se trouver au moins dans la fourchette de 0,95 à 1,0.

Par définition, lorsque nous envisageons que des animaux puissent prédire la réaction chez l'homme lors des essais de médicaments et des recherches sur les maladies, nous répondons aux risques de prédictions erronées et à la question du niveau de risque que la société est prête à tolérer. Le troglitazone (Rezulin™) est un bon exemple de la marge d'erreur tolérée à l'heure actuelle par la société en matière médicale. Le troglitazone a été administré à plus d'1 million de personnes, dont moins de 1 % ont souffert d'insuffisance hépatique, et pourtant ce médicament a été retiré en raison de cet effet secondaire [33]. (Il est intéressant de noter au passage que les essais sur les animaux n'ont pas réussi à reproduire l'insuffisance hépatique due au troglitazone [34].) Le rofecoxib (Vioxx™) est un autre exemple montrant combien le pourcentage de morbidité ou de mortalité toléré dans le domaine médical est faible lors de l'introduction d'un nouveau médicament. Les chiffres varient et sont controversés, mais il semble aujourd'hui établi que moins de 1 % des personnes ayant pris du rofecoxib ont subi une crise cardiaque ou une attaque cérébrale liée au médicament, pourtant celui-ci a également été retiré [35]. Ceci signifie que même si un essai doté d'une Vp+ de 0,99 avait assuré à l'industrie pharmaceutique que le rofecoxib et le troglitazone étaient sans danger, l'essai n'aurait pas été suffisamment précis au regard des critères de la société. Il s'agit-là d'un point important. La médecine ne tolère pas les risques (la probabilité de se tromper) acceptables dans un contexte d'expérience en laboratoire. Lors de recherches fondamentales, on peut donner le feu vert à une étude sur le simple fait qu'il est plus probable d'obtenir un résultat que de ne pas l'obtenir ; cela reste acceptable dans un contexte de recherche fondamentale. Mais les erreurs dans le domaine médical ont des conséquences :

des gens meurent. Les standards actuels de la société dans le domaine médical exigent des valeurs de sensibilité, spécificité, prédiction positive et prédiction négative très élevées pour les essais. Nous verrons un peu plus loin l'application de tout cela aux modèles animaux.

Les standards relatifs à la prédiction que nous venons de décrire ne doivent pas être confondus avec ceux en vigueur dans d'autres secteurs de la société, par exemple les jeux de hasard. Si nous mettions au point une méthode permettant d'obtenir des résultats corrects 51 % du temps, nous l'appliquerions sans hésiter à la table de blackjack ou à la roulette, et nous ferions sauter la banque... Parfois, trouver la bonne réponse 51 % du temps est fabuleux !

À la lumière de ce qui précède, il est courant de s'appuyer sur plusieurs tests lors de l'évaluation de l'état d'un patient ou de l'effet d'un médicament. Si quelqu'un suggère qu'un animal, par exemple une souris, est capable de prédire la réaction humaine à une molécule sur le plan de la carcinogenèse, ce quelqu'un doit pouvoir fournir des données cohérentes avec celles requises par le tableau 2. Peut-être qu'un animal seul n'est pas capable de prédire la réaction chez l'homme, mais lorsque le même résultat est obtenu chez deux espèces, par exemple singe et souris, alors les résultats ont peut-être une valeur prédictive. Ou bien, peut-être que combinée à d'autres types de données, les données animales peuvent se traduire par une valeur prédictive élevée. Mais là encore, si cela s'avérait le cas, la personne à l'origine de l'affirmation devrait être en mesure de présenter des données capables de résister à l'évaluation par la règle d'or définie dans le tableau 2. Or, à notre connaissance, il n'existe aucune donnée de cet ordre.

Prédiction des réactions chez l'homme

Nous allons à présent nous intéresser aux données réellement à notre disposition. Les données issues d'essais de six médicaments sur des animaux ont été comparées aux données sur l'être humain [36]. Les essais sur les animaux ont donné lieu à une sensibilité de 0,52 et une valeur de prédiction positive de 0,31. Cette valeur de sensibilité correspond à celle d'un jeu de pile ou face ; la $Vp+$ est encore plus faible que cela. On ne qualifierait pas de tels résultats de prédictifs, au sens scientifique du terme. Des valeurs de cet ordre correspondent plutôt à des conjectures. Ce type de données conduit les utilisateurs des modèles animaux à employer parfois l'expression *taux de concordance* ou *taux de concordance positive vrai* lors de l'évaluation d'essais sur les animaux. Ces expressions ne relèvent pas de la terminologie normale de la prédiction et sont habituellement employées dans le sens de corrélation, ce qui n'a rien à voir avec la prédiction comme nous le verrons plus loin.

Deux études des années 1990 ont révélé ce qui suit : (1) dans 4 cas de toxicité sur 24 seulement, cette toxicité a été révélée d'abord par les données chez l'animal [36] ; et (2) dans 6 cas sur 114 seulement, les toxicités cliniques sont corrélées par des données chez l'animal [37]. La sensibilité, la spécificité, la $Vp+$ et la $Vp-$ des modèles animaux de ces études sont de toute évidence loin d'être optimales.

Une étude de 1994 portant sur 64 médicaments disponibles sur le marché, réalisée par l'association japonaise des fabricants de produits pharmaceutiques, rapporte que 39 sur 91 (43 %) des toxicités cliniques n'avaient pas été prévues par les études sur les animaux [38]. (Cette étude, comme de nombreuses autres, a considéré comme une prédiction positive *tout* résultat animal corrélé à la réaction observée chez l'homme. Cette pratique est hypocrite car elle revient à sélectionner les données.) Sans connaître les données brutes il est impossible de calculer des valeurs $Vp+$ et $Vp-$ véritables, mais même utilisé tel quel, un ratio de 43 % d'erreurs pour 57 % de résultats corrects n'est pas prédictif.

Les figures 1 et 2 illustrent graphiquement notre affirmation selon laquelle les animaux ne sont pas prédictifs. Ces deux figures comparent les données de biodisponibilité chez trois espèces animales et chez l'homme. (On définit habituellement la biodisponibilité comme la fraction d'un médicament qui atteint la circulation systémique et reflète un certain nombre de variables différentes. Quelle que soit la corrélation ou l'absence de corrélation des variables, la biodisponibilité du médicament constitue le déterminant final de la quantité de médicament atteignant le récepteur ou le site actif.) La figure 1 a été compilée par Harris à partir d'un article de Grass et Sinko publié dans *Advanced Drug Delivery Reviews*. Comme peut le constater le lecteur, la biodisponibilité de différents médicaments est mesurée chez l'homme et chez trois espèces animales (parmi les primates, les rongeurs et les chiens) et les résultats représentés sous forme graphique. Certains des médicaments ayant présenté une biodisponibilité élevée chez le chien étaient à contrario faiblement biodisponibles chez l'homme, et inversement. Ce type de résultat est observé quel que soit le médicament et quelle que soit l'espèce. Certains résultats étaient corrélés entre espèces, mais dans l'ensemble il n'existe aucune corrélation entre l'effet d'un médicament chez l'homme et l'effet sur l'une quelconque des espèces animales, seules ou combinées. La figure 2 a été compilée par Harris à partir d'une contribution de Mandagere et Jones à l'ouvrage intitulé *Drug Bioavailability: Estimation of Solubility, Permeability, Absorption and Bioavailability (Methods and Principles in Medicinal Chemistry)* ; ceux-ci ont effectué les mêmes mesures et tiré les mêmes conclusions que Grass et Sinko.

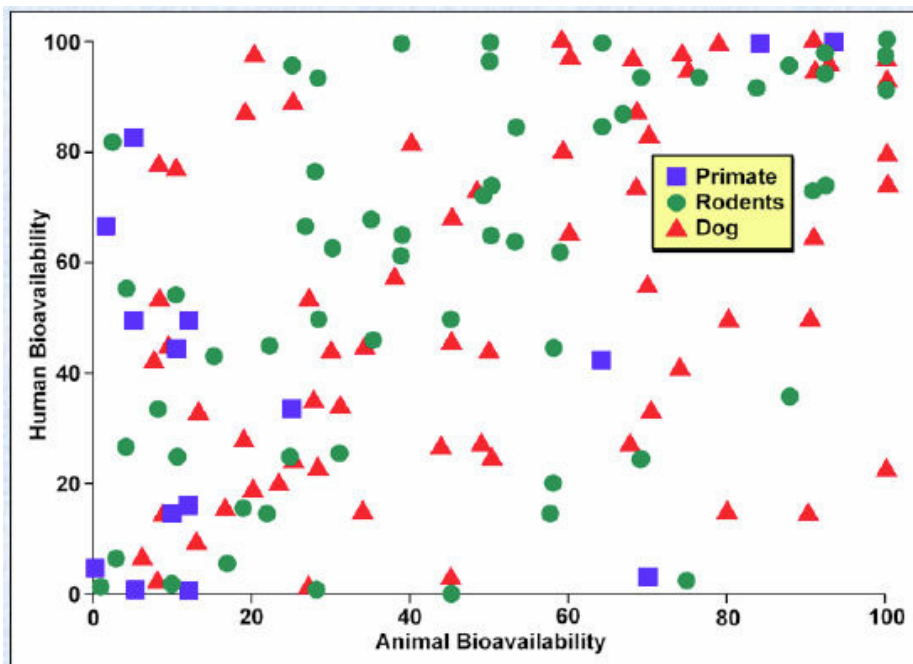


Figure 1

La biodisponibilité chez l'homme et chez l'animal (1). Graphique gracieusement fourni par James Harris, PhD, qui l'a présenté lors de la conférence du Center for Business Intelligence intitulée *6^e Forum sur les outils ADME/Tox prédictifs* et organisée à Washington DC du 27 au 29 septembre 2006. Ce graphique reprend et adapte les données publiées dans Grass GM, Sinko PJ., « Physiologically-based pharmacokinetic simulation modelling », *Adv Drug Deliv Rev.*, 31 mars 2002 ; 54(3) : 433–5.

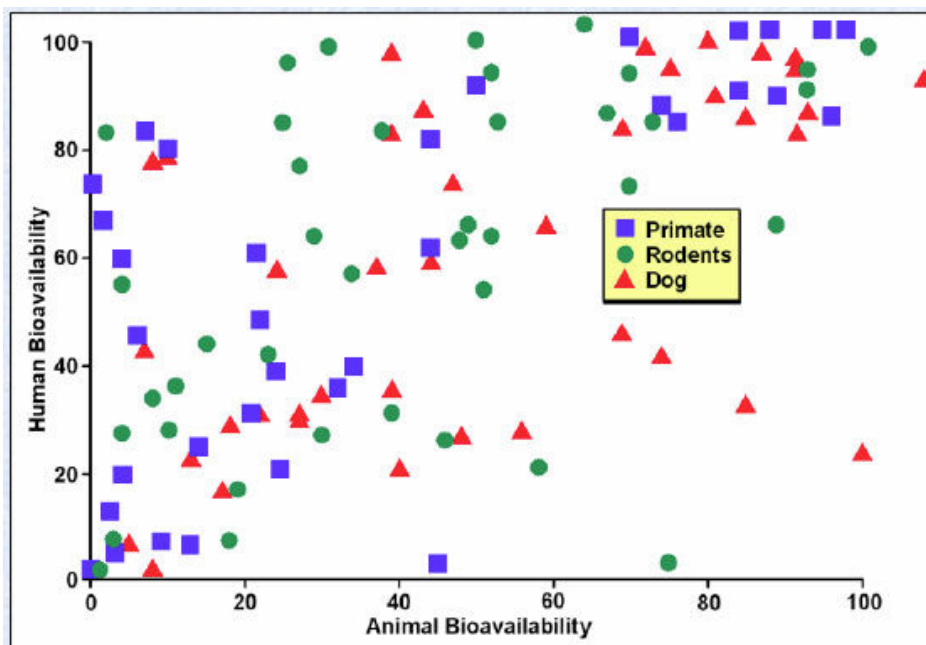


Figure 2

La biodisponibilité chez l'homme et chez l'animal (2). Graphique gracieusement fourni par James Harris, PhD, qui l'a présenté lors de la conférence du Center for Business Intelligence intitulée *6^e Forum sur les outils ADME/Tox prédictifs* et organisé à Washington DC du 27 au 29 septembre 2006. Ce graphique reprend et adapte les données publiées dans (Eds) Han van de Waterbeemd, Hans Lennernäs, Per Artursson et Raimund Mannhold. *Drug Bioavailability: Estimation of Solubility, Permeability, Absorption and Bioavailability (Methods and Principles in Medicinal Chemistry)*, Wiley-VCS 2003, 444–60.

Comme on peut le voir, il n'existe qu'une faible corrélation entre les données animales et les données humaines. Dans certains cas, la biodisponibilité est élevée chez l'homme comme chez le chien tandis que dans d'autres, les données obtenues pour l'homme et le chien varient considérablement. Les résultats chez les deux espèces sont de type dispersés, c'est-à-dire semblables à ce que l'on verrait sur une cible après avoir tiré un coup de fusil à plombs : les impacts sont multiples et dépourvus de précision et de fidélité. Ce type de résultat dispersé est également appelé schéma de dispersion ou diagramme de corrélation, car il correspond à ce que l'on pourrait attendre d'associations aléatoires. Ce qui précède met en lumière la raison pour laquelle il est problématique d'éliminer des médicaments en développement sur la foi de résultats d'essais sur les animaux. Sankar déclarait en 2005, dans *The Scientist* :

Typiquement, un composé entrant en phase I d'essais cliniques a déjà fait l'objet d'essais précliniques rigoureux pendant une dizaine d'années, et pourtant à ce stade il n'a que 8 % de chances d'être mis sur le marché. Ce faible taux est dû en partie aux toxicités qui ne se déclarent qu'à la fin des essais cliniques – ou pire, après l'autorisation du médicament. Une partie du problème est dû au fait que la toxicité est évaluée dans les phases ultimes du développement du médicament, après le criblage d'un grand nombre de composés candidats pour sélectionner celui offrant la meilleure activité et solubilité, et la fabrication du composé retenu en quantités suffisantes pour réaliser les études sur les animaux. Howard Jacob note que les rats et les êtres humains sont identiques à 90 % au niveau génétique ; or, la majorité des médicaments s'étant avérés sans danger chez l'animal échoue lors des essais cliniques. « Le pouvoir prédictif n'est que de 10 %, puisque 90 % des composés testés échouent lors des essais chez l'homme », lors des essais toxicologiques classiques effectués sur le rat. À l'inverse, certains composés prometteurs peuvent être éliminés en raison de leur toxicité chez le rat ou le chien, alors qu'ils pourraient en réalité présenter un profil de risque acceptable chez l'être humain [39]. (Souligné par nous.)

Là encore, cet argument suffirait à régler une fois pour toute la question de la prédiction. Mais continuons plus avant.

La sensibilité n'est pas la même chose que la prédiction. Bien qu'il soit établi que tous les carcinogènes humains connus et étudiés de manière adéquate se sont avérés cancérogènes chez au moins une espèce animale [40-42], un aphorisme irrévérencieux connu en biologie sous le nom de loi de Morton et selon lequel « des rats soumis à des expériences développeront un cancer » est également vrai. La loi de Morton est similaire à la loi de Karnofsky en tératologie, laquelle indique que tout composé peut être tératogène lorsqu'il est administré à la bonne espèce, à la bonne dose et au bon moment de la gestation. Ce que veulent dire ces lois est qu'il est très facile d'obtenir des résultats de carcinogénicité et tératogénicité positifs ; c'est-à-dire une sensibilité élevée. Toutefois, ces résultats ne veulent rien dire si l'on ne connaît pas également la spécificité, la valeur de prédiction positive et la valeur de prédiction négative.

Carcinogénèse

Quelle est l'efficacité des modèles animaux en matière de prédiction de la carcinogénèse ? Les composés potentiellement carcinogènes sont recensés dans la base de données chimique IRIS (Integrated Risk Information System) gérée par l'EPA, l'agence américaine de protection de l'environnement. Selon Knight et al. [43], au 1^{er} janvier 2004 IRIS était incapable de déterminer la carcinogénicité de 93 composés sur 160 testés uniquement sur des animaux. L'OMS (Organisation mondiale pour la santé) classe également les produits chimiques en fonction de leur carcinogénicité établie par l'IARC (le Centre international de recherches sur le cancer).

Knight et al. écrivaient en 2006 :

Sur les 128 produits chimiques également évalués par le Centre international de recherches sur le cancer de l'Organisation mondiale de la santé, seuls les 17 possédant des données (même limitées) concernant l'homme ($p = 0,5896$) avaient un classement de carcinogénicité compatible avec celui de l'EPA. Pour les 111 composés reposant principalement sur des données animales, l'EPA était beaucoup plus susceptible que l'IARC de noter la carcinogénicité comme présentant un risque plus élevé pour l'homme ($p < 0,0001$) [43].

Cet écart est embarrassant. Knight et al. ont examiné une étude de Tomatis et Wilbourn [44] datant de 1993. Tomatis et Wilbourn ont étudié les 780 agents chimiques ou circonstances d'exposition recensés dans les volumes 1 à 55 de la série de monographies de l'IARC [45]. Ils ont trouvé que « 502 (64,4 %) étaient classés comme présentant des preuves nettes ou limitées de carcinogénicité chez l'animal, et 104 (13,3%) étaient classés comme carcinogènes probables ou avérés chez

l'homme (...). Environ 398 composés carcinogènes pour l'animal étaient considérés comme non carcinogènes probables ou avérés pour l'homme. »

Knight et al. poursuivent :

(...) sur la base de ces chiffres de l'IARC, la valeur prédictive positive des essais animaux pour les carcinogènes probables et avérés pour l'homme n'est que de 7 % environ (104/502), tandis que le taux de faux positifs s'élève de manière inquiétante à 79,3 % (398/502) [43].

Des classements plus récents réalisés par l'IARC indiquent très peu de changement de la prédictivité positive des essais sur les animaux des carcinogènes pour les humains. En janvier 2004, dix ans plus tard, seuls 105 agents supplémentaires avaient été ajoutés au chiffre de 1993, soit un total de 885 agents ou circonstances d'exposition recensés dans les monographies de l'IARC [46]. Sans surprise, la proportion de carcinogènes avérés ou probables pour l'homme était voisine des 13,3 % de 1993. En effet, en 2004, seuls 9,9 % de ces 885 étaient classés comme carcinogènes avérés, et seuls 7,2 % comme carcinogènes probables pour l'homme, soit 17,1 % au total.

Dans une étude publiée en 2000, Haseman [47] révélait que 250 (53,1 %) des substances chimiques listées dans la base de données de carcinogénicité NTP étaient carcinogènes dans au moins un groupe sexe-espèce. Il en concluait que le nombre réel de substances présentant un risque important de carcinogénicité pour l'homme était sans doute bien plus faible. Environ la moitié de toutes les substances testées sur des animaux et inscrits dans la base de données de potentiel de carcinogénicité de Berkeley (CPDB) étaient carcinogènes [48].

Knight et al. en concluent :

En appliquant une interprétation visant à éviter les risques, selon laquelle tout résultat positif chez le rat ou la souris mâle ou femelle est considéré comme positif, alors 9 sur 10 carcinogènes humains connus parmi les centaines de substances chimiques testées par le NTP sont positifs, tout comme le sont également un taux non plausible de 22 % de toutes les substances soumises à essai. En appliquant une approche moins sensible au risque, selon laquelle seules les substances chimiques ayant donné lieu à un résultat positif chez le rat et chez la souris sont considérées comme positives, alors seulement trois des six carcinogènes humains connus testés chez les deux espèces sont positifs. La première interprétation peut conduire au rejet non nécessaire de substances chimiques présentant une utilité potentielle pour la société, la seconde pouvant conduire à une exposition humaine de grande ampleur à des carcinogènes insoupçonnés [43].

À ce stade du débat, d'aucuns affirmeront que les modèles animaux peuvent être utiles à la science et à la recherche scientifique, et tenteront de faire l'amalgame entre les termes *prédictif* et *utile*. Ce procédé est retors pour de nombreuses raisons, la première étant que le terme *utile* est trop ambigu pour signifier quoi que ce soit. Utile pour qui ? Utile comment ? Presque tout peut être utile selon le point de vue que l'on adopte. Si une personne se fait payer pour deviner l'avenir, on peut dire que la divination est utile à cette personne. Mais que la divination soit un moyen fiable de prédire l'avenir est une toute autre question... Nous ne nions pas que les modèles animaux puissent être *utiles* dans certaines circonstances, mais cela n'a rien à voir avec leur éventuelle nature prédictive. Deuxièmement, ce procédé est un exemple typique de leurre publicitaire : vantez les modèles animaux en tant que prédictifs chez l'homme, puis justifiez leur utilisation par leur *utilité*, car ils ne sont pas prédictifs. Freeman et St Johnston illustrent ce type de procédé trompeur lorsqu'ils indiquent :

De nombreux scientifiques travaillant sur des organismes modèles, y compris nous, créent de toute pièce une connexion avec une maladie humaine afin d'obtenir un financement ou de publier un article. Cela peut se comprendre ; après tout, les parallèles sont véridiques, mais les connexions souvent indirectes. Nous en verrons plus d'exemples dans les chapitres ultérieurs [49].

Troisièmement, le verbe prédire a une signification bien précise dans le domaine des sciences, et c'est d'ailleurs la notion de prédiction qui distingue les sciences des pseudosciences. En faisant l'amalgame entre *utile* et *prédictif*, nous portons atteinte à la respectabilité des sciences en général, en les rabaisant au niveau du négoce de véhicules d'occasion. Et de nouveau, nous reconnaissons que l'étude des animaux peut conduire à de nouvelles connaissances ; nous ne contestons pas ce point.

Le cas du thalidomide

Regardons de plus près un médicament et les essais sur animaux qui *auraient pu* être réalisés, afin d'évaluer ce que ces essais auraient pu nous apprendre. Il existe de nombreux exemples de modèles animaux donnant des résultats avec une variance extrême par rapport aux humains, et même d'un modèle animal à l'autre ; le thalidomide est un exemple parmi d'autres, mais nous l'avons choisi en raison de sa place particulière dans l'histoire. Le thalidomide était un sédatif, qui a été prescrit aux femmes enceintes à la fin des années 1950 et au début des années 1960. Les enfants de certaines de ces femmes sont nés dépourvus de membres, une malformation appelée phocomélie. La tragédie du thalidomide aurait-elle pu être prédite et évitée grâce à des essais sur les animaux, comme l'ont affirmé Gad [29] et d'autres ? Examinons les preuves. Schardein, qui a étudié cette tragédie, observe ce qui suit :

Chez environ 10 souches de rats, 15 souches de souris, 11 races de lapins, 2 races de chiens, 3 souches de hamsters, 8 espèces de primates et autres espèces diverses telles que chats, tatous, cobayes, porcs et furets sur lesquelles a été testé le thalidomide, des effets tératogènes n'ont été constatés que de façon occasionnelle [50].

Nous rappelons au lecteur que ces résultats, ainsi que ceux qui suivent, provenaient d'essais réalisés après la survenue des effets du thalidomide sur l'être humain. Schardein observe également :

Les essais de tératogénicité les plus décevants sont ceux conduits sur les primates, en raison de l'utilisation potentielle de ces animaux en tant que modèle prédictif. Bien que neuf primates non humains (tous à l'exception du galago) aient présenté, après administration de thalidomide, les défauts des membres observés chez les humains, les résultats obtenus avec 83 autres agents testés sur les primates se sont avérés loin d'être parfaits. Sur les 15 tératogènes potentiels pour l'être humain recensés, testés chez des primates non humains, seuls huit se sont avérés également tératogènes chez l'une au moins des différentes espèces [51].

Manson et Wise résumant comme suit les essais réalisés sur le thalidomide :

Un résultat inattendu fut que le rat et la souris se sont montrés résistants, le lapin et le hamster ont présenté des réactions variables et certaines souches de primates se sont montrées sensibles à la toxicité développementale du thalidomide. Différentes souches de la même espèce d'animaux ont également montré une sensibilité très variable au thalidomide. Des facteurs tels que les différences d'absorption, de distribution, de biotransformation et de transfert placentaire ont été écartés en tant que causes de la variabilité de sensibilité entre espèces et entre souches [52].

L'utilisation de modèles animaux aurait-elle pu prédire les effets tératogènes du thalidomide ? Même si tous les animaux mentionnés plus haut avaient été étudiés, la réponse est non. Les différentes espèces ont présenté des réactions très variées au thalidomide. Ici encore, si vous pariez sur plusieurs chevaux à la fois vous tomberez probablement sur le gagnant – de même, si vous sélectionnez les données, vous trouverez un résultat positif. Dans le cas du thalidomide, ses effets sur les humains étaient déjà connus ; il est ainsi facile de sélectionner les données. Les modèles animaux pour le thalidomide évoqués plus haut avaient pour objet de simuler rétroactivement des effets *commus* chez l'homme. Et pourtant, peu de modèles animaux ont réussi. Si les effets sur l'homme avaient été *incommus*, quelle valeur prédictive auraient eu ces essais ? Dans le cas présent, pour poursuivre l'analogie avec les chevaux de course, nous nous serions trouvés face à un très grand nombre de chevaux, sans aucune indication sur le futur vainqueur. Il y aura bien entendu un gagnant (ce qui n'est pas forcément le cas avec les essais sur les animaux ayant pour but de reproduire ou de deviner une réaction chez l'homme), mais lequel ? Il nous est impossible de le savoir avant d'observer les résultats factuels, alors comment juger *préalablement* du cheval gagnant – ou du modèle animal à choisir ? Quelles espèces seraient les plus aptes à modéliser l'homme avant d'avoir collecté des données chez ce dernier ? Cette question est loin d'être anodine, car la proximité dans l'évolution n'augmente pas la valeur prédictive du modèle. Caldwell souligne que des différences biologiques relativement faibles entre deux sujets testés peuvent donner lieu à des résultats très différents :

Il est évident depuis un certain temps qu'il n'y a généralement aucune base liée à l'évolution derrière l'aptitude d'une espèce donnée à métaboliser un médicament donné. De fait, chez les rongeurs et les primates, des espèces proches sur le plan zoologique présentent des mécanismes de métabolisme sensiblement différents [53].

L'exemple du thalidomide montre à quel point l'hypothèse généralisée selon laquelle les animaux ont une valeur prédictive pour l'homme est fautive. De nouveau, cette hypothèse généralisée est à l'opposé de l'utilisation d'animaux en tant que dispositifs heuristiques, qui permettent de dégager des hypothèses lesquelles sont, ensuite, soumises à essai.

Même en sélectionnant a posteriori tous les animaux ayant réagi de la même manière que les humains au thalidomide, il nous aurait été impossible d'affirmer que ces animaux étaient capables de prédire les effets chez l'homme, car leur corrélation avec les réactions humaines à d'autres médicaments varie considérablement. Or pour effectuer des prédictions relatives à des essais de médicaments dans le cadre de recherches sur les maladies, il convient de disposer d'un historique de réussite. Des conjectures correctes isolées ne constituent pas des prédictions. Les primates non humains en sont un bon exemple. Ils ont plus ou moins réagi de la même manière que les humains au thalidomide (on leur accorde donc le bénéfice du doute en acceptant qu'ils correspondent aux humains dans ce cas précis). Toutefois, soumis à d'autres médicaments, ils ont *prédict* les réactions chez l'homme avec autant de succès qu'à pile ou face. Si l'on ajoute à cela le fait que la descendance atteinte de phocomélie suite à l'administration de thalidomide à la génération parentale l'a été uniquement lorsque les parents ont reçu des doses 25 à 150 fois supérieure à la dose humaine [54-56] : il semble que ni aucun animal, ou groupe d'animaux, ni que le modèle animal lui-même aurait pu être utilisé pour prédire la tératogénicité chez l'être humain. (Ironie du sort, c'est la triste affaire du thalidomide qui a permis de faire passer un grand nombre de réglementations exigeant des essais sur les animaux.)

L'histoire controversée du thalidomide ne doit pas interférer avec notre analyse, car cette histoire ne rentre pas dans le champ de notre question. La controverse en question porte sur les animaux utilisés pour les essais, sur la question de savoir si des animaux gravides ont été testés, de ce que savait le fabricant du médicament, à quel moment il l'a su, etc. Rien de cela ne nous concerne, car nous analysons les données comme si elles étaient disponibles avant la mise sur le marché du médicament. Nous accordons au modèle animal le plus grand bénéfice du doute possible, et nous arrivons à la conclusion que même si toutes les données disponibles aujourd'hui l'avaient été à l'époque, la décision d'autoriser ou non le médicament n'aurait pas été *fondée* sur les essais sur l'animal. La loi de Karnofsky s'applique ici. Tout médicament est tératogène s'il est donné au bon animal, au bon moment. Étant donné le profil du thalidomide aujourd'hui, les médecins recommanderaient aux femmes enceintes de ne pas prendre ce médicament, ce qu'ils font de toute façon pour tous les médicaments non vitaux, quels que soient les résultats des essais sur les animaux.

L'affirmation selon laquelle les effets du thalidomide ont été ou auraient pu être prévus par les animaux est un exemple de sélection des données.

La controverse du quantitatif et du qualitatif

Intéressons-nous à présent à la controverse du quantitatif et du qualitatif. Il existe une tendance, chez certains chercheurs, de voir les différences entre espèces comme de simples différences quantitatives, c'est-à-dire des différences que l'on suppose pouvoir compenser dans le contexte de la prédiction. Selon Vineis et Melnick :

Nous sommes toutefois en désaccord avec Shanks et Pyles à propos de l'utilité des essais sur les animaux pour prédire les risques pour l'homme. Si l'on se base sur l'observation darwinienne des variations, entre espèces et entre individus d'une même espèce, de toutes les fonctions biologiques, Shanks et Pyles suggèrent que les essais sur les animaux ne peuvent pas être utilisés pour identifier les risques pour la santé humaine. Nous affirmons que bien que l'activité enzymatique puisse varier d'un individu et d'une espèce à l'autre, ceci ne signifie pas que des événements critiques dans les processus des maladies survenant après l'exposition à des agents dangereux diffèrent qualitativement entre les modèles animaux et l'homme... Les différences dans la manière dont les animaux de laboratoire et les êtres humains métabolisent les agents environnementaux, ou dont ces agents interagissent avec les cibles moléculaires (ADN, enzymes, récepteurs nucléaires...) sont en grande partie de nature quantitative [27].

Cette manière de penser, très newtonienne, ne tient pas compte des effets de l'évolution biologique, ni du fait que les animaux sont des systèmes complexes.

Les toxicologues savent depuis longtemps que les différences entre espèces peuvent être *quantitatives* ou *qualitatives* [53,57]. Prenons une substance modèle telle que le phénol. Les êtres humains et les rats éliminent le phénol selon deux mécanismes : par conjugaison avec les sulfates et par conjugaison avec l'acide glucuronique. Il existe une différence quantitative entre les humains et les rats dans la mesure où le ratio entre la conjugaison avec les sulfates et la conjugaison avec l'acide glucuronique est différent pour les deux espèces. Mais il existe également des différences d'ordre qualitatif. Par exemple, la conjugaison à l'acide glucuronique est impossible chez le chat, qui élimine donc le phénol exclusivement par conjugaison avec les sulfates. C'est l'inverse chez le cochon, chez qui la conjugaison aux sulfates n'existe pas et qui élimine le phénol avec l'acide glucuronique. (Il est intéressant de noter qu'il existe au moins sept voies métaboliques exclusives aux primates – par exemple l'aromatation de l'acide quinique [57].)

Ce que l'on peut retenir de cet exemple est que même lorsqu'une même fonction est réalisée par deux espèces (ici, l'élimination du phénol), cette fonction n'est pas automatiquement réalisée par les mêmes mécanismes causaux sous-jacents. En toxicologie et en pharmacologie, ces différences sur le plan des mécanismes peuvent avoir une grande importance lors de l'évaluation de la sécurité et de l'utilité pharmacologique.

D'autres voix que la nôtre

Nous ne sommes pas les seuls à nous intéresser à la question du caractère prédictif des modèles animaux. Au sein même de la communauté scientifique, les avis diffèrent quant à l'utilité prédictive des modèles animaux. Nous allons à présent examiner ce que disent certains de ces scientifiques au sujet de la capacité des modèles animaux à prédire les réactions chez l'homme. Si les citations qui suivent (et celles qui précèdent, de Leavitt et Sankar) ne constituent pas des preuves au sens mathématique du terme, elles offrent malgré tout un aperçu de ce que pensent des personnes compétentes en la matière et doivent donc être prises avec le sérieux qui convient. Ces avis devraient donner à réfléchir à tous ceux qui pensent que la question de la prédiction ne donne pas lieu à une controverse sensée.

R.J. Wall et M. Shani observent :

La vaste majorité des animaux utilisés comme modèles le sont lors d'essais précliniques biomédicales. La valeur prédictive de ces études sur l'animal est soigneusement surveillée, ce qui en fait un réservoir de données idéal pour évaluer l'efficacité des modèles animaux. En moyenne, les résultats extrapolés à partir d'études portant sur des dizaines de millions d'animaux ne réussissent pas à prédire avec exactitude les effets chez l'homme (...) Nous en concluons qu'il est probablement plus sûr d'utiliser les modèles animaux pour émettre des hypothèses, que dans le but de réaliser des extrapolations [58].

Curry note :

L'échec, sur le plan clinique, d'au moins quatorze agents neuroprotecteurs potentiels censés favoriser la récupération après une attaque cérébrale, dont des essais sur les animaux avaient prédit qu'ils fonctionneraient chez l'homme, est examiné selon les principes de l'extrapolation, à l'homme, des données obtenues sur l'animal [59].

Ces citations prouvent deux choses : premièrement, que certains membres au moins de la communauté de l'expérimentation animale connaissent la signification du terme prédire ; et deuxièmement, qu'ils savent aussi que les modèles animaux ne sont pas prédictifs. Leurs analyses et leurs conclusions, qui révèlent l'échec des modèles animaux, ne sont ni nouvelles, ni surprenantes. L'histoire nous apprend que les écarts entre les études chez l'homme et l'animal – et même entre études chez l'animal – remontent à plusieurs siècles. Percival Pott a démontré en 1776 le caractère carcinogène du goudron de houille pour l'homme. Yamagiwa et Ichikawa ont montré que cette substance était carcinogène chez certains animaux en 1915. Mais même à cette époque, les lapins n'ont pas réagi comme les souris [60]. En 1980, il existait environ 1600 substances chimiques connues dont la carcinogénicité était avérée chez la souris et d'autres rongeurs, mais la carcinogénicité chez l'homme était prouvée pour seulement quinze de ces composés environ [61]. Le Council on Scientific Affairs déclarait en 1981, lors d'une publication dans le *Journal of the American Medical Association* :

Les consultants du Council reconnaissent que l'identification d'une carcinogénicité lors d'essais sur les animaux ne constitue pas en tant que telle une prédiction d'un risque ou d'un résultat lors d'essais sur l'homme (...) Le Council s'inquiète des centaines de millions de dollars dépensés chaque année (tant dans le secteur public que privé) pour déterminer la carcinogénicité des substances chimiques. Cette inquiétude est particulièrement justifiée au regard de la valeur scientifique contestable de ces essais en matière de prédiction des résultats chez l'homme [62]. (Souligné par nous.)

David Salsburg de la société Pfizer écrivait en 1983 qu'un rapport du National Cancer Institute ayant examiné 170 substances chimiques concluait que les études d'alimentation à long terme (vie entière) de rongeurs manquaient de sensibilité et de spécificité. Selon ses termes :

Si nous limitons notre attention aux études d'alimentation à long terme sur le rat ou la souris, seuls 7 des 19 agents carcinogènes pour l'homme par des voies autre que l'inhalation (36,8 %) ont provoqué des cancers. Si nous ajoutons à ces études les études par inhalation et examinons les 26 produits testés, seuls 12 (46,2 %) ont provoqué des cancers chez le rat ou la souris après une exposition chronique par l'alimentation ou par inhalation. La probabilité des études de nourrissage à long terme chez le rat et la souris d'identifier des substances

carcinogènes avérées pour l'homme est donc inférieure à 50 %. Si l'on en croit la théorie des probabilités, on aurait obtenu de meilleurs résultats en jouant à pile ou face [63]. (Souligné par nous.)

Devons-nous pour autant écarter tous les médicaments qui provoquent des cancers chez les animaux ? Le paracétamol, le chloramphénicol et le métronidazole sont connus pour être cancérigènes pour certaines espèces animales [64,65]. Le phénobarbital et l'isoniazide sont cancérigènes pour les rongeurs [60,66,67]. Cela signifie-t-il qu'ils n'auraient pas dû être autorisés ? Le diphenylhydantoïne (ou phénytoïne) est cancérigène pour l'homme, mais pas pour le rat et la souris [68-70]. L'exposition en milieu professionnel à la 2-naphtylamine semble provoquer des cancers de la vessie chez les humains. Les chiens et les singes exposés par voie orale à cette même substance développent également des cancers de la vessie, tandis qu'elle provoque des hépatomes chez la souris. Par contre, elle ne semble pas avoir d'effets cancérigènes chez le rat et le lapin. Ces différences de nature qualitative sont dues à des différences au niveau du métabolisme des amines aromatiques [71]. Il apparaît également que l'induction d'un cancer dépend d'un moins grand nombre d'événements génétiques, épigénétiques ou d'expression génique chez le rongeur que chez l'homme [72-74]. (Anisimov et al. [72] ont procédé à une revue intéressante des différences qui existent entre espèces en matière de carcinogénèse, en expliquant à quoi sont dues ces différences.)

Il existe aussi des différences au sein même des espèces. Le clofibrate, la nafénopine, le phénobarbital et la réserpine provoquent des cancers chez le rat âgé, mais pas chez les jeunes [68,75].

Ces médicaments, qui induisent des cancers chez certains animaux, auraient-ils dû être interdits ? Si l'hypothèse nulle est qu'il n'existe aucune association entre les carcinogènes pour l'animal et les carcinogènes pour l'homme qui soit suffisamment forte pour conférer une valeur prédictive au modèle animale, alors les preuves étayant cette hypothèse sont nombreuses, tandis que celles qui la réfutent sont rares, voire inexistantes.

Ce que nous souhaitons montrer ici est qu'il existe des scientifiques (trop nombreux pour pouvoir être mentionnés dans ces pages) qui remettent en question la valeur prédictive et/ou clinique des essais sur les animaux, et que l'histoire leur donne raison. Comme nous l'avons dit plus haut, l'avis d'un scientifique ne prouve rien en lui-même. De plus, certains avis présentés ici pourraient être écartés pour leur caractère anecdotique, mais ce serait une erreur. D'une part, les études auxquelles il est fait référence dans la partie précédente sont justement cela : des études scientifiques, et non des anecdotes. D'autre part, les exemples présentés sont repris en référence, tandis que les anecdotes ne le sont pas (sauf lorsqu'il s'agit de cas signalés, et n'oublions pas que l'effondrement du thalidomide a commencé par une série de signalements). Mais là encore, il convient de se demander à qui incombe de faire la preuve. Nous croyons la seconde loi de la thermodynamique car elle n'a jamais été contredite par un seul exemple. Un seul contre-exemple réfuterait la loi. Si les partisans du modèle animal affirment que ce modèle est prédictif, qu'ils expliquent alors les exemples et les études qui indiquent le contraire. Les exemples ponctuels tels que les signalements de cas comptent lorsqu'il s'agit d'infirmer une affirmation, surtout lorsque ces cas sont ensuite étayés par des études, et qu'ils ne peuvent être utilisés comme preuve par ceux qui soutiennent l'hypothèse généralisée. C'est ainsi que fonctionne la science. Ce n'est pas nous qui en avons défini les règles. En résumé, il existe de nombreuses preuves permettant de remettre en question, voire même de réfuter totalement, l'hypothèse généralisée selon laquelle les modèles animaux sont prédictifs pour les êtres humains.

Pour poursuivre dans le même ordre d'idée, nous devons chercher quelles conditions devraient être satisfaites pour que les animaux puissent servir d'outils de prédiction des phénomènes biomédicaux chez l'homme. Une telle question est d'ordre théorique et touche au domaine de la philosophie des sciences.

Analyse théorique

Les exigences à satisfaire pour réaliser de véritables prédictions de causalité (et non de simples corrélations) sur les membres d'une espèce, en se basant sur les résultats d'essais réalisés sur les membres d'une autre espèce, sont très difficiles à satisfaire (et au mieux, elles ne peuvent qu'être approchées dans quelques cas). Les modèles, ou toute modalité prétendant à la prédictibilité, ont des propriétés de causalité identiques. Comme l'expliquent les chercheurs Carroll et Overmier dans leur ouvrage récent *Animal Research and Human Health* [76], ainsi que LaFollette et Shanks dans *Brute Science* [3], les animaux utilisés en recherche biomédicale sont fréquemment utilisés en tant que modèles à analogie de causalité (MAC). Si le cœur sert à pomper le sang chez le chimpanzé, on déduit par analogie qu'il fait de même chez l'homme. Si le fenfluramine-phentermine est sans danger pour le cœur des animaux, on déduit par analogie qu'il le sera également pour le cœur des êtres humains [77]. Carroll et Overmier déclarent :

Lorsque l'expérimentateur détermine des problèmes pour l'animal et étudie une chaîne de causalité qui, par analogie, peut être vue comme parallèle aux problèmes posés à l'homme, alors l'expérimentateur utilise un modèle animal [76].

Ces exemples illustrent l'emploi des animaux en tant que MAC ou modèles prédictifs au sens de la définition traditionnelle du terme prédiction, qui est celle utilisée dans le présent article. Nous reviendrons plus en détail sur les MAC dans la partie relative à la théorie.

Les modèles animaux sous cette acception impliquent un raisonnement par analogie de causalité. Tout d'abord, qu'est-ce qu'un modèle à analogie de causalité (MAC) et comment fait-il appel au raisonnement par analogie de causalité ? La première condition à respecter pour qu'un objet puisse être considérée comme un MAC est celle-ci : « X (le modèle) est semblable à Y (l'objet modélisé) sur les plans $\{a...e\}$. » Si « X présente une propriété additionnelle f , alors bien que f n'ait pas été observée directement chez Y, Y possède vraisemblablement également la propriété f [3]. » Cette dernière affirmation doit être mise à l'épreuve. Dans le cas présent, cela signifie qu'il faut disposer de données chez l'homme.

Cette première condition n'est pas suffisante. Par exemple, le chimpanzé et l'être humain ont (a) tous deux un système immunitaire, (b) 99 % de leur ADN en commun, (c) la capacité de contracter des virus, etc. Le VIH se réplique très lentement chez le chimpanzé ; on s'attend donc à ce qu'il fasse de même chez l'homme [3]. Par conséquent, si le VIH se réplique lentement chez le chimpanzé, les partisans de l'expérimentation animale déduisent *par analogie* qu'il en sera de même chez l'homme. Or cela n'est pas le cas.

Les MAC doivent répondre à deux autres conditions : (1) les propriétés communes (a, \dots, e) doivent être des propriétés de causalité, lesquelles (2) sont liées par un lien de cause à effet à la propriété (f) que nous souhaitons modéliser – en d'autres termes, (f) doit être la ou les cause(s) ou les effet(s) des propriétés (a, \dots, e) du modèle. L'utilisation des animaux en tant que modèles à analogie de causalité, qui permet de passer des résultats chez le modèle au système modélisé, est appelé *raisonnement par analogie de causalité* [3].

Il n'est toutefois pas suffisant de souligner les similitudes pour justifier une extrapolation d'une espèce à une autre dans le cadre du raisonnement par analogie de causalité. Dans les systèmes interactifs et complexes tels que les organismes, nous devons savoir s'il existe des différences causales pertinentes, c'est-à-dire des dissemblances de causalité (par rapport aux mécanismes et aux voies) qui compromettent la valeur du raisonnement par analogie. En d'autres termes, pour qu'un MAC ait une valeur prédictive, il ne doit exister aucune dissemblance de causalité entre le modèle et l'objet modélisé. Par exemple, il ne doit exister aucune propriété $\{g, h, i\}$ unique au modèle ou à l'objet modélisé qui interfère de manière causale avec les propriétés communes $\{a...e\}$, car de telles propriétés compromettraient vraisemblablement l'utilité prédictive du modèle.

L'idée en cause ici n'est pas nouvelle. Elle concerne le *déterminisme causal*, un concept qui a joué un rôle fondamental dans le développement de la science moderne. Le déterminisme causal repose sur deux principes fondamentaux : (1) *le principe de causalité*, qui énonce que tout événement a une cause ; et (2) *le principe d'uniformité* selon lequel, pour des systèmes identiques sur le plan qualitatif, toutes choses égales par ailleurs, une même cause est toujours suivie du même effet.

Ces idées ont joué un rôle dans notre discussion précédente de la mécanique newtonienne, au début du présent article. D'une certaine manière, la question entière de la prédiction se résume au principe d'uniformité. Les animaux utilisés pour formuler des prédictions sur les êtres humains sont-ils identiques à ces derniers sur le plan qualitatif, une fois prises en compte les différences de poids corporel ou de surface ? Aucune personne sensée, comprenant la biologie de l'évolution et sachant, par exemple, que les rongeurs et les humains ont suivi des trajectoires évolutives très différentes depuis que les lignées conduisant à l'homme et aux rongeurs modernes ont divergé il y a plus de 70 millions d'années, ne s'attend à une identité qualitative entre les deux espèces. Mais l'identité qualitative est peut-être un idéal qui ne peut être qu'approché. Les humains et les animaux employés pour les modéliser sont-ils suffisamment similaires pour pouvoir formuler des prédictions approchées ? Les nombreuses études citées plus haut à titre de référence répondent non. Pour quelle raison ?

Les vertébrés sont des systèmes complexes évolués. De tels systèmes peuvent manifester des réactions différentes à un même stimulus en raison : (1) de différences entre gènes ou allèles ; (2) de différences au niveau des mutations sur un même gène (une espèce possédant un orthologue d'un gène présent chez une autre espèce) ; (3) de différences relatives aux protéines et à leur activité ; (4) de différences liées à la régulation des gènes ; (5) de différences liées à l'expression

des gènes ; (6) de différences au niveau des interactions protéine-protéine ; (7) de différences au niveau des réseaux génétiques ; (8) de différences liées à l'organisation de l'organisme (les êtres humains et les rats peuvent être des systèmes intacts, mais de façon différente) ; (9) de différences au niveau des expositions environnementales ; et enfin, (10) de différences sur le plan des histoires évolutives. Ces raisons sont parmi les raisons importantes pour lesquelles les membres d'une espèce réagissent souvent différemment aux médicaments et aux toxines, et souffrent de maladies différentes. Ces dix faits suffiraient à eux seuls pour conclure que les modèles animaux ne peuvent pas être prédictifs pour l'être humain ; que l'extrapolation entre espèces est impossible en ce qui concerne la réaction à un médicament et les maladies, en particulier à la lumière des critères exigés par la société actuelle. (Critères qui ne sont pas trop élevés. Si vous pensez qu'ils le sont, posez-vous la question suivante : si vous aviez pris du rofecoxib et en aviez subi les effets nocifs, une Vp+ de 0,99 vous aurait-elle semblé acceptable ?)

Le domaine biomédical ne peut se payer le luxe des mathématiciens en déclarant, en préambule à la modélisation des êtres humains et des rongeurs : « soient l'homme et le rongeur, tout deux représentés par des sphères. » Si seulement c'était aussi simple ! Au lieu de cela, nous sommes face à une riche matière théorique qui remet en question la nature prédictive des modèles animaux. Bien entendu, un tel raisonnement théorique pourrait être rejeté au motif qu'il n'est justement que théorique. La véritable question est celle des preuves. Nous avons examiné les preuves à charge et les avons trouvées irréfutables, mais nous allons à présent examiner les preuves citées comme démontrant la nature prédictive des animaux. Pour cela, nous allons examiner la célèbre étude Olson, dont beaucoup pensent qu'elle a réglé clairement la question en établissant le caractère prédictif du modèle animal pour l'homme.

La célèbre étude Olson

L'étude Olson [78] a pour objet (et elle a certainement été citée à cet égard) d'apporter des preuves de l'utilité prédictive étendue des modèles animaux lors des études de toxicité chez l'homme. En réponse à un article de Shanks et al. [79], Conn et Parker citaient l'étude Olson et indiquaient :

Les auteurs ont simplement fait abstraction de l'étude classique (Olson Harry et al., 2000. "Concordance of the Toxicity of Pharmaceuticals in Humans and in Animals." *Regul Toxicol Pharmacol* 32, 56–67), laquelle compile les résultats de 12 sociétés pharmaceutiques internationales sur le caractère prédictif des essais animaux lors des études de toxicité chez l'homme. Bien qu'imparfaite, cette étude arrive à la conclusion générale, sur 150 composés et 221 événements de toxicité chez l'homme, que l'expérimentation animale possède une valeur prédictive significative en ce qu'elle détecte la plupart – mais pas tous – des domaines de toxicité humaine [80]. (Souligné par les auteurs)

Nous invitons le lecteur à examiner l'étude Olson dans son intégralité. Nous en reproduisons quelques extraits représentatifs importants, suivis de nos commentaires. Le lecteur excusera la longueur de la citation, mais en raison de l'importance accordée par beaucoup à cette publication, il nous a semblé qu'un examen poussé s'imposait.

Le présent rapport résume les résultats d'une étude réalisée par des multinationales pharmaceutiques, ainsi que le résultat d'un atelier de l'ILSI (international Life Sciences Institute) d'avril 1999, dont le but était de mieux comprendre la concordance entre la toxicité de composés pharmaceutiques observée chez les humains et celle observée chez les animaux de laboratoire. Ont participé à cet atelier des universitaires, des représentants de l'industrie pharmaceutique internationale et des scientifiques d'organismes réglementaires internationaux. L'objectif principal de ce projet était d'examiner les forces et les faiblesses des études animales en matière de prédiction de la toxicité humaine (TH). La base de données a été constituée à partir d'une étude ayant porté uniquement sur les composés ayant donné lieu à toxicité chez l'homme lors de la phase clinique du développement de nouveaux médicaments, et qui a permis de déterminer si les études de toxicité chez l'animal avaient identifié les toxicités sur organes cibles concordantes chez les êtres humains (...)

Les résultats ont montré un taux de concordance positive réelle des TH de 71 % pour des espèces de rongeurs et des espèces autres, lesdites espèces autres prises considérées séparément ayant prédit 63 % des TH et les rongeurs considérés séparément en ayant prédit 43 %. L'incidence la plus élevée de concordance générale des TH a été observée sur les plans hématologique, gastro-intestinal et cardiovasculaire, et la plus faible sur le plan de la toxicité cutanée. Dans les cas où les modèles animaux, d'une ou plusieurs espèces, ont identifié des TH concordantes, la première observation de ces TH a eu lieu, dans 94 % des cas, lors d'études d'une durée de 1 mois ou moins. Ces résultats démontrent la valeur des études de toxicologie in vivo pour prédire de nombreuses TH associées aux médicaments, et ils ont contribué à identifier les catégories de TH susceptibles de bénéficier de méthodes améliorées (...)

L'objectif principal était d'examiner à quel degré les toxicités observées lors des études précliniques sur les animaux seraient capables de prédire les toxicités humaines réelles pour un certain nombre d'organes cibles spécifiques, en utilisant une base de données constituée d'informations existantes (...)

Bien que des efforts considérables aient été faits pour rassembler des données permettant une comparaison directe de la toxicité chez l'homme et chez l'animal, il a été reconnu dès le début que les données ne pouvaient répondre complètement à la question visant à déterminer avec quel degré de réussite les études animales prédisent globalement les effets chez l'homme. Pour cela, il faudrait des informations dans les quatre quadrants de la Fig. 1, ce qui n'était pas possible en pratique à ce stade. L'ampleur de l'effort de recueil de données nécessaire pour cela a été considérée comme impossible en pratique à ce stade. La présente analyse est une première étape, au cours de laquelle les données recueillies appartenaient uniquement à la colonne de gauche de la Fig. 1: les vrais positifs et les faux négatifs. [Voir la figure 3.] Par définition, la base de données contient donc uniquement les composés testés chez l'homme (et pas ceux qui n'ont jamais atteint ce stade car considérés comme trop toxiques chez les animaux, ou qui ont été retirés pour des raisons non liées à la toxicité). Malgré cette limitation, on a estimé qu'il était utile de poursuivre, en considérant que les conclusions éventuelles qu'il serait possible de tirer répondraient à certaines questions clés et montrer du doigt certaines des forces et des faiblesses des études sur les animaux (...)

		<u>Toxicité humaine</u>	
		<u>Pos.</u>	<u>Nég.</u>
<u>Toxicité animale</u>	<u>Pos.</u>	Vrai	Faux {perte de NCE}
	<u>Nég.</u>	Faux {risques évités}	Vrai

Figure 3 : La Figure 1 de l'étude Olson

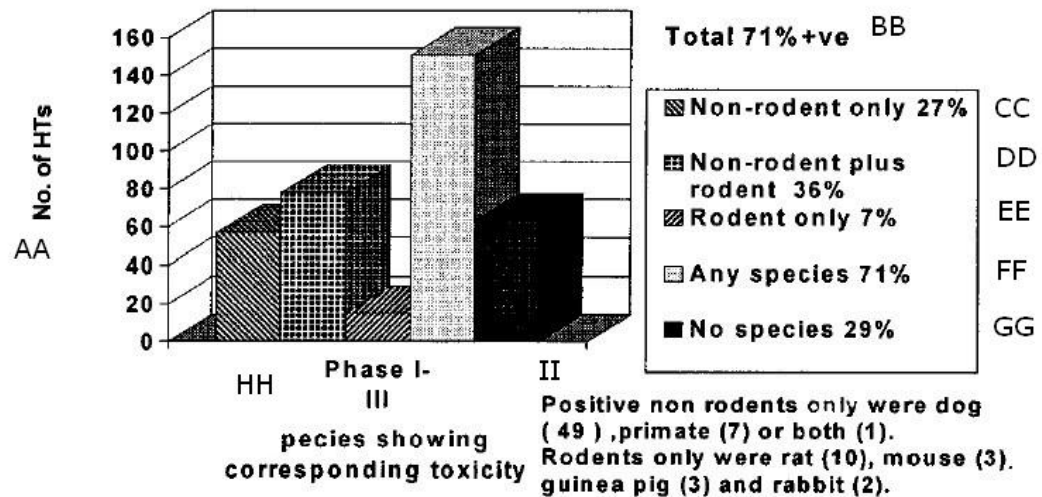


Figure 4 : La figure 3 de l'étude Olson

- AA Nombre de TH
- BB Total 71 % + ve
- CC Non-rongeurs uniquement 27 %
- DD Rongeurs et non-rongeurs 36 %
- EE Rongeurs uniquement 7 %
- FF Toutes les espèces 71 %
- GG Aucune espèce 29 %
- HH Espèces présentant une toxicité correspondante lors des phases I à III
- II Les cas de non-rongeurs positifs uniquement étaient des chiens (49), des primates (7) ou les deux (1). Les cas de rongeurs uniquement étaient des rats (10), des souris (3), des cobayes (3) et des lapins (2).

Un groupe de cliniciens de sociétés participant à l'étude a développé des critères visant à inclure les TH « significatives » dans l'analyse. Pour pouvoir être prise en compte, une TH devait (a) être responsable de l'arrêt du développement, (b) avoir donné lieu à une limitation de la dose, (c) avoir nécessité une surveillance du taux de médicament et un éventuel ajustement de la dose, ou (d) avoir conduit à limiter la population de patients cibles. Le seuil de gravité des TH pouvait être modulé en fonction de la classe thérapeutique du composé (médicament anticancéreux ou anti-inflammatoire, par exemple). Ceci permettait d'exclure la myriade « d'effets secondaires » moindres qui accompagnent toujours le développement de nouveaux médicaments mais ne sont pas suffisants pour restreindre ce développement. Les jugements des cliniciens impliqués étaient définitifs quant à la validité d'inclure ou non un composé. Étaient relevés la phase d'essais cliniques au cours de laquelle la TH avait été détectée pour la première fois, ainsi que la relation ou non de la TH avec la pharmacologie. Les TH étaient classées en fonction du système d'organes concerné et des symptômes détaillés, conformément à la nomenclature standard (COSTART, National Technical Information Service, 1999) (...)

Concordance avec une ou plusieurs espèces : globale et par TH. Globalement, le taux de concordance positive réelle (sensibilité) était de 70 % pour une ou plusieurs espèces animales modèles au stade préclinique (d'étude de sûreté pharmacologique ou d'étude de sûreté toxicologique) montrant des signes de toxicité sur organe cible sur les mêmes systèmes d'organes que les TH [Voir la figure 4].

La présente étude n'a pas tenté d'évaluer le caractère prédictif des données expérimentales précliniques pour l'homme. Elle s'est contentée d'évaluer la concordance entre les effets toxiques détectés au stade clinique et les données issues de l'expérimentation animale (toxicologie préclinique) [78]. (Souligné par nous.)

L'étude Olson, comme il est dit plus haut, a été utilisée par les chercheurs pour justifier leurs affirmations quant à l'utilité prédictive des modèles animaux. Nous pensons toutefois que cette étude offre moins de substance qu'il n'y paraît, pour les raisons suivantes :

1. L'étude a été réalisée et publiée majoritairement par l'industrie pharmaceutique. Ce fait seul n'invalide pas l'étude ; toutefois, il ne faut pas perdre de vue le fait qu'elle a été élaborée par des parties qui avaient un intérêt direct dans ses conclusions. Si ce fait était le seul élément fâcheux, on pourrait peut-être ne pas en tenir compte ; or nous allons voir que l'étude est entachée de défauts plus sérieux.

2. L'étude indique dès le départ que son objectif est de mesurer la fiabilité prédictive des modèles animaux. Les auteurs reconnaissent plus loin que leurs méthodes ne sont en réalité pas capables de réaliser cet objectif. Ce qui nous pousse à nous demander combien de personnes, parmi toutes celles qui citent cette étude, l'ont réellement lue dans son intégralité.

3. Les auteurs de l'étude ont inventé une nouvelle terminologie statistique pour décrire les résultats. Le terme clé ici est le « taux de concordance positive réelle », qui ressemble à « valeur prédictive positive réelle » (ce qui aurait dû être mesuré par l'étude mais ne l'a pas été).

Une recherche Google sur l'expression anglaise « true positive concordance rate » donne douze résultats (en comptant les résultats similaires), qui font tous référence à l'étude Olson (voir la figure 5). Au moins sept des douze résultats trouvés par Google qualifient l'expression « true positive concordance rate » du terme « sensibilité », qui est lui un concept statistique bien connu. En effet, ces termes sont synonymes. Les auteurs de l'étude savent sans doute que la « sensibilité » ne mesure pas la « valeur prédictive réelle ». Il faudrait en outre disposer d'informations sur la « spécificité », etc., pour pouvoir déterminer ladite valeur prédictive réelle. Si l'étude Olson s'est contentée de mesurer la sensibilité, alors ses conclusions sont pour l'essentiel sans rapport avec la question centrale qui est celle de la prédiction.

4. Tout animal ayant réagi de la même manière que les êtres humains a été compté comme résultat positif. Donc, si sur six espèces testées, l'une d'entre elles réagissait comme les êtres humains, elle était comptée comme résultat positif. L'étude Olson a donc porté principalement non sur la prédiction, mais sur la reproduction a posteriori de résultats préalablement connus chez l'homme.

5. Seuls des médicaments ayant atteint le stade des essais cliniques ont été étudiés. Une grande partie des médicaments testés n'atteignent en réalité jamais ce stade, car ils échouent lors des essais sur les animaux.

6. « Ceci permettait d'exclure la myriade « d'effets secondaires moindres qui accompagnent toujours le développement de nouveaux médicaments mais ne sont pas suffisants pour restreindre ce développement. » Un effet secondaire moindre est un effet qui touche quelqu'un d'autre. Par exemple, si la toxicité hépatique est un effet secondaire majeur, les effets secondaires moindres (mais non moins importants pour le patient) sont les nausées sévères, les bourdonnements d'oreille, la pleurésie, les maux de tête, etc. Ce qui conduit à se demander si les critères utilisés pour classer les effets secondaires en tant qu'effet principal ou effet mineur (moindre) avaient bénéficié d'une quelconque validation scientifique indépendante.

7. Même si toutes les données sont bonnes – et il se peut qu'elles le soient – une sensibilité (c'est-à-dire un taux de concordance positive réelle) de 70 % ne permet pas de répondre à la question de la prédiction. La sensibilité n'est pas synonyme de prédiction, et même si l'on considérait que ces 70 % représentaient la valeur de prédiction positive, alors une valeur égale à 70 % serait inadéquate pour prédire la réaction chez l'homme. Dans les études de carcinogénicité, la sensibilité avec les rongeurs pourrait être de 100 % que la spécificité serait une toute autre histoire. C'est la raison pour laquelle il n'est pas possible de dire que les rongeurs permettent de prédire le résultat chez l'homme, dans ce contexte biomédical particulier.

L'étude Olson n'est pas dénuée d'intérêt, mais même dans ses modalités, elle n'apporte aucun argument démontrant la nature prédictive des modèles animaux pour l'homme. Nous pensons donc qu'il faut la citer avec prudence. Une recherche de citations (également sur Google, réalisée 23 juillet 2008), a trouvé 114 citations de cette étude. Nous nous demandons si toutes ces citations font preuve de prudence.

Web Images Maps News Shopping Gmail more Sign in

Advanced Search Preferences

Web Books Results 1 - 12 of 12 for " true positive concordance rate" . (0.30 seconds)

Concordance of the Toxicity of Pharmaceuticals in Humans and in ...
 Overall, the true positive concordance rate (sen- sitivity) was 70% for one or more preclinical animal. TABLE 2. Distribution of Compounds by Therapeutic ...
[linkinghub.elsevier.com/retrieve/pii/S027323000913990](#) - Similar pages

European Journal of Pharmaceutical Sciences : Early microdose drug ...
 There was a true positive concordance rate of 71% for comparable target organs in rodent plus non-rodent studies and identified human toxicities for 150 ...
[linkinghub.elsevier.com/retrieve/pii/S092809870300040X](#) - Similar pages

Histopathology of Preclinical Toxicity Studies: Interpretation and ... - Google Books Result
 by Peter Greaves - 2007 - Medical - 953 pages
 Overall, the true positive concordance rate (sensitivity) is of the order of 70% with 30% of human toxicities not predicted by safety pharmacology or ...
[books.google.com/books?isbn=0444527710...](#)

First dose of potential new medicines to humans: how animals help ...
 Overall, the true positive concordance rate (sensitivity) of the data derived from conventional studies is of the order of 70%, with 30% of human toxicities ...
[www.nature.com/nrd/journal/v3/n3/full/nrd1329.html](#) - Similar pages

First dose of potential new medicines to humans: how animals help ...
 Overall, the true positive concordance rate (sensitivity) of the data derived from conventional studies is of the order of 70%, with 30% of human toxicities ...
[www.nature.com/nrd/journal/v3/n3/full/nrd1329.html?sessionId=AC7F28279806834829F7F100EEEE4E6E](#) - Similar pages

7100048a 440.447
 showed the true positive concordance rate of 70%. for rodent and no n-rodent species, with non-rodents. alone being predictive for 63% of HTs and rodents ...
[www.ingentaconnect.com/content/sage/ het/2000/00000019/00000008/7100048a?crawler=true](#) - Similar pages

7100048a 440.447
 showed the true positive concordance rate of 70%. for rodent and no n-rodent species, with non-rodents. alone being predictive for 63% of HTs and rodents ...

Google

[www.ingentaconnect.com/content/ am/het/2000/00000019/00000008/7100048a](#) - Similar pages

First dose of potential new medicines to humans: how animals help ...
 Overall, the true positive concordance rate (sensitivity) of the data derived from conventional studies is of the order of 70%, with 30% of human toxicities ...
[www.nature.com/nrd/journal/v3/n3/full/nrd1329.html](#) - Similar pages

"Pre-clinical Safety Evaluation". In: Pharmacovigilance
 Overall, the true positive concordance rate was. 70% for the pre-clinical animal species to show. target organ toxicity in the same organ system as ...
[doi.wiley.com/10.1002/0470853093.ch5](#) - Similar pages

First dose of potential new medicines to humans: how animals help ...
 Overall, the true positive concordance rate (sensitivity) of the data derived from conventional studies is of the order of 70%, with 30% of human toxicities ...
[www.emboj.org/nrd/journal/v3/n3/full/nrd1329.html](#) - Similar pages

First dose of potential new medicines to humans: how animals help ...
 Overall, the true positive concordance rate (sensitivity) of the data derived from conventional studies is of the order of 70%, with 30% of human toxicities ...
[intl.emboj.org/nrd/journal/v3/n3/full/nrd1329.html](#) - Similar pages

FIRST DOSE OF POTENTIAL NEW MEDICINES TO HUMANS: HOW ANIMALS HELP
 File Format: PDF/Adobe Acrobat - View as HTML
 true positive concordance rate (sensitivity) of the data. derived from conventional studies is of the order of. 70%, with 30% of human toxicities not ...
[npg.nature.com/nrd/journal/v3/n3/pdf/nrd1329.pdf](#) - Similar pages

Search within results | Language Tools | Search Tips | Dissatisfied? Help us improve | Try Google Experimental
 Google Home - Advertising Programs - Business Solutions - Privacy - About Google

Figure 5 : Résultats de recherche Google

Conclusion

Mark Kac déclarait, « une preuve est ce qui convainc une personne sensée. » Même si ce n'est pas à nous de prouver que les modèles animaux ne sont *pas* prédictifs, nous pensons avoir présenté ici des preuves à même de convaincre toute personne sensée.

Il existe des différences quantitatives et qualitatives entre les espèces. Ceci n'est pas surprenant, compte-tenu de ce que nous savons de la biologie de l'évolution et développementale, de la régulation et de l'expression géniques, de l'épigénétique, de la théorie de la complexité et de la génomique comparative. Les hypothèses génèrent des prédictions, qui sont ensuite confirmées ou infirmées. Le terme prédire a une signification bien précise dans le domaine scientifique, et d'aucuns affirment qu'il est à la base même de la science. Prédire ne signifie par trouver a posteriori un animal qui réagit comme l'être humain à des stimuli donnés, et en déduire que cet animal prédit la réaction chez l'homme ; il n'est pas non plus synonyme de sélection soigneuse des données, ni de tomber juste de temps à autre.

Lorsque l'on tient pour vrai un concept tel que « les modèle animaux peuvent prédire la réaction chez l'homme », ce concept n'est plus une hypothèse. Nous l'avons considéré comme une hypothèse généralisée, mais on aurait aussi pu l'appeler un postulat infondé. Un postulat ou une hypothèse généralisée peuvent être vrais, mais leur véracité doit être prouvée. Si une modalité telle que l'expérimentation animale ou l'emploi des animaux pour prédire la pathophysiologie des maladies humaines est qualifiée de prédictive, alors toutes les données générées par ladite modalité doivent présenter

une probabilité très élevée d'être vraie chez les humains. Or les modèles animaux des maladies et des réactions aux médicaments ne remplissent pas cette condition.

Dans le domaine médical, même une valeur de prédiction positive de 0,99 peut s'avérer inadéquate pour certains essais – et les modèles animaux ne permettent même pas d'approcher une telle valeur. Les modèles animaux ne sont pas des outils de prédiction des réactions chez l'homme. Il arrive que des animaux réagissent de la même manière que les humains à un stimulus donné. Mais comment savoir a priori quel animal reproduira les réactions humaines ? Ceux qui persistent à attribuer une valeur prédictive aux animaux confondent sensibilité et prédiction. Les animaux en tant que groupe sont extrêmement sensibles pour la carcinogénicité ou d'autres phénomènes biologiques. Soumettez cent souches ou espèces différentes à des essais, il est très probable que l'une d'entre elle réagisse comme les humains. Mais la spécificité est très faible ; les valeurs de prédiction positive et négative également. (Même si la science décidait d'abandonner l'utilisation correcte au plan historique du verbe prédire, à chaque fois qu'un défenseur du modèle animal déclarait qu'une espèce animale a prédit une réaction humaine, il lui faudrait également admettre que les espèces animales A, B, C, D, E, etc. ont donné des prédictions erronées. De fait, la justification de l'emploi d'animaux pour la raison que les modèles animaux *per se* augmentent la sécurité de nos médicaments ou prédisent des faits relatifs aux maladies chez l'homme ne serait pas vraie.)

Certains ont suggéré qu'il ne faut pas critiquer les modèles animaux si nous n'avons pas de meilleure suggestion pour la recherche et les essais [27]. Ce n'est pas à nous de reporter nos critiques du caractère non prédictif des modèles animaux en attendant que des modèles prédictifs *in silico*, *in vitro* ou *in vivo* soient disponibles. Ce n'est pas non plus à nous d'inventer de tels moyens. L'astrologie n'est pas prédictive au sens qu'elle ne peut pas dire l'avenir, et nous la critiquons pour cette raison, bien que nous n'ayons aucune piste pour proposer un moyen similaire de prédire l'avenir.

Certains suggèrent aussi que les modèles animaux deviendront peut-être un jour prédictifs, et que nous devrions admettre cette possibilité. Bien que cela soit vrai, car tout est possible, une telle éventualité semble tout de même très improbable, car on a constaté que les organismes génétiquement modifiés souffrent des mêmes problèmes de prédiction que ceux évoqués ici [16,81-87] et que d'autre part, comme nous l'avons mentionné, des êtres humains différents présentent des réactions très différentes aux médicaments et aux maladies. Compte-tenu de notre connaissance actuelle des systèmes complexes et de l'évolution, il serait surprenant qu'une espèce puisse être utilisée pour prédire un résultat chez une autre, étant donné le niveau de précision des études actuelles des maladies et des médicaments, et l'exactitude que la société exige des sciences médicales.

Le malentendu autour de la signification du terme prédiction a des conséquences directes et indirectes. Si nous devions retirer du marché tous les produits chimiques ou médicaments présentant des effets cancérigènes, tératogènes ou secondaires graves chez une quelconque espèce, il n'y aurait plus aucun médicament ni aucun produit chimique sur le marché. Cela a également un coût d'écarter du marché des composés chimiques utiles par ailleurs. Nous perdons des traitements, parfois même des moyens de guérison ; de possibles gains financiers ; sans parler de la somme de connaissances nouvelles qu'il aurait été possible de recueillir sur le composé en question. Ces pertes ne sont pas négligeables. Nos connaissances actuelles de la médecine personnalisée nous montrent que les êtres humains réagissent différemment aux maladies et aux médicaments, et qu'un être humain ne permet pas de prédire l'effet d'un médicament sur un autre être humain ; il semble donc illogique de chercher des modèles prédictifs basés sur des espèces totalement différentes de l'espèce humaine. Si nous voulons réellement disposer de méthodes de recherches et d'essais prédictifs (ce que nous souhaitons), alors il semblerait logique de commencer à chercher à l'intérieur de l'espèce, et non dans d'autres espèces.

Intérêts concurrents

Les auteurs déclarent l'absence d'intérêts concurrents.

Contributions des auteurs

Tous les auteurs ont contribué de manière égale au présent article, et ont lu et approuvé le manuscrit.

À propos des auteurs

Le professeur Niall Shanks, PhD, est titulaire de la chaire Curtis D. Gridley d'histoire et philosophie des sciences à l'université d'état de Wichita, où il est également professeur de philosophie, maître de conférence en sciences biologiques, maître de conférence en physique, et membre associé du corps professoral de deuxième cycle. Il était président, en 2008/2009 de la branche Southwest & Rocky Mountain de l'American Association for the Advancement of Science. Il est titulaire d'un doctorat en philosophie de l'université d'Alberta.

Le docteur Ray Greek a obtenu son doctorat de médecine à l'Université d'Alabama à Birmingham, puis sa spécialisation en anesthésiologie à l'université du Wisconsin à Madison. Il a enseigné à l'université du Wisconsin et à l'université Thomas Jefferson de Philadelphie. Il est actuellement président de l'association à but non lucratif Americans For Medical Advancement <http://www.curedisease.com>.

Jean Greek, docteur en médecine vétérinaire, est diplômée de l'école vétérinaire de l'université du Wisconsin à Madison et a obtenu une spécialisation en dermatologie à l'université de Pennsylvanie. Elle a enseigné à l'université du Missouri et exerce à présent son activité à titre libéral.

Remerciements

Cet article a pu voir le jour notamment grâce à une subvention de la National Anti-Vivisection Society of Chicago Illinois, É.-U., <http://www.navs.org>, accordée à l'AFMA (Americans For Medical Advancement) <http://www.curedisease.com>, association à but non lucratif et à visée éducative, dont la position sur l'utilisation des animaux est clairement résumée dans le présent article. Contrairement à la NAVS, l'AFMA n'est pas opposée à toutes les recherches utilisant des animaux. Tous les auteurs sont des membres du comité exécutif ou du comité directeur de l'AFMA.

Références bibliographiques

1. Sarewitz D, Pielke RA Jr. **Prediction in Science and Policy**. Dans *Prediction: Science, Decision Making, and the Future of Nature* Dirigé par Sarewitz D, Pielke RA Jr, Byerly R Jr. Island Press. 2000:11-22.
2. Quine W: *Quiddities" An Intermittently Philosophical Dictionary* Cambridge : The Belknap Press of Harvard University Press. 2005.
3. LaFollette H, Shanks N. *Brute Science: Dilemmas of animal experimentation*. London et New York : Routledge. 1996.
4. Greek R, Greek J. *Specious Science*. New York : Continuum Int. 2002.
5. **Xceleron** [<http://www.xceleron.com/metadot/index.pl>]
6. Altman L. *Who Goes First? The Story of Self-Experimentation in Medicine*. University of California Press. 1998.
7. Bruder CE, Piotrowski A, Gijssbers AA et al. **Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles**. *Am J Hum Genet* 2008, **82**:763-771.
8. Fraga MF, Ballestar E, Paz MF et al. **Epigenetic differences arise during the lifetime of monozygotic twins**. *Proc Natl Acad Sci USA*. 2005. **102**:10604-10609.
9. Weiss ST, McLeod HL, Flockhart DA, et al. **Creating and evaluating genetic tests predictive of drug response**. *Nat Rev Drug Discov* 2008, **7**:568-574.
10. Kaiser J. **Gender in the pharmacy: does it matter?** *Science* 2005. **308**:1572.
11. Willyard C. **Blue's clues**. *Nat Med* 2007. **13**:1272-1273.
12. Couzin J. **Cancer research. Probing the roots of race and cancer**. *Science* 2007. **315**:592-594.
13. Holden C. **Sex and the suffering brain**. *Science* 2005. **308**:1574.
14. Salmon W. **Rational Prediction**. *Philosophy of Science*. 1998:433-444.
15. Butcher EC. **Can cell systems biology rescue drug discovery?** *Nat Rev Drug Discov* 2005, **4**:461-467.
16. Horrobin DF. **Modern biomedical research: an internally selfconsistent universe with little contact with medical reality?** *Nat Rev Drug Discov*. 2003. **2**:151-154.
17. Pound P et al. **Where is the evidence that animal research benefits humans?** *BMJ* 2004, **328**:514-517.
18. Éditorial. **The time is now**. *Nat Rev Drug Discov*. 2005. **4**:613.
19. Littman BH, Williams SA: **The ultimate model organism: progress in experimental medicine**. *Nat Rev Drug Discov*. 2005, **4**:631-638.
20. Uehling M. **Microdoses of Excitement over AMS, 'Phase 0' Trials**. *Bio-IT World* 2006. **2006**.
21. Dixit R, Boelsterli U. **Healthy animals and animal models of human disease(s) in safety assessment of human pharmaceuticals, including therapeutic antibodies**. *Drug Discovery Today*. 2007. **12**:336-342.
22. Greek R, Greek J. *Sacred Cows and Golden Geese: The Human Cost of Experiments on Animals*. New York : Continuum Int; 2000.
23. Greek J, Greek R. *What Will We Do if We Don't Experiment on Animals*. Continuum. 2004.
24. **FDA Issues Advice to Make Earliest Stages Of Clinical Drug Development More Efficient** [<http://www.fda.gov/bbs/topics/news/2006/NEW01296.html>]
25. Knight A. **The beginning of the end for chimpanzee experiments?** *Philos Ethics Humanit Med*. 2008. **3**:16.
26. Shanks N, Pyles RA. **Evolution and medicine: the long reach of "Dr. Darwin"**. *Philos Ethics Humanit Med*. 2007. **2**:4.
27. Vineis P, Melnick R. **A Darwinian perspective: right premises, questionable conclusion. A commentary on Niall Shanks and Rebecca Pyles' "evolution and medicine: the long reach of "Dr. Darwin"'**. *Philos Ethics Humanit Med*. 2008. **3**:6.
28. **Débat : Animals are predictive for humans** [<http://video.google.com/videoplay?docid=-8464924004908818871&q=madison+debate+animal&total=5&start=0&num=30&so=0&type=search&plindex=0>]
29. Gad S. **Préface**. Dans *Animal Models in Toxicology* Dirigé par Gad S. CRC Press. 2007:1-18.
30. Fomchenko EI, Holland EC. **Mouse models of brain tumors and their applications in preclinical trials**. *Clin Cancer Res*. 2006, **12**:5288-5297.
31. Hau J. **Animal Models**. Dans *Handbook of Laboratory Animal Science Animal Models Volume II*. 2^e édition. Dirigé par : Hau J, van Hoosier GK Jr. CRC Press. 2003 :2-8.
32. Comité de rédaction : **Of Mice...and Humans**. *Drug Discovery and Development* 2008. **11**:16-20.
33. **FDA panel recommends continued use of controversial diabetes drug** [<http://www.cnn.com/HEALTH/9903/26/rezulin.review.02/index.html>]
34. Masubuchi Y. **Metabolic and non-metabolic factors determining troglitazone hepatotoxicity: a review**. *Drug Metab Pharmacokinet*. 2006. **21**:347-356.
35. Topol EJ. **Failing the public health – rofecoxib, Merck, and the FDA**. *N Engl J Med*. 2004. **351**:1707-1709.
36. Heywood R. **Clinical Toxicity – Could it have been predicted? Post-marketing experience**. *Animal Toxicity Studies: Their Relevance for Man*. 1990:57-67.

37. Spriet-Pourra C, Auriche M. **Drug Withdrawal from Sale.** *New York* 2^e édition. 1994.
38. Igarashi T. **The duration of toxicity studies required to support repeated dosing in clinical investigation – A toxicologists opinion.** Dans *CMR Workshop: The Timing of Toxicological Studies to Support Clinical Trials.* Dirigé par : Parkinson CNM, Lumley C, Walker SR. Boston/UK : Kluwer. 1994:67-74
39. Sankar U. **The Delicate Toxicity Balance in Drug Discovery.** *The Scientist.* 2005. **19**:32.
40. Wilbourn J, Haroun L, Heseltine E et al. **Response of experimental animals to human carcinogens: an analysis based upon the IARC Monographs programme.** *Carcinogenesis.* 1986. **7**:1853-1863.
41. Rall DP. **Laboratory animal tests and human cancer.** *Drug Metab Rev.* 2000. **32**:119-128.
42. Tomatis L et al. **Human carcinogens so far identified.** *Jpn J Cancer Res.* 1989. **80**:795-807.
43. Knight A, Bailey J, Balcombe J. **Animal carcinogenicity studies: 1. Poor human predictivity.** *Altern Lab Anim.* 2006, **34**:19-27.
44. Tomatis L, Wilbourn L. **Evaluation of carcinogenic risk to humans: the experience of IARC.** Dans *New Frontiers in Cancer Causation.* Dirigé par : Iversen. Washington, DC : Taylor and Francis. 2003:371-387.
45. IARC : *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans* Lyon : IARC. 1972.
46. **IARC monographs programme on the evaluation of carcinogenic risks to humans** [http://monographs.iarc.fr]
47. Haseman JK. **Using the NTP database to assess the value of rodent carcinogenicity studies for determining human cancer risk.** *Drug Metab Rev.* 2000. **32**:169-186.
48. Gold LS, Slone TH, Ames BN. **What do animal cancer tests tell us about human cancer risk? Overview of analyses of the carcinogenic potency database.** *Drug Metab Rev.* 1998. **30**:359-404.
49. Freeman M, St Johnston D. **Wherefore DMM? Disease Models & Mechanisms.** 2008. **1**:6-7.
50. Schardein J. *Drugs as Teratogens.* CRC Press. 1976.
51. Schardein J. *Chemically Induced Birth Defects.* Marcel Dekker. 1985.
52. Manson J, Wise D. **Teratogens. Casarett and Doull's Toxicology.** 4^e édition. 1993:228.
53. Caldwell J. **Comparative Aspects of Detoxification in Mammals.** Dans *Enzymatic Basis of Detoxification Volume 1.* Dirigé part : Jakoby W. New York : Academic Press. 1980.
54. Runner MN. **Comparative pharmacology in relation to teratogenesis.** *Fed Proc.* 1967. **26**:1131-1136.
55. Keller SJ, Smith MK. **Animal virus screens for potential teratogens. I. Poxvirus morphogenesis.** *Teratog Carcinog Mutagen.* 1982. **2**:361-374.
56. Staples RE, Holtkamp DE. **Effects of Parental Thalidomide Treatment on Gestation and Fetal Development.** *Exp Mol Pathol.* 1963. **26**:81-106.
57. Caldwell J. **Problems and opportunities in toxicity testing arising from species differences in xenobiotic metabolism.** *Toxicol Lett.* 1992. **64–65**(n° spéc.) :651-659.
58. Wall RJ, Shani M. **Are animal models as good as we think?** *Theriogenology.* 2008, **69**:2-9.
59. Curry SH. **Why have so many drugs with stellar results in laboratory stroke models failed in clinical trials? A theory based on allometric relationships.** *Ann N Y Acad Sci.* 2003. **993**:69-74. Discussion 79–81
60. Shubick P. **Statement of the Problem.** Dans : *Human Epidemiology and Animal Laboratory Correlations in Chemical Carcinogenesis.* Dirigé par : Coulston F, Shubick P. Ablex Pub. 1980:5-17.
61. Coulston F. **Final Discussion.** Dans *Human Epidemiology and Animal Laboratory Correlations in Chemical Carcinogenesis* Dirigé par : Coulston F, Shubick P. Ablex. 1980:407.
62. Council on Scientific Affairs : **Carcinogen regulation.** *JAMA.* 1981. **246**:253-256.
63. Salsburg D. **The lifetime feeding study in mice and rats – an examination of its validity as a bioassay for human carcinogens.** *Fundam Appl Toxicol.* 1983. **3**:63-67.
64. IARC : **IARC Working group on the evaluation of carcinogenic risks to humans.** *Lyon.* 1972. **1–78**..
65. Sloan DA et al. **Increased incidence of experimental colon cancer associated with long-term metronidazole therapy.** *Am J Surg.* 1983. **145**:66-70.
66. Clemmensen J, Hjalgrim-Jensen S. **On the absence of carcinogenicity to man of phenobarbital.** Dans *Human Epidemiology and Animal Laboratory Correlations in Chemical Carcinogenesis.* Dirigé par : Alex Pub. Coulston F, Shubick S. 1980:251-265.
67. Clayton D. **The carcinogenic action of drugs in man and animals.** Dans *Human Epidemiology and Animal Laboratory Correlations in Chemical Carcinogenesis* Dirigé par : Coulston F, Shubick P. Ablex Pub. 1980:185-195.
68. Anisimov V. *Carcinogenesis and Aging.* Boca Rotan. CRC Press. 1987.
69. Anisimov V. *Molecular and Physiological Mechanisms of Aging.* Saint Pétersbourg : Nauka. 2003.
70. Dilman VM, Anisimov VN. **Effect of treatment with phenformin, diphenylhydantoin or L-dopa on life span and tumour incidence in C3H/3Sn mice.** *Gerontology.* 1980. **26**:241-246.
71. IARC. **Some aromatic amines, hydrazine and related substances, n-nitroso compounds and miscellaneous alkylating agents.** *IARC monograph on the evaluation of carcinogenic risks to humans.* Lyon : 1974. **4**.
72. Anisimov VN, Ukraintseva SV, Yashin AI. **Cancer in rodents: does it tell us about cancer in humans?** *Nat Rev Cancer.* 2005. **5**:807-819.
73. Hahn WC, Weinberg RA. **Modelling the molecular circuitry of cancer.** *Nat Rev Cancer.* 2002. **2**:331-341.
74. Rangarajan A, Weinberg RA. **Opinion: Comparative biology of mouse versus human cells: modelling human cancer in mice.** *Nat Rev Cancer.* 2003. **3**:952-959.
75. Anisimov VN. **Age as a risk in multistage carcinogenesis.** Dans *Comprehensive Geriatric Oncology* 2^e édition. Dirigé par : Balducci L, Ershler WB, Lyman GH. M E : Informa Healthcare. Taylor and Francis. 2004:75-101. 157–178.
76. Overmier JB, Carroll ME. **Basic Issues in the Use of Animals in Health Research.** Dans *Animal Research and Human Health.* Dirigé par : Carroll ME, Overmier JB. American Psychological Association. 2001:5.
77. Kolata G. **2 Top Diet Drugs Are Recalled Amid Reports of Heart Defects.** *New York Times.* New York : 1997.
78. Olson H, Betton G, Robinson D et al. **Concordance of the toxicity of pharmaceuticals in humans and in animals.** *Regul Toxicol Pharmacol.* 2000. **32**:56-67.
79. Shanks N, Greek R, Nobis N, Greek J. **Animals and Medicine: Do Animal Experiments Predict Human Response?** *Skeptic.* 2007. **13**:44-51.
80. Conn P, Parker J. **Letter. Animal research wars.** *Skeptic* 2007. **13**:18-19.
81. Van Regenmortel MH. **Reductionism and complexity in molecular biology. Scientists now have the tools to unravel biological and overcome the limitations of reductionism.** *EMBO Rep.* 2004. **5**:1016-1020.
82. Morange M. **A successful form for reductionism.** *The Biochemist.* 2001. **23**:37-39.
83. Morange M. *The misunderstood gene.* Cambridge : Harvard University Press. 2001.
84. Mepham TB, Combes RD, Balls M et al. **The Use of Transgenic Animals in the European Union: The Report and Recommendations of ECVAM Workshop 28.** *Altern Lab Anim.* 1998. **26**:21-43.
85. Liu Z, Maas K, Aune TM. **Comparison of differentially expressed genes in T lymphocytes between human autoimmune disease and murine models of autoimmune disease.** *Clin Immunol.* 2004. **112**:225-230.

86. Dennis C. **Cancer: off by a whisker.** *Nature.* 2006. **442**:739-741.
87. Houdebine LM. **Transgenic animal models in biomedical research.** *Methods Mol Biol.* 2007. **360**:163-202.