

Le dossier – Comment le HDL n'est plus devenu...

Comment des données maintenant remises en cause ont pu être acceptées comme fiables ?

RÉSUMÉ : Comment une hypothèse médicale aujourd'hui reconnue erronée a-t-elle pu être acceptée comme valide pendant plusieurs décennies ? Cet article analyse deux pistes.

Une première piste est socio-psychologique : c'est le biais de confirmation d'hypothèse qui conduit à croire, préalablement à la vérification de la validité des faits observés, qu'une hypothèse est valide parce que certaines données peuvent la soutenir. Ce biais de raisonnement annihile l'analyse critique de la validité de ces faits, analyse qui aurait pu ou dû faire reconnaître la faiblesse de leur niveau de preuve.

Dès lors que l'on est convaincu de la validité d'une hypothèse, la communication autour de faits pouvant la soutenir, même s'ils sont de faible niveau de preuve, a été conduite davantage pour faire "passer un message", c'est-à-dire supporter l'hypothèse, que pour inciter à en analyser la faiblesse. La seconde piste est l'analyse des critères montrant *a posteriori* en quoi les données initiales ayant soutenu l'hypothèse étaient faibles : études de faible puissance utilisant des tests statistiques ayant des limites.

F. DIÉVART
Clinique Villette, DUNKERQUE.

Entre la publication de l'étude princeps de Framingham en 1977 et la première méta-analyse publiée en 2009 remettant pour partie en cause la relation entre HDL-c et risque coronaire, les données épidémiologiques présentées aux médecins dans les congrès et celles publiées dans des revues à comité de lecture ont été admises comme valides. Comme il peut maintenant être reconnu qu'elles ne l'étaient pas, une question émerge : quels sont les éléments qui ont fait que ces données ont été admises comme valides par les médecins ?

Certaines des données de la recherche en psychologie, en sociologie et en biostatistique peuvent fournir quelques réponses.

Du côté des spectateurs : la pensée PowerPoint

Comment la présentation de données qui, *a posteriori*, manquent de fiabilité a-t-elle pu induire le sentiment qu'il y a une relation linéaire et inverse entre le HDL-c et le risque coronaire et que, de ce fait, il peut être utile d'augmenter le HDL-c avec un moyen pharmacologique ?

Il y a assurément plusieurs façons possibles d'envisager les réponses à cette question. Nous n'en envisagerons ici qu'une parmi d'autres. Celle-ci est fournie par deux ordres d'éléments : ce que certains appellent "la pensée PowerPoint" et ce que les psychologues ont analysé concernant la communication permettant d'exploiter le caractère

faillible et les immenses possibilités de mettre en défaut la réflexion.

Un certain type de pensée est qualifié de pensée “PowerPoint”. Qu’est-ce que c’est ? C’est la façon de faire passer un message et la façon dont il est reçu quand il est véhiculé par un jeu de diapositives brièvement présentées dans un congrès, une présentation ou une conférence. PowerPoint étant le logiciel le plus utilisé pour construire des diapositives sur un ordinateur, la pensée PowerPoint est donc une forme de néologisme désignant les modes de pensée induits par une présentation orale supportée par une projection de diapositives.

Quels en sont les ressorts ?

Pour être efficaces en termes de communication, les conférenciers/orateurs ont appris qu’un message doit être simple et très “visuel”. Il doit donc être résumé aussi graphiquement que possible à ses principaux éléments. En cela, le logiciel des ordinateurs permettant de construire des diapositives porte probablement bien son nom : PowerPoint. Dans une tentative de traduction en français, il pourrait ainsi se comprendre avec les sous-entendus suivants : le “point puissant” ou la “puissance du point”. En d’autres termes, “le point qui tue” ?

Comme l’écrivait Édouard Tétreau dans son livre “20 000 milliards de dollars” (Grasset, 2010) : “Finies, les notes compliquées avec des phrases longues et complètes. Terminées, les nuances subtiles et discursives apportées à un raisonnement ou une démonstration. PowerPoint, c’est : une idée et une illustration par page. Éventuellement, compléter par des sous-arguments, mais qui vont toujours dans le même sens. Celui du “point” que l’on veut mettre en avant. Avec PowerPoint, interdiction absolue de formuler des doutes, d’émettre des hypothèses contraires, d’interroger à charge et à décharge. Le PowerPoint ne fait pas bon ménage avec les sceptiques. Il est là pour servir la démonstration et l’action. Il ne

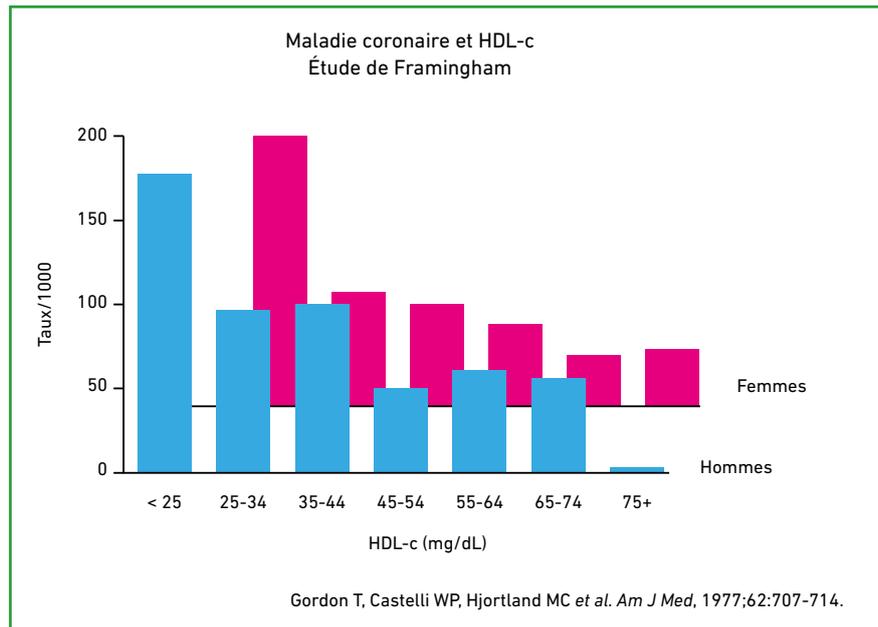


Fig. 1 : La diapositive reprenant les données de l’étude initiale de Framingham sur la relation entre HDL-c et risque d’événement coronaire.

saurait s’encombrer de détails superflus : ils n’ont pas de place sur ces belles diapositives projetées à l’écran, comme un film d’entertainment.”

En d’autres termes, pour que des présentations orales appuyées par des diapositives soient parlantes, elles seront efficaces si elles annulent les données chiffrées et qu’elles prennent une forme graphique claire. Ainsi, si l’on regarde la **figure 1**, et sa variante encore plus simplifiée de la **figure 2**, qui sont des diapositives utilisées dans de très nombreuses présentations de congrès et issues de l’étude de Framingham, on imagine instantanément la relation entre HDL-c et risque coronaire. Cette relation est graphique, imagée ; elle parle directement à ce que Daniel Kahneman appelle le système 1, ce mode de fonctionnement de l’esprit qui s’appuie sur des intuitions, une certaine aisance, mais sans approfondir ce qui lui est montré.

Lorsque l’on regarde la diapositive de la **figure 1** lors d’une conférence, visuellement, on relie d’instinct la grande barre de gauche à la petite barre de droite et

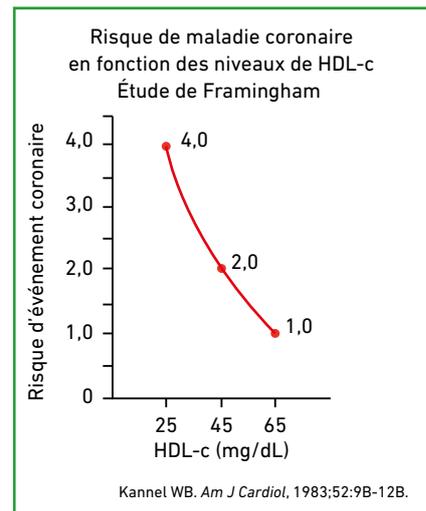


Fig. 2 : Une autre forme de diapositive reprenant les données de l’étude initiale de Framingham sur la relation entre HDL-c et risque d’événement coronaire.

on s’imprègne d’une droite virtuelle montrant bien que plus le HDL-c est bas, plus le risque coronaire est élevé et que ce risque décroît proportionnellement à l’augmentation du HDL. Et, dans la **figure 2**, la barre virtuelle est directement représentée, rendant le message encore plus parlant.

Le dossier – Comment le HDL n'est plus devenu...

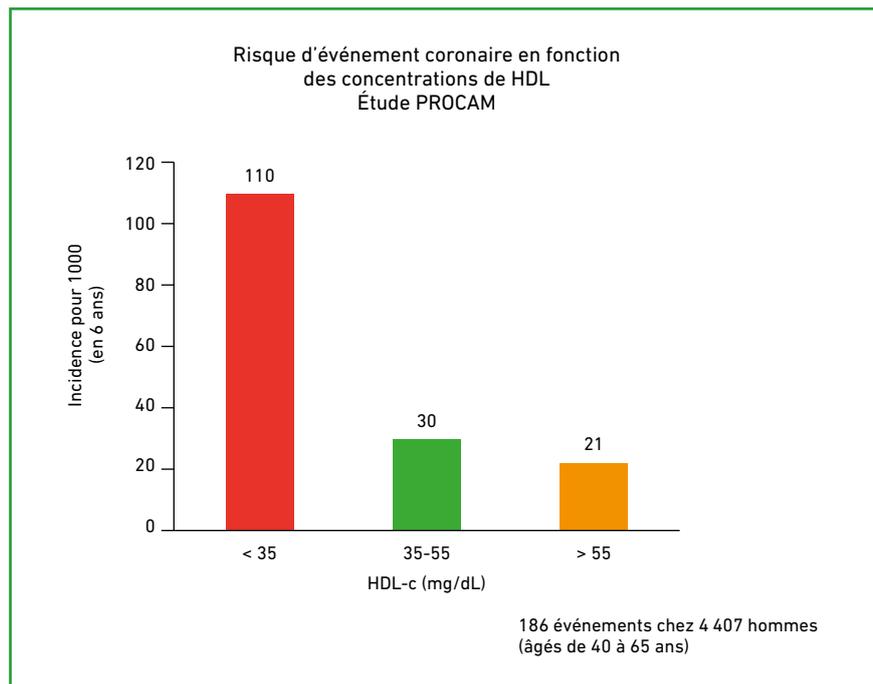


Fig. 3 : Diapositive reprenant les données de l'étude PROCAM concernant la relation entre HDL-c et risque d'événement coronaire.

Quand, de plus, une autre démonstration prend la même allure, comme dans l'étude PROCAM sur la **figure 3**, instantanément, l'esprit fait le parallèle avec les données déjà vues de la **figure 1**. Et à force de présentations, il devient familier avec cette représentation. Or, les psychologues nous enseignent que l'esprit a de grandes difficultés à faire la distinction entre familiarité et vérité. La donnée devenue familière prend donc l'allure d'une vérité, et ce d'autant plus qu'elle est présentée par des personnes ayant le statut d'experts. C'est comme cela que se crée la "pensée PowerPoint" qui, d'une belle image, tend à faire une vérité sans que les éléments spécifiques permettant de démontrer qu'il s'agit d'un fait aient besoin d'être présentés : une image et une parole d'autorité sont là pour garantir la validité du message.

Du côté des experts : le pourquoi d'éventuelles erreurs

Reprenons maintenant à nouveau les éléments de la **figure 1**. Comme il a été

dit, cette figure a été utilisée dans de nombreuses présentations faites dans de nombreuses conférences, au point qu'il est très facile de la trouver sur Internet et de l'utiliser pour une présentation personnelle. En l'utilisant, on se l'approprie plus encore qu'en la regardant d'ailleurs. Comme on la commente favorablement, on en fait une vérité qui devient progressivement une conviction qu'il sera de ce fait d'autant plus difficile à remettre en cause.

Si cette diapositive/présentation de résultats a des aspects attrayants, qui permettent une communication simple et aisée et frappent l'esprit, elle fournit néanmoins deux éléments ayant les caractères scientifiques requis : elle fournit une valeur du risque coronaire en taux pour 1 000 et elle possède une référence bibliographique. Plus encore, cette référence montre qu'elle est issue d'une étude prestigieuse, l'étude de Framingham, un des fondements – pour ne pas dire un "mythe" – de la cardiologie moderne.

Alors, où est le problème ? En d'autres termes, pourquoi le message qu'elle délivre n'a en fait, et à bien y regarder, qu'une portée scientifique limitée au point que ce message sera finalement invalidé dans les 40 ans qui suivront son élaboration ?

1. Peu de patients, des moyennes et pas d'intervalle de confiance

Le premier élément troublant est que cette diapositive ne fournit aucun chiffre absolu et surtout aucune déviation standard et/ou intervalle de confiance : il n'est pas possible de savoir en la regardant si elle prend en compte les données d'un million de personnes ayant eu 300 000 événements coronaires ou seulement celles de 300 personnes ayant eu 30 événements coronaires. Ce qui, en termes de puissance, peut tout changer.

Ainsi, par exemple, imaginons que l'incidence des infarctus associée à un HDL-c inférieur à 0,35 g/L soit de 2 % et celle d'un HDL-c supérieur à 0,75 g/L soit aussi de 2 %. Si, dans la population incluse, le nombre de patients ayant un HDL-c inférieur à 0,35 g/L est de 50 et que, dans la période d'observation, il survient dans ce groupe 5 infarctus – ce qui est tout à fait compatible avec un taux réel attendu de 2 % et est ici lié au hasard de l'échantillonnage – en face de ma colonne "inférieur à 0,35 g/L" il est possible de mettre une barre d'incidence moyenne qui monte jusqu'à 10 %, soit un taux de 100 pour 1 000. Par le même effet hasard, supposons que le groupe de patients ayant des valeurs de HDL-c supérieures à 0,75 g/L soit aussi de 50 patients et que, dans la période d'observation, il ne survienne dans ce groupe aucun infarctus. En face de cette valeur, il est donc possible d'afficher une incidence moyenne de 0 % ou 0 pour 1 000. Au final, on construit donc une droite de régression allant de 10 % à 0 % qui donne l'illusion d'une vraie régression.

Ce qui manque à ce graphique, ce sont donc des données chiffrées : les numé-

teurs, c'est-à-dire le nombre de patients par groupe de HDL-c (ce qui permettrait de juger de la puissance de l'étude), le dénominateur (c'est-à-dire le nombre d'événements par groupe de HDL-c) aboutissant à la même remarque et, enfin, la déviation standard ou l'intervalle de confiance à 95 % de l'incidence observée pour chaque groupe de HDL-c. Avec ce dernier élément, il serait alors peut-être possible de constater que la valeur la plus basse de l'intervalle de confiance associée à la valeur la plus basse de HDL-c est relativement proche de la valeur la plus haute de l'intervalle de confiance de la valeur la plus haute de HDL-c et que donc il n'y a peut-être pas une différence aussi grande dans les incidences d'événements coronaires entre les patients ayant des valeurs différentes de HDL-c. On remarquera, dans l'article précédent de ce dossier, que dans les résultats graphiques des méta-analyses de 2007, de 2009 et dans ceux de l'étude CANHEART, les intervalles de confiance à 95 % des résultats pour chaque valeur de HDL-c sont présentés, ce qui n'est pas le cas dans les représentations graphiques des études de Framingham et PROCAM.

Pour revenir à Framingham, en recourant à la publication de 1977 dont est issue la diapositive, il est possible de connaître les chiffres qui ont un intérêt pour analyser la valeur de l'étude. Cela justifie un effort de recherche puis un travail critique afin d'en évaluer la valeur probante. À la lecture de l'article source, on comprend que l'étude n'a porté que sur 2 815 hommes et femmes âgés de 49 à 82 ans ayant eu, dans la période d'observation, 142 événements coronaires, ce qui est relativement peu. Surtout quand ces 142 événements sont répartis en 14 groupes de patients (7 tranches de HDL-c chez l'homme et 7 tranches chez la femme) et que les résultats affichés sur la diapositive et publiés dans l'article ne portent que sur les moyennes des incidences d'événements par groupe, sans aucune indication sur les intervalles de confiance. Là encore, un travail critique

(qui est aussi un effort) permet de calculer ces intervalles de confiance.

Le calcul de ces intervalles de confiance, en ne prenant que quelques valeurs de HDL-c chez l'homme – et déjà, sans calculer si les différences entre les groupes sont significatives – est cependant instructif. Ainsi, pour les valeurs de HDL-c comprises entre 0,65 et 0,74 g/L chez l'homme, l'incidence des événements est de 0,025 (soit 2,5 % ou 25 pour 1 000), la population prise en compte étant de 40 patients, le calcul de l'intervalle de confiance à 95 % donne une valeur de $\pm 0,04$. L'incidence des événements dans ce groupe de patients a donc 95 % de probabilité d'être proche de zéro mais peut aller jusqu'à 6,5 %. Pour un HDL-c compris entre 0,25 et 0,34 g/L chez l'homme, l'incidence des événements a été observée à 0,1 (soit 10 %), la population prise en compte étant de 170 patients, et le calcul de l'intervalle de confiance à 95 % donne une valeur de $\pm 0,045$. Donc, l'incidence des événements dans ce groupe de patients a 95 % de probabilité d'être comprise entre 6 et 14 %. On constate que l'intervalle de confiance des incidences d'événements des patients ayant un HDL-c bas chevauche légèrement celui des patients ayant un HDL-c élevé. Il ne peut donc être exclu que l'incidence des événements dans les deux groupes les plus extrêmes, si l'étude avait été conduite dans une très grande population, soit finalement proche, c'est-à-dire peu différente. On comprend ici que faire des projections à partir de faibles échantillons de population fournit des chiffres aléatoires.

Dans l'étude de Framingham, prenons maintenant le groupe d'hommes le plus nombreux : l'analyse de la publication montre qu'il s'agit des patients ayant un HDL-c compris entre 0,35 et 0,44 g/L. Dans ce groupe de 335 patients, l'incidence observée des événements a été de 0,104 (soit 10,4 %). L'intervalle de confiance à 95 % est ici beaucoup plus étroit puisque la taille de la population est plus importante : cet intervalle est

de $\pm 0,009$, l'incidence des événements coronaires dans ce groupe a donc 95 % de probabilité d'être comprise entre 9,1 et 10,9 %. On constate ici que l'incidence probable des événements coronaires observables chez les patients de ce groupe est comprise dans l'intervalle de confiance de l'incidence des événements coronaires des patients ayant un HDL-c compris entre 0,25 et 0,34 g/L et qu'il pourrait être possible, si la taille de l'échantillon était beaucoup plus importante, que l'incidence des événements du groupe ayant un HDL-c inférieur à 0,34 g/L ne soit pas différente de celle du groupe ayant un HDL-c compris entre 0,35 et 0,44 g/L. En revanche, l'incidence réelle des événements dans le groupe à HDL-c compris entre 0,65 et 0,74 g/L a une forte probabilité d'être différente de celle des patients ayant un HDL-c compris entre 0,35 et 0,44 g/L.

On entrevoit donc que la taille de l'échantillon et le nombre des événements observés permettent déjà d'entrevoir des tendances de ce que peut être la relation épidémiologique entre HDL-c et événements coronaires, mais ne permettent pas encore de dire que le résultat fourni est indiscutable et démontrent la relation inverse observée.

2. La loi des petits nombres

Lorsque l'on observe la **figure 1**, il est nécessaire d'avoir à l'esprit ce que l'on qualifie de "loi des petits nombres". Cette approche complète les données fournies au paragraphe précédent. Ainsi, en regardant la figure, l'impression visuelle qui se dégage est qu'il y a une forte incidence d'événements chez les patients ayant un HDL-c très bas et très peu d'événements chez ceux ayant un HDL-c très élevé, et qu'il y a un aspect tendant vers l'horizontale dans les valeurs intermédiaires de HDL-c.

Or, la répartition des valeurs de HDL-c dans la population est normale, ce qui, en d'autres termes, veut dire qu'elle prend l'aspect d'une courbe de Gauss. Qu'est-ce

Le dossier – Comment le HDL n’est plus devenu...

que cela veut dire ? Tout simplement qu’aux valeurs extrêmes de la distribution, le nombre de patients est très faible. Si leur incidence d’événements est faible, les incertitudes sur la fiabilité du résultat observé dans les valeurs extrêmes seront grandes, ce que doit traduire la représentation des intervalles de confiance. Plus encore, si la taille des populations de l’extrémité de la courbe de Gauss est faible, si l’incidence des événements y est faible, la modalité de déroulement d’une étude prospective tend à rendre encore plus aléatoire le résultat qui sera observé.

Ainsi, dans ce type d’étude, les patients ont été suivis à partir du moment où un dosage du HDL-c avait été fait, donc inclus progressivement dans le temps. Inversement, l’analyse a été faite en arrêtant la surveillance des données à un temps t précis et unique pour tous les patients. Les patients ayant les valeurs extrêmes de HDL-c dans la distribution étant plus rares que ceux ayant des valeurs moyennes dans cette distribution gaussienne, il y a une plus grande probabilité pour que les patients ayant des valeurs extrêmes aient eu un suivi plus court que ceux dont les valeurs étaient proches de la moyenne. De ce fait, les patients ayant des valeurs extrêmes ont toute probabilité d’avoir été suivis moins longtemps en moyenne que ceux ayant des valeurs autour de la moyenne, ce qui a diminué la probabilité d’observer avec une fiabilité acceptable un taux d’événements valide chez les patients ayant des valeurs extrêmes, créant un biais.

Dès lors, la probabilité d’un résultat dû au hasard dans la construction des points extrêmes devient très grande. C’est ce qui est montré par le calcul des intervalles de confiance à partir du moment où il est possible de disposer du nombre de patients par échantillon et du nombre d’événements dans ces échantillons.

Pourtant, dans cette branche de l’étude de Framingham, parue en 1977, il a été conclu qu’il y a une relation inverse “démontrée” entre le HDL-c et le risque

coronaire alors qu’aucune valeur de pente éventuelle de régression n’est fournie. À ce stade, il faut donc analyser la méthode statistique utilisée dans cette partie de l’étude de Framingham. Si celle-ci a pour objectif de comparer des moyennes entre groupes, le choix d’un test statistique dépend de plusieurs variables : répartition gaussienne ou non d’une variable, tailles similaires ou non des groupes comparés, capacité discriminante du test... Dans l’étude de 1977 de Framingham, le test statistique utilisé (c’est-à-dire choisi par les investigateurs) a bien reposé sur une comparaison des moyennes d’événements par groupe.

Question : le test choisi prend-t-il en compte les écarts-types qui sont très importants aux extrémités de la distribution ? Surtout, le test statistique utilisé dans cette étude, créé en 1955 par DB Duncan, s’il est bien un test permettant des comparaisons de multiples moyennes de données, est un test critiqué. En effet, s’il a été créé pour diminuer la possibilité d’obtenir des faux négatifs (ou erreurs de type II), cela s’est fait au prix d’une plus grande probabilité d’obtenir... des faux positifs (ou erreurs de type I). L’utilisation de ce test augmente donc la probabilité que le résultat observé soit... faussement positif. Cela peut donc conduire à affirmer une corrélation alors qu’il n’y en a pas.

Que retenir d’une analyse critique a posteriori des résultats de l’étude de Framingham ?

Une première leçon est qu’il est nécessaire d’avoir recours aux données de la publication (s’entend par là, les données chiffrées), puis à leur analyse avec un regard critique. Cela nécessite du temps et aussi de savoir remettre en cause plusieurs éléments : le mythe Framingham, le fait qu’un article puisse être publié dans une revue de référence mais ne pas avoir tous les critères de fiabilité utiles, le langage véhiculé dans moult confé-

rences, recommandations voire parfois utilisés par soi-même...

Ce regard critique nous apprend que :

- le nombre absolu de patients par tranche de HDL-c et le nombre absolu d’événements survenus dans l’étude de référence de Framingham a été relativement faible ;
- ce qui a été pris en compte pour juger de la relation entre HDL-c et risque coronaire a été uniquement une moyenne des événements observés par groupe de HDL-c, sans intervalle de confiance, sans, *a priori*, d’ajustement ;
- la comparaison (l’utilisation du test statistique) a été faite à partir des moyennes alors que la variation possible de celles-ci peut être importante aux extrémités de la distribution ;
- l’analyse a été faite avec un test dont les limites sont maintenant reconnues.

On comprendra donc que l’utilisation de méthodes potentiellement plus adaptées et que le recours à des échantillons beaucoup plus importants pourra contribuer à relativiser les données issues de cette publication, pour ne pas dire à les rendre fausses, c’est-à-dire à les falsifier, pour reprendre une terminologie conçue par Karl Popper.

Alors, comment comprendre que l’on ait pu dire que “dans la publication de 1977 de l’étude de Framingham, il y a une analyse statistique qui démontre, de façon indiscutable, que le risque coronaire est inversement proportionnel à la valeur du HDL-c” ? En fait, dans cette assertion, le terme “démontre” doit être considéré comme relatif pour plusieurs raisons. Quant au terme “indiscutable”, il s’agit d’un jugement de valeur, donc encore plus relatif.

Le terme “démontre” est relatif et c’est bien le principe des statistiques : une analyse statistique ne démontre pas : elle indique que la probabilité qu’un résultat soit lié au hasard est plus ou moins grande. En d’autres termes, elle réduit l’incertitude mais ne fournit pas de certitude.

Ensuite, une analyse statistique repose sur un choix, celui notamment d'un test, test qui résulte lui-même de multiples choix.

Une deuxième leçon est la valeur encore plus relative qu'à une présentation orale sous forme de diapositives. On aura approché, par cet exemple, ce que peut être la "pensée PowerPoint".

Du côté de l'émetteur du message, en voulant que ce message porte, par souci de simplicité, il retire des informations essentielles pour ne laisser qu'un graphique démonstratif. Le paradoxe est que les informations retirées sont essentielles en ce sens qu'elles pourraient invalider la valeur du message porté par le graphique.

Du côté du récepteur du message, l'information visuelle est forte, simple à comprendre, graphique, elle porte sur le mode du fonctionnement du cerveau (le système 1 de Daniel Kahneman) qui préfère fonctionner sans effort.

Et ainsi, progressivement, il est tout à fait compréhensible, *a posteriori*, que dans l'esprit du public médical ait pu s'instiller la notion que le HDL-c est un facteur de risque, qu'il peut être utile – voire qu'il doit être utile – de l'augmenter avec un traitement pharmacologique. Et, le jour où des avis d'experts et/ou des recommandations indiquent qu'il ne faut pas proposer de traitement ayant comme objectif d'augmenter le HDL-c, y compris chez les diabétiques, il peut y avoir une relative incompréhension de la part du public médical, incompréhension qui sera définitive chez certains, transitoire et de durée variable chez d'autres.

■ En synthèse : Framingham 1977 remise en cause

En utilisant ce que certains appellent par ironie un "rétroscopomètre", il est possible de comprendre ce qui peut être considéré comme les limites de

l'étude de Framingham. Ainsi, à la lumière de données nouvelles et de plus forte puissance, il est possible de reconnaître que les données issues de l'étude de Framingham concernant le HDL-c avaient finalement un faible niveau de preuve. Cependant, elles ont longtemps été les seules disponibles. Si les données venant ensuite, avec un tout aussi faible niveau de preuve, ont semblé confirmer les données initiales de l'étude de Framingham, toutes les données dans ce domaine, notamment les données discordantes, ont-elles été publiées ? Et si oui, ces données discordantes ont-elles été mises en avant ?

L'histoire du HDL-c illustre une évolution de la médecine : celle où l'arrivée de données gigantesques, les *big data*, vont commencer soit à confirmer avec un niveau de preuve plus élevé des données antérieures, soit, dans plusieurs cas, à remettre en cause les données antérieures.