**La pêche aux alphas, niveau 2 : cours de perfectionnement**

Nous avons vu, dans un article précédent, comment un astrologue pouvait montrer (presque) scientifiquement un lien entre signe astrologique et cancer, grâce à une technique infaillible : la pêche aux alphas, c’est-à-dire la mise en avant de résultats statistiquement significatifs, mais obtenus en fait uniquement par l’effet du hasard.

Pour être franc, la méthode employée dans cet exemple était quand même un peu basique. S’il veut faire carrière, un scientifique doit parfois convaincre des publics un peu moins conciliants que les jurys de thèse d’Elisabeth Tessier ou des frères Bogdanov. Nous allons donc étudier aujourd’hui une méthode de pêche aux alphas plus sophistiquée, qui nous permettra cette fois de passer le barrage du comité de lecture d’une revue de bonne réputation. Vantardise ? Pas du tout : l’exemple que nous allons voir provient d’un article authentique[[1]](#footnote-1), publié dans la revue Environmental Health Perspective (impact factor de 7,977 en 2014). Un article tellement sérieux qu’il a été cité par le magazine La Recherche en mars 2016, dans son dossier « accablant » sur les pesticides. Il a également inspiré l’enquête Exppert 6 de Générations Futures, ce qui confirme son impact dans les plus hautes sphères scientifiques (non, là on plaisante…).

**Le principe de l’étude**

L’étude décrite dans cet article part d’une idée intéressante : utiliser les aspirateurs comme instrument de mesure des pollutions domestiques. Les chercheurs ont récupéré le contenu des sacs à poussière des aspirateurs de 682 personnes atteintes de lymphôme non-hodgkinien (LNH), une forme de cancer des cellules sanguines. Ils ont recherché dans ces poussières les traces de 27 produits cancérogènes avérés ou soupçonnés. Ils ont ensuite comparé ces résultats avec ceux obtenus dans les appartements de personnes en bonne santé, vivant dans les mêmes régions. Cette étude a été réalisée dans 4 sites : Los Angeles, Seattle, Detroit, et l’Etat agricole de l’Iowa. Les auteurs ont d’abord déterminé, pour chaque produit individuellement, si les cas de LNH étaient plus nombreux dans les appartements les plus contaminés par ce produit. Leurs résultats sont exprimés en Odds Ratio (OR) : un OR de 1,30 signifie que les personnes atteintes de LNH sont plus nombreuses de 30% dans les appartements les plus contaminés par le produit considéré, par rapport à celles habitant dans les appartements les moins contaminés par ce produit ; à l’inverse, un OR de 0,70 signifie que l’on a un déficit de 30% de LNH.

Avant de passer aux résultats, notons une particularité intrigante de cet article : il date de 2015, alors que l’étude de terrain date de 1999/2001. Pourquoi un tel délai ? Nous verrons l’explication dans la suite…

**Les résultats, première lecture : un cas typique de pêche aux alphas**

Les auteurs ont étudié 27 produits chimiques dans 4 sites différents, ce qui nous fait au total 108 couples (produit x site). Cela nous fait un vivier assez grand pour que nous ayons de bonnes chances d’y pêcher quelques alphas bien frais : **même si l’ensemble des produits étaient parfaitement inoffensifs, nous devrions normalement en trouver 4 avec des OR significativement différents de 1**, pour le risque de 1ère espèce  = 5%. Dans ce cas, on devrait également trouver une distribution symétrique des résultats, avec autant d’OR supérieurs à 1 que d’OR inférieurs à 1 (y compris parmi les résultats « significatifs »). S’il y a réellement des produits ayant eu un effet cancerignène dans nos sacs à poussière, on devrait trouver un excès de résultats significatifs, mais seulement pour les OR >1. Et, de façon générale, on devrait également trouver plus d’OR>1 que d’OR <1. Qu’en est-il dans les résultats de nos auteurs ?

Une petite surprise nous attend quand nous recherchons ces OR par produit et par site. Bien qu’il s’agisse du résultat fondamental de l’étude, ils ne figurent pas dans le corps de l’article : il faut aller les chercher dans le « Supplemental Material ». **Les auteurs ont ici appliqué la règle d’or du dur métier de pêcheur d’alphas : mettre dans les annexes, et en vrac, tous les résultats embarrassants, qui pourraient détourner l’attention du lecteur des alphas bien frétillants mis en avant dans l’article lui-même.** Ils obligent ainsi les reviewers (s’ils en ont le courage…) à regrouper les tableaux S1 et S2, et à sortir leurs stabilos pour s’y retrouver un peu. Après cet effort de rangement, nous obtenons le tableau suivant :



Les OR supérieurs à 1 ont été surlignés en rouge, les OR inférieurs à 1 en vert, **ce qui nous permet de voir tout de suite des différences de comportement frappantes selon les produits** :

* Les 5 premiers produits sont des polychlorobiphenyles (PCB), des produits (non pesticides) dont le caractère cancerogène est établi de longue date, et qui sont d’ailleurs interdits depuis longtemps (1987 en France). Ces produits ont tous des OR supérieurs à 1 dans chaque site, l’un d’eux, le PCB 180, obtient même un résultat hautement significatif, quand on fait la moyenne de tous les résultats.
* **Pour les 22 autres produits, les résultats ne montrent aucune cohérence**. Nos fidèles lecteurs reconnaitront ces chiffres (s’ils ont une très bonne mémoire) : ce sont ceux que nous avions utilisés dans notre jeu-test [« Statisticien, épidémiologiste, astrologue : quel scientifique êtes-vous ? »](http://www.forumphyto.fr/2016/04/01/pours-sourire-et-sinstruire-statisticien-epidemiologiste-astrologue-quel-scientifique-etes-vous/). Si on fait le décompte des 88 résultats obtenus pour ces 22 produits dans les 4 sites, on trouve une parité parfaite entre les OR supérieurs et inférieurs à 1 :



La distribution des résultats significatifs présente une légère anomalie, avec un excès de résultats significatifs à 5%, mais une proportion quasi normale de résultats significatifs à 10 ou 1%. De plus, ce léger excès de résultats significatifs est symétrique (il y a autant de résultats significatifs parmi les OR>1 que parmi les OR <1), ce qui ne serait pas le cas s’il y avait un ou plusieurs produits cancérogènes cachés parmi les produits inoffensifs.

Si on reprend la grille d’interprétation de notre jeu-test, **tout statisticien aurait donc conclu que les analyses montrent un lien très probable entre contamination aux PCB et LNH (en particulier pour le PCB 180), mais qu’il n’y a aucun indice d’un effet cancérogène pour les 22 autres molécules**. Des épidémiologistes feraient sans doute de même s’il s’agissait d’astrologie. Mais comme on parle ici de produits hautement suspects, ils ouvriraient sans doute le parapluie en notant que 3 organo-chlorés (α-chlordane, γ-chlordane et DDE) ont en moyenne des OR non significatifs, mais relativement élevés, et mériteraient donc des études complémentaires. **C’est d’ailleurs la conclusion à laquelle arrivait la première publication sur ces analyses de poussière, en 2005[[2]](#footnote-2).**

Notons qu’il y a tout de même une différence importante entre les résultats ci-dessus, selon qu’ils s’appliquent au signe astrologique des personnes enquêtées, ou aux produits chimiques trouvés dans leur appartement : chaque personne n’a qu’un signe astrologique, les résultats obtenus par chaque signe astrologique sont donc indépendants entre eux. Ce n’est pas le cas pour les produits chimiques, car certains peuvent être correllés entre eux : si certains produits, dont un seul est cancérigène, sont fréquemment associés, on va trouver une liaison positive avec le LNH pour les produits qui lui sont associés, même s’ils sont inoffensifs. On peut se poser cette question pour les PCB, qui sont des produits chimiquement proches entre eux : les PCB autres que le PCB 180 ont des résultats moins clairs que lui, il est possible que leurs OR soient simplement causés par leur association fréquente avec le PCB 180. En sens inverse, si des produits ont une liaison négative avec le PCB 180 (c’est-à-dire qu’ils sont moins présents là où le PCB 180 est le plus présent), ils vont obtenir un OR <1, ce qui pourrait expliquer le léger excès d’OR significativement inférieurs à 1 que nous avons observé. La question se pose en particulier pour les pesticides à usage exclusivement agricole analysés dans cette étude : le chlorpyrifos, le 2-4 D, le dicamba et le diazanon. Ces produits ont tous quatre des OR inférieurs à 1, proche du seuil de significativité à 5%. Il est donc possible que le LNH dans cette étude ait été causé par un ou des facteurs d’origine plutôt industrielle ou urbaine, et soit moins fréquent dans les secteurs agricoles, ce qui expliquerait la liaison négative avec ces 4 pesticides (les autres pesticides sont ou ont été également utilisés comme insecticides à usage domestique ou pour la protection des charpentes, on peut donc les trouver aussi bien en ville qu’à la campagne).

Il y avait donc de vraies et bonnes raisons de ne pas s’arrêter à l’examen des OR de chaque produit, et de faire des analyses supplémentaires pour redresser les résultats en fonction des corrélations éventuelles entre produits**. C’est l’objectif officiel de cette publication de 2015…mais nous allons voir que la méthode choisie avait également une autre vertu moins avouable : appliquée à ce type de données, c’est un « aspirateur à alphas » aussi efficace que discret…**

**Une nouvelle méthode au secours des pêcheurs d’alphas**

Dans cette publication de 2015, les données ont été réanalysées à l’aide d’une nouvelle méthode statistique **: la weighted quantile sum regression[[3]](#footnote-3)**. Cette méthode vise à hiérarchiser l’effet de chaque molécule dans un cocktail de produits potentiellement dangereux, en éliminant l’effet des corrélations éventuelles entre eux. Dans cette régression, on construit un indice (le weighted quantile sum index, WQSI), qui fait la somme des influences de chaque molécule sur la maladie étudiée, en attribuant à chaque produit un coefficient qui optimise la liaison entre le WQSI et cette maladie. Les produits qui ont le coefficient le plus élevé dans le WQSI sont donc supposés être les plus influents sur la maladie étudiée. Un WQSI a été calculé pour chacun des 4 sites d’étude, et un cinquième pour le regroupement de l’ensemble des résultats. Tous ces WQSI ont un OR significatif ou hautement significatif pour le LNH (sauf celui calculé pour Seattle, mais qui n’est pas très loin d’être significatif) (tableau 2 de Czarnota et al 2015). C’est-à-dire que les personnes habitant dans un logement où le WQSI est élevé ont significativement plus de risques d’être atteints de LNH. **Il semble donc cette fois que nous ayons une bonne méthode pour mettre en évidence un lien entre le cocktail de substances présentes dans les poussières et la maladie étudiée…mais les choses se gâtent quand on regarde le détail des résultats par molécule et par site.**

Dans cette méthode, c’est le coefficient attribué à chaque produit dans le WQSI qui mesure son effet relatif sur le lymphôme non hodgkinien (ou tout au moins son association avec cette maladie). La somme des 27 coefficients calculés pour chaque WQSI est de 1; si tous les produits avaient un effet équivalent, ils auraient donc un coefficient de 0,037 (1/27). Ceux qui ont un coefficient plus élevé sont fortement corrélés au LNH (et donc suspects de le provoquer), et ceux qui ont des coefficients nuls ou très faibles ne montrent aucune liaison avec le LNH. On pourrait donc espérer que cette fois les coefficients du WQSI soient plus cohérents que les OR de la publication de 2005, mais ce n’est pas du tout le cas : **leur distribution est tout aussi anarchique, et même encore plus que celle des OR**, puisque cette fois même les PCB ont des coefficients tantôt supérieurs, tantôt inférieurs à la moyenne, selon les sites :



***Données du tableau 3 de Czarnota et al 2015 : coefficients attribués aux 27 produits testés dans les WQS Index, dans les 4 sites individuellement, puis en moyenne. Les produits ayant un coefficient élevé sont sensés être les plus liés au lymphôme non-hodgkinien. Les coefficients supérieurs à la moyenne sont surlignés en rouge, les autres en vert. En comparant ces données au tableau 1, on constate :***

* ***que la distribution de ces coefficients est tout aussi incohérente que l’était la distribution des OR : un produit peut avoir un fort coefficient dans un site, et très faible dans un autre. Les coeffficients obtenus n’ont donc aucune logique toxicologique***
* ***que pour chaque WQSI, les 3 produits aux coefficients les plus élevés (en gras) totalisent à eux seuls plus de 50% du poids dans la régression WQS, et qu’il s’agit à chaque fois des produits qui avaient eu les OR les plus élevés pour le site étudié.***

***Au bout du compte, la régression WQS fait donc ressortir les mêmes produits que les OR (dont nous avons vu que la distribution était conforme au hasard pour les 22 produits non PCB), et ne fournit pas de test d’évaluation du risque de première espèce qui permettrait de juger de la crédibilité des résultats obtenus.***

Cela jette d’emblée un sérieux doute sur la pertinence de la méthode employée : certes, on a réussi à obtenir dans chaque site un index WQSI fortement corrélé au LNH, mais en appliquant des coefficients ajustés ad hoc pour chaque ville, sans aucune cohérence entre sites. Par exemple, l’insecticide propoxur (matière active du Baygon) a un coefficient très faible dans le WQSI de Los Angeles (le site où il a pourtant été détecté de loin avec les plus fortes concentrations), et très élevé dans le WQSI de l’Iowa (site où les concentrations étaient les plus faibles). **D’un point de vue toxicologique, on a du mal à comprendre comment une molécule pourrait avoir des effets aussi différents d’un site à l’autre.**

Ce manque de cohérence des coefficients des WQSI peut paraitre surprenant, mais il a une explication toute simple, que les auteurs glissent en cours de discussion comme s’il s’agissait d’un détail : « A limitation of WQS regression is that it cannot identify associations in different directions for the components of the index. ». En clair, la régression WQS ne devrait être utilisée que si tous les produits étudiés ont un effet de même sens (c’est-à-dire qu’ils favorisent tous la maladie, ou bien qu’ils en protègent tous). Or nous avons vu que, dans les résultats de base, nous avions très exactement 50% de molécules avec un 0R > 1 (et donc une liaison positive avec le LNH), et 50% avec un OR<1 (donc une liaison négative avec le LNH). Qu’est-il advenu des molécules à OR < 1 ? Les auteurs se félicitent avec une certaine candeur de ce qu’elles ont toutes eu des coefficients négligeables dans le calcul des WQSI. Le contraire aurait été surprenant, vu que par construction la régression ne pouvait pas leur attribuer de coefficient négatif : elle leur a donc attribué des coefficients proches de 0, pour ne pas dégrader la relation entre WQSI et LNH. **On voit donc ici la 1ère vertu de la régression WQS pour les pêcheurs d’alphas : elle élimine automatiquement du WQS Index Ies produits à OR <1. En d’autres termes, elle sélectionne donc les alphas positifs !** De plus, on constate que, dans chaque site, trois produits cumulent à eux seuls plus de la moitié du poids des coefficients du WQSI… et qu’à chaque fois il s’agit précisément de ceux qui avaient eu les OR les plus élevés pour ce site ! **Au bout du compte, l’analyse des coefficients des WQSI n’apporte donc aucun élément nouveau : elle a simplement agi comme un filtre pour faire ressortir exactement les mêmes croisements molécule\*site que l’analyse initiale des OR, avec les mêmes incohérences entre sites**. Tout en faisant disparaitre discrètement les OR < 1 de cette étude initiale.

**Une méthode sans contrôle du risque de 1ère espèce**

Concrètement, pour les 22 produits autres que les PCB, la nouvelle et très complexe méthode statistique employée donne finalement les mêmes résultats que l’analyse initiale basée sur les OR. Or nous avons vu que la distribution de ces OR ne différait pas de celle que l’on aurait obtenu avec 22 produits inoffensifs ! Le résultat obtenu avec la régression WQS est-il lui aussi une simple conséquence du risque de 1ère espèce ? Impossible de le savoir : dans la régression WQS, l’évaluation de chaque produit dépend de son coefficient de régression, et non de son OR. Pour les OR, nous avons vu qu’on calculait une probabilité critique et un intervalle de confiance à 95%, qui permettaient de s’assurer de la significativité des écarts à la normale, et donc de démasquer les alphas. **Or il n’existe pas de test statistique pour vérifier la significativité des écarts des coefficients du WQSI par rapport à la normale. Nous touchons là la 2ème vertu essentielle de cette méthode : il est impossible de distinguer la pêche aux alphas de résultats « réellement » significatifs !** Ce serait déjà gênant si les résultats de la régression WQS différaient radicalement de ceux des OR. Cela l’est encore plus, quand, après moult détours compliqués, l’analyse des coefficients WQSI nous ressort finalement les mêmes résultats que les OR !

**Si on fait le bilan de cette étude, on constate donc que :**

* **Les auteurs sont partis de données qui présentaient quelques résultats significatifs, mais dont la distribution laissait penser (pour les 22 produits non PCB) qu’il s’agissait d’un simple effet du risque de 1ère espèce.**
* **Ils ont réalisé une nouvelle analyse statistique, avec une méthode inappropriée, puisqu’elle ne pouvait normalement s’appliquer qu’à des produits ayant tous une liaison positive avec le LNH, ce qui n’était pas le cas de leur jeu de données.**
* **Après ce traitement inapproprié, ils obtiennent des résultats identiques à ceux de l’analyse initiale, mais par une méthode qui ne permet pas de vérifier le risque de 1ère espèce.**

Des chercheurs jeunes et naïfs pourraient penser qu’ils ont perdu leur temps, et que leurs résultats n’ont aucune chance d’être publiés. C’est oublier que les revues scientifiques ont des intérêts communs avec eux. Les chercheurs ont désespérément besoin de résultats significatifs (d’où la tentation de la pêche aux alphas). Les revues scientifiques cherchent tout aussi désespérément des articles susceptibles de faire monter leur impact factor. De ce point de vue, les pesticides sont clairement un meilleur sujet que les PCB, dont le sort est réglé depuis longtemps. Avec un minimum d’intelligence, il est donc possible de satisfaire tout le monde. Pour cela, quelques précautions simples ont suffi pour les auteurs :

* Employer une méthode suffisamment complexe et nouvelle pour décourager la curiosité de la majorité des lecteurs
* Mettre en annexe (plutôt que dans le corps du texte) les résultats embarrassants
* Glisser eux-mêmes dans le texte les objections que l’on pourrait faire à leur méthode, de façon suffisamment elliptique pour laisser croire qu’elles sont mineures.

Ce 3ème point est essentiel : tout le succès de l’opération repose sur le doigté avec lequel les auteurs réussissent à expliquer qu’ils ont employé une méthode statistique normalement inapplicable aux données étudiées, tout en restant assez elliptiques pour que l’importance de cette objection ne soit pas trop claire. Une fois ce postulat accepté par la revue, il n’y a plus aucun risque : on tient un bel article sur un sujet scientifique vendeur, sans risque de contre-attaque trop cinglante. Si par hasard un collègue grincheux envoie une lettre aux éditeurs pour rappeler que la méthode statistique est inadaptée, il suffira de répondre que cette objection était signalée par les auteurs, mais qu’il restait intéressant de vérifier que cette nouvelle méthode statistique ne donnait pas de résultats aberrants dans ce contexte. Avec un peu de chance, il est plus probable que d’autres épidémiologistes perçoivent plutôt le potentiel de cette méthode pour doper leurs résultats, et qu’elle devienne vite suffisamment employée pour devenir inattaquable.

**Quand les statistiques sont plus belles que les données, publiez les statistiques**

L’air de rien, cette publication est révélatrice d’une tendance de plus en plus fréquente : l’effacement des données derrière les statistiques. Les auteurs ne montrent dans l’article que les résultats de leur analyse statistique, les données de base étant reléguées dans les documents annexes. Il est vrai que d’autres chercheurs font encore plus fort : dans un article récent sur l’effet des néonicotinoïdes sur les abeilles[[4]](#footnote-4), les auteurs réussissent carrément à disserter sur 3 pages sur des surmortalités d’abeilles calculées par des redressements fort complexes…sans jamais publier les chiffres de mortalité bruts !

Autre tendance révélatrice : la complexité de la méthode statistique employée occulte toute réflexion sur la crédibilité des résultats obtenus. Passons sur le fait que la régression WQS n’aurait jamais dû être employée sur ce jeu de données. Même si on acceptait la validité de cette méthode, il devrait être clair que les coefficients calculés n’ont aucun sens sur le plan toxicologique, puisque tous les produits ont des résultats contradictoires selon les sites d’expérimentation. Cette question n’est même pas abordée dans la discussion des résultats. Du moment que les WQS Index ont une liaison significative ou hautement significative, peu importe que ces résultats puissent avoir ou non une signification dans le monde réel : c’est digne d’être publié.

Ces tendances ne s’observent pas dans toutes les revues scientifiques, mais certaines s’illustrent particulièrement dans ce domaine : rappelons que cet article vient d’Environmental Health Perspectives (EHP), revue particulièrement ouverte à la créativité statistique. Nous l’avons déjà croisée à propos des effets supposés des organophosphorés sur le système nerveux, avec deux articles qui faisaient tout le sel d’une récente étude du CNRS sur le coût caché des perturbateurs endocriniens[[5]](#footnote-5) :

* L’une[[6]](#footnote-6), qui démontrait que l’exposition prénatale aux organophosphorés faisait chuter le QI de 5 points…par rapport à une population témoin au QI de 107
* L’autre[[7]](#footnote-7), qui après moult redressements statistiques en fonction de l’origine ethnique, arrivait à la conclusion que ces mêmes organophosphorés faisaient chuter le QI des enfants noirs ou hispaniques…mais aussi augmenter celui des enfants blancs !
* De tels résultats feraient simplement sourire… s’ils ne sortaient que dans des revues obscures. Mais nous avons vu qu’EHP a un impact factor tout-à-fait honorable, qui leur assure un nombre considérable de citations dans d’autres articles par la suite… et leur reprise ultérieure dans des méta-analyses comme celle du CNRS déja citée, qui assureront définitivement le blanchiment de toutes leurs approximations statistiques.

1. Czarnota J, Gennings C, Colt JS, De Roos AJ, Cerhan JR, Severson RK, Hartge P, Ward MH, Wheeler DC. 2015. Analysis of environmental chemical mixtures and non-Hodgkin lymphoma risk in the NCI-SEER NHL Study. Environ Health Perspect 123:965–970

   <http://ehp.niehs.nih.gov/1408630/#tab1>

   Ne pas oublier de télécharger aussi le Supplemental Material :

   <http://ehp.niehs.nih.gov/1408630/#tab3>

   C’est là que figurent tous les éléments vraiment importants… et distrayants ! [↑](#footnote-ref-1)
2. Colt JS, Severson RK, Lubin J, Rothman N, Camann D, Davis S, et al. 2005. Organochlorines in carpet dust and non-Hodgkin lymphoma. Epidemiology 16(4):516–525. [↑](#footnote-ref-2)
3. Carrico C, Gennings C, Wheeler DC, Factor-Litvak P. 2014. Characterization of a weighted quantile sum regression for highly correlated data in a risk analysis setting. J Agric Biol Environ Stat [↑](#footnote-ref-3)
4. <http://rspb.royalsocietypublishing.org/content/282/1819/20152110> [↑](#footnote-ref-4)
5. <http://press.endocrine.org/doi/full/10.1210/jc.2014-4323> [↑](#footnote-ref-5)
6. Bouchard MF, Chevrier J, Harley KG, et al. (2011) : Prenatal exposure to organophosphate pesticides and IQ in 7-year-old children. EnvironHealth Perspect. (119) 1189–1195 [↑](#footnote-ref-6)
7. Engel SM, Wetmur J, Chen J, et al., 2011 : Prenatal exposure to organophosphates, paraoxonase 1, and cognitive development in childhood.Environ Health Perspect. (119) 1182–1188 [↑](#footnote-ref-7)