

[Adjustment for energy intake in the assessment of
dietary risk factors]

Anne Thiébaud, Emmanuelle Kesse, Catherine Com-Nougué, Françoise
Clavel-Chapelon, Jacques Bénichou

► To cite this version:

Anne Thiébaud, Emmanuelle Kesse, Catherine Com-Nougué, Françoise Clavel-Chapelon, Jacques Bénichou. [Adjustment for energy intake in the assessment of dietary risk factors]. *Epidemiology and Public Health / Revue d'Epidémiologie et de Santé Publique*, Elsevier Masson, 2004, 52 (6), pp.539-57. inserm-00172736

HAL Id: inserm-00172736

<https://www.hal.inserm.fr/inserm-00172736>

Submitted on 18 Sep 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Ajustement sur l'apport énergétique dans l'évaluation des facteurs de risque alimentaires

Adjustment for energy intake in the assessment of dietary risk factors

A. Thiébaud^{1,2}, E. Kesse¹, C. Com-Nougué^{1,3}, F. Clavel-Chapelon¹, J. Bénéchou²

¹INSERM, E3N-EPIC, Institut Gustave-Roussy, Villejuif.

²Unité de Biostatistique, CHU et Faculté de Médecine-Pharmacie de Rouen, 1, rue de Germont, 76031 Rouen Cedex.
Email biostatistique@chu-rouen.fr

³Université Paris-VI.

Résumé Dans les études épidémiologiques examinant l'association entre une mesure d'état de santé et des facteurs nutritionnels se pose le problème de l'ajustement sur l'apport énergétique. En effet, comme la plupart des apports en nutriments sont fortement corrélés à l'apport énergétique qui peut lui-même être associé au risque de maladie, il est indispensable d'ajuster sur ce facteur pour déterminer l'effet d'un nutriment. Pour éviter les problèmes d'estimation et d'interprétation liés à l'utilisation de la méthode standard qui ajuste directement sur l'apport énergétique, plusieurs autres méthodes ont été proposées : la méthode des densités qui utilise le quotient de l'apport en nutriment sur l'apport énergétique total, la méthode des résidus qui repose sur les résidus de la régression de l'apport en nutriment sur l'apport énergétique total et la méthode de partition qui ajuste l'énergie apportée par le nutriment sur l'énergie apportée par les autres sources. Ces méthodes conduisent à l'estimation d'effets différents mais aucune ne mesure directement l'effet spécifique du nutriment d'intérêt. Les effets estimés sont des combinaisons d'effets spécifiques des nutriments et d'un effet générique de l'énergie qu'ils apportent et reflètent l'addition ou la substitution d'un nutriment à un autre. Cet article propose une revue de ces méthodes et de leur interprétation, ainsi que leur application à l'analyse de la relation entre apport protéique et adénome colorectal dans la cohorte E3N-EPIC. Cet exemple illustre l'avantage d'utiliser plusieurs de ces méthodes plutôt qu'une seule afin d'obtenir une interprétation qui puisse guider les recommandations nutritionnelles.

Mots-clés : Épidémiologie nutritionnelle, Nutriment, Colinéarité, Résidu, Densité, Addition, Substitution, Adénome colorectal, Apport protéique

Abstract *Epidemiologic studies assessing the association between health status and nutritional factors raise the issue of adjusting for energy intake. Indeed, as most nutrients are highly correlated with energy intake which can itself be associated with disease risk, energy intake needs to be adjusted for upon assessing the effect of a specific nutrient. To avoid problems of estimation and interpretation incurred by the use of the standard method which rests on directly adjusting for energy intake, several other methods have been suggested. Namely, the density method uses the ratio of nutrient intake over total energy intake, the residual method relies on the residuals from the regression of nutrient intake on total energy intake, and the partition method fits energy from the nutrient and energy from other sources. These methods yield estimates of different effects but do not allow direct estimation of specific nutrient effects. Estimated effects combine specific and generic energy effects of nutrients and reflect effects of adding or substituting one nutrient for another. We review and apply these methods to the assessment of the association between protein intake and colorectal adenoma occurrence in the E3N-EPIC cohort. This example illustrates how considering findings from all of these methods rather than one single method can lead to a more in-depth understanding of such associations and provide useful guidance for nutritional recommendations.*

Keywords: *Nutritional epidemiology, Nutrient, Colinearity, Residual, Density, Addition, Substitution, Colorectal adenoma, Protein intake*

Introduction

L'alimentation est une composante de notre comportement, par là-même modifiable. Son rôle dans l'apparition de nombreuses pathologies, telles que les maladies cardio-vasculaires et les cancers, a donc été souvent étudié à des fins de prévention primaire. Dans les études épidémiologiques recherchant les facteurs de risque alimentaires d'une pathologie donnée, différentes dimensions de l'alimentation peuvent être prises en considération : la consommation d'aliments spécifiques, la consommation de groupes d'aliments, des profils alimentaires associant plusieurs caractéristiques de l'alimentation (par exemple un profil végétarien) ou encore l'apport de nutriments spécifiques provenant de plusieurs aliments.

La quantité totale d'aliments ingérés, qui peut être estimée par l'apport énergétique (ou calorique, ces deux termes étant employés ici de manière équivalente) total, peut aussi être impliquée comme facteur de risque de maladie. D'un individu à l'autre, cet apport est principalement modulé par trois facteurs : la corpulence, l'efficacité métabolique (la capacité à utiliser plus ou moins d'énergie pour assimiler les nutriments et maintenir le corps à température constante) et l'activité physique^{1,2}. Par conséquent, l'apport calorique total peut être lié au risque de maladie soit directement soit indirectement comme résultat de l'association avec un ou plusieurs de ces trois facteurs¹. De plus, les apports en nutriments sont pour la plupart positivement corrélés à l'apport calorique total. C'est le cas des macronutriments (lipides, glucides, protéines, éthanol) et de leurs composantes qui contribuent directement à l'apport calorique. C'est aussi le cas de la plupart des nutriments non énergétiques (vitamines et minéraux) car les individus ayant un apport calorique élevé ont tendance à consommer de plus grandes quantités de tous les aliments et donc de tous les nutriments³. Donc, si l'apport calorique total est lié au risque de maladie, pratiquement tous les apports nutritionnels tendent aussi à lui être associés, dans la même direction que l'apport calorique total. En conséquence, avant d'attribuer un effet propre à un nutriment donné (le nutriment *per se* comme facteur de risque de la pathologie considérée), il est indispensable de démontrer que l'association de ce nutriment avec le risque de pathologie est indépendante de l'apport calorique². En pratique, dans l'analyse de données d'épidémiologie nutritionnelle, cela signifie qu'il est indispensable d'ajuster sur ce facteur.

Les premières études d'association entre pathologies chroniques et apports en nutriments spécifiques exprimaient leurs résultats soit en gramme d'apport de nutriment soit en densités de nutriment (une unité d'apport en nutriment rapportée à une unité d'apport calorique total). Les densités nutritionnelles étaient utilisées pour corriger les différences d'apport nutritionnel dues à des différences d'apport calorique entre individus. À partir de la seconde moitié des années 1980, deux nouvelles formes d'expression sont apparues pour ajuster sur l'apport calorique : les résidus de nutriment⁴ et la partition de l'apport calorique total en ses composantes⁵. Ces méthodes ont fait l'objet d'un grand nombre de discussions méthodologiques dans la littérature épidémiologique et clinique^{1,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20} que nous synthétisons ici, après la présentation des différents modèles. Nous insistons sur l'interprétation de ces modèles qui nous semble trop souvent négligée dans les études épidémiologiques, notamment en ce qui concerne les types d'effets (substitution et addition) qu'ils permettent d'estimer. Nous abordons également le cas où les variables alimentaires sont introduites sous forme catégorielle plutôt que continue dans les modèles, ce qui induit des différences supplémentaires. Tous les modèles et leur interprétation sont illustrés à l'aide de données issues de la cohorte E3N-EPIC sur l'apport protéique et la survenue d'adénome colorectal, comme décrit ci-après.

Un exemple d'étude de facteurs de risque nutritionnels

L'étude de cohorte E3N (Étude Épidémiologique auprès de femmes de la Mutuelle Générale de l'Éducation Nationale) est une enquête prospective ayant pour objectif principal l'étude de la relation entre facteurs comportementaux et cancer chez la femme²¹. Une partie de la cohorte, composée de 74 524 femmes ayant complété entre 1993 et 1997 un questionnaire alimentaire validé²², constitue la composante française de l'étude européenne EPIC (*European Prospective Investigation into Cancer and Nutrition*)²³. Le questionnaire, de type histoire alimentaire, permettait d'évaluer les habitudes alimentaires au cours de l'année précédente pour 208 aliments, recettes et boissons dont on demandait la fréquence de consommation et la quantité consommée avec l'aide d'un livret de photographies également validé²⁴. Les fréquences et les quantités ont été transformées pour obtenir les consommations journalières d'aliments qui ont ensuite été converties en apports nutritionnels à l'aide d'une table de composition *ad hoc*, dérivée de la table REGAL²⁵. À titre d'illustration, nous proposons d'étudier la relation entre survenue d'adénome colorectal et apport protéique sans chercher néanmoins à contrôler les autres facteurs de risque non alimentaires éventuels. Pour cette étude, le suivi a été défini de la date de remplissage du questionnaire alimentaire jusqu'à la date de point fixée au 31 décembre 1997 et réalisé au moyen de deux questionnaires auto-administrés, en 1994-1995

et 1997. Les cas incidents sont les sujets ayant déclaré des polypes intestinaux au cours du suivi et dont le compte-rendu anatomo-pathologique a confirmé le caractère adénomateux et l'absence de dysplasie sévère. Pour sélectionner les femmes indemnes d'adénome tout au long du suivi (non-cas), la population a été restreinte aux femmes ayant déclaré dans un questionnaire ultérieur envoyé en juin 2000 avoir effectué une coloscopie après le 31 décembre 1997 et pour lesquelles il a pu être vérifié que cette coloscopie était normale, de même que les éventuelles coloscopies effectuées entre l'enquête alimentaire et cette date. Au final, 517 cas incidents d'adénome colorectal et 4 695 sujets indemnes ont été retenus pour l'analyse. Sur l'ensemble de ces 5 212 femmes, l'apport calorique total moyen était de 2 155 kcal/j et l'apport protéique moyen de 90,2 g/j (soit 360,7 kcal/j), ce qui représentait 17 % de l'apport calorique total.

Principes des modèles nutritionnels

Supposons que l'on cherche à modéliser l'association de l'état de santé avec l'apport en un certain nutriment contribuant à l'apport calorique (macronutriment). L'apport en ce nutriment, N , peut être exprimé en kilocalories/jour (kcal/j) en multipliant la quantité brute, en grammes/jour (g/j), par sa valeur énergétique, en kcal/g (9 kcal/g pour les lipides, 3,75 kcal/g pour les glucides, 4 kcal/g pour les protéines et 7 kcal/g pour l'éthanol). Le cas où le nutriment ne contribue pas à l'apport calorique sera abordé dans la discussion. On note T l'apport calorique total.

Le modèle le plus simple associant l'état de santé à l'apport en nutriment d'intérêt peut s'écrire de la façon suivante : $g(D|N) = a_0 + a_1N$. (modèle 0)

L'interprétation de la fonction $g(\cdot)$ varie selon le contexte mais les principes gouvernant les modèles présentés sont généraux²⁶. Dans le cas où l'état de santé est mesuré par une variable continue D , par exemple le taux de cholestérol dans le sang, et que le modèle 0 correspond à un modèle de régression linéaire simple, $g(\cdot)$ désigne l'espérance mathématique conditionnelle de la variable D , $E(D|N)$. Souvent, l'état de santé est mesuré par une variable binaire « malade/non malade » dont on modélise la probabilité de survenue p à l'aide d'un modèle de régression logistique, par exemple dans une étude cas-témoins : $g(\cdot)$ est alors égale à $\text{logit } p$. L'état de santé peut aussi être mesuré par le délai de survenue d'un événement que l'on modélise à l'aide d'un modèle de Cox ou de Poisson, par exemple dans une étude de cohorte : $g(\cdot)$ vaut alors $\ln h(\cdot)$ où h désigne la fonction de risque instantané d'événement. La particularité dans ce dernier cas est que a_0 est une fonction du temps (continue non spécifiée pour le modèle de Cox, constante par intervalle pour le modèle de Poisson). En référence aux deux dernières situations qui sont les plus fréquemment rencontrées dans les études épidémiologiques, nous parlerons dans la suite de risque de maladie plutôt que d'état de santé en général.

Quelle que soit la formulation, le modèle 0 ignore l'apport calorique et ne permet donc aucun contrôle de l'effet de confusion de ce facteur, ce qui est pourtant indispensable. Plusieurs modèles ont été proposés afin de contrôler l'effet de l'apport calorique.

Méthode standard (« standard method »)

Cette méthode repose sur le modèle, classique pour le contrôle des facteurs de confusion en épidémiologie, dans lequel est introduit un terme pour l'apport calorique total en plus du terme de l'apport brut en nutriment, selon la forme : $g(D|N,T) = a_0 + a_1N + a_2T$. (modèle 1)

Néanmoins, on constate fréquemment une forte corrélation entre N et T ³. Ainsi, dans notre exemple, la corrélation (coefficient de Pearson) entre l'apport protéique et l'apport calorique total était de 0,83. Cette corrélation pose des problèmes pratiques du fait de difficultés de convergence des algorithmes utilisés dans les logiciels statistiques pour estimer les paramètres. Elle pose également un problème conceptuel car elle rend délicate l'interprétation des paramètres a_1 et a_2 du modèle, d'autant que N est contenu dans T . Trois autres méthodes d'ajustement sur l'apport calorique ont de ce fait été proposées afin de limiter ou supprimer ce problème de colinéarité.

Méthode des densités (« nutrient density method »)

Pour un nutriment contribuant à l'apport calorique, la densité nutritionnelle (ou fraction énergétique) représente la part d'énergie apportée par ce nutriment. Elle est calculée en divisant la quantité de nutriment par l'apport calorique total. Du fait de cette division, on s'attend à une corrélation moindre entre N/T et T qu'entre N et T .

Les densités nutritionnelles peuvent être utilisées sans aucun ajustement supplémentaire sur l'apport calorique total, selon la forme : $g(D|N,T) = b_0 + b_1(N/T)$. (modèle 2a)

Malgré la faible corrélation entre N/T et T , l'interprétation du paramètre de N/T dans ce modèle n'est pas indépendante de T . En effet, un accroissement de N/T peut signifier soit une augmentation de N soit une diminution de T , ce qui rend l'interprétation du paramètre b_1 difficile. Il est donc préférable d'utiliser les densités nutritionnelles avec un ajustement sur l'apport calorique total, selon la forme : $g(D|N,T) = b_0 + b_1(N/T) + b_2T$. (modèle 2b)

Dans ce modèle, la variation de N/T correspond à la variation de N seul puisque T est maintenu constant. En outre, étant donné la corrélation plus faible entre N/T et T qu'entre N et T , les problèmes conceptuels et d'estimation rencontrés avec la méthode standard se trouvent réduits. Dans notre exemple, la corrélation de l'apport protéique avec l'apport calorique total est effectivement diminuée lorsque les apports bruts sont remplacés par les densités nutritionnelles puisque l'on passe alors à une corrélation de $-0,21$ (restant toutefois significativement non nulle).

Dans chacun des modèles 2a et 2b, l'interprétation du paramètre b_1 associé à la densité N/T dépend de l'unité choisie pour cette quantité. Le choix le plus fréquent est le pourcentage d'énergie totale, obtenu à partir du rapport des calories apportées par le nutriment sur les calories totales.

Méthode des résidus (« residual method »)

En alternative aux densités de nutriment, Willett et Stampfer⁴ ont proposé d'utiliser les résidus de la régression linéaire simple de la quantité brute de nutriment (variable expliquée) sur l'apport calorique total (variable explicative). À partir des estimations $\hat{\alpha}_0$ et $\hat{\alpha}_1$ des paramètres du modèle $E(N|T) = \alpha_0 + \alpha_1 T$, on calcule l'apport résiduel N' (« calorie-adjusted nutrient intake »). Pour un individu donné, N' est la différence entre son apport effectif en nutriment et l'apport en nutriment prédit par son apport calorique total, soit : $N' = N - \hat{E}(N|T) = N - (\hat{\alpha}_0 + \hat{\alpha}_1 T)$.

Une fois les résidus N' estimés pour chaque sujet, l'apport calorique total est introduit en plus de l'apport résiduel en nutriment dans le modèle associant le risque de maladie à l'apport nutritionnel, selon la forme :

$$g(D|N',T) = g(D|N,T) = c_0 + c_1 N' + c_2 T. \text{ (modèle 3)}$$

Par définition, les résidus N' sont de moyenne nulle et peuvent donc prendre des valeurs négatives. Ceci peut être gênant puisque l'on est habitué à des apports alimentaires positifs. Pour faciliter l'interprétation, on ajoute souvent au résidu une constante égale à la quantité de nutriment attendue pour un individu avec un apport calorique standard (par exemple la moyenne observée de ce nutriment sur l'échantillon analysé, soit $\bar{T} = \hat{\alpha}_0 + \hat{\alpha}_1 \bar{T}$). L'ajout d'une constante au résidu ne modifie que la valeur de la constante c_0 dans le modèle 3, pas celles des paramètres c_1 et c_2 associés à N' et T , respectivement. Il faut noter que des valeurs négatives peuvent néanmoins subsister, en particulier lorsque l'apport en nutriment considéré est modérément corrélé à l'apport calorique total (par exemple l'apport en éthanol)⁶.

Par construction, ces résidus translatés ont la même moyenne que la quantité brute de nutriment et ont une variance inférieure ou égale à celle de l'apport brut (Annexe 1). De plus, ils sont indépendants de l'apport calorique total, ce qui se vérifie dans l'exemple par une corrélation nulle entre apport calorique total et apport résiduel en protéines. Les problèmes conceptuels et d'estimation rencontrés avec la méthode standard sont ainsi entièrement évités par la méthode des résidus. Malgré l'indépendance entre N' et T , il est préférable de conserver le terme de l'apport calorique total dans le modèle, surtout en cas d'association entre le risque de maladie et l'apport calorique total, pour éviter un biais dans l'estimation de l'effet du nutriment d'intérêt^{10,27}. Dans le cas contraire où l'apport calorique total n'est pas lié au risque de maladie, ceci permet de faciliter l'interprétation et de préserver l'équivalence des modèles que nous évoquons plus loin.

Méthode de partition (« energy partition method »)

Cette méthode a été proposée par Howe *et al.*^{5,8} comme une alternative à la méthode des résidus de Willett et Stampfer⁴. On introduit dans le modèle l'énergie apportée par le nutriment considéré, N , et l'énergie apportée par les autres nutriments, $M = T - N$, soit :

$$g(D|N,M) = d_0 + d_1 N + d_2 M. \text{ (modèle 4)}$$

Ainsi, avec les protéines comme nutriment d'intérêt, M représente l'énergie apportée par les lipides, les glucides et l'éthanol. Dans ce modèle, N et M ne sont pas directement corrélés puisque l'un n'est pas une partie de l'autre et qu'ils pourraient contribuer indépendamment à l'apport calorique total. Dans la pratique, du fait que ces nutriments peuvent provenir des mêmes aliments, on s'attend cependant à une corrélation entre N et M mais moindre qu'entre N et T . Dans l'exemple, la corrélation de l'apport protéique brut avec l'apport calorique des autres macronutriments est effectivement diminuée par rapport à celle avec l'apport calorique total mais reste néanmoins élevée (0,76 contre 0,83). Les problèmes conceptuels et d'estimation rencontrés avec la méthode standard peuvent donc persister.

Interprétation des modèles

Effets spécifiques des nutriments et effet générique de l'énergie

Partant du postulat qu'une calorie provenant d'un macronutriment particulier peut avoir deux effets sur le risque de maladie, un effet *générique* en tant que source d'énergie et un effet *spécifique* de l'apport en ce nutriment, l'objectif d'une étude épidémiologique portant sur les déterminants alimentaires d'une pathologie devrait être d'isoler l'effet spécifique de l'effet générique de l'apport en macronutriment¹⁶. Considérons pour simplifier les apports en deux macronutriments seulement, les protéines (N) et les sources d'énergie non protéiques (M), en plus de l'apport calorique total ($T = N + M$). Il faudrait alors distinguer les trois effets séparés suivants sur le risque de maladie : un effet spécifique des protéines, un effet spécifique des macronutriments non protéiques et un effet générique de l'énergie, commun à toutes les calories.

Le modèle le plus simple pour séparer ces différents effets s'écrit : $g(D|N,M,T) = \beta_0 + \beta_N N + \beta_M M + \beta_T T$. (modèle 5)

Dans ce modèle, les paramètres des apports en nutriment d'intérêt N et en autres nutriments M , respectivement β_N et β_M , mesurent les effets spécifiques associés à l'augmentation d'une unité (1 kcal) d'apports en protéines et en macronutriments non protéiques, alors que le paramètre de l'apport calorique total, β_T , mesure l'effet générique associé à l'augmentation d'une unité (1 kcal) d'apport énergétique quelle qu'en soit l'origine. Ainsi, si l'on montrait que β_N ou β_M étaient statistiquement différents de 0, on aurait mis en évidence l'existence d'un effet spécifique qui ne peut pas être attribué à l'énergie, ce qui satisferait la condition définie par Willett pour attribuer à un macronutriment la causalité d'un effet².

Malheureusement, un tel modèle n'étant pas identifiable (du fait de l'égalité $T = N + M$), l'une des trois variables N , M ou T doit être exprimée en fonction des deux autres (par exemple $M = T - N$) qui, seules, restent dans le modèle. On retrouve alors les modèles 1, 3 et 4 présentés ci-dessus. Ainsi la méthode standard et la méthode des résidus ne considèrent pas l'effet des macronutriments non protéiques, tandis que la méthode de partition ignore l'effet de l'apport calorique total. Les paramètres estimés sont alors des combinaisons linéaires de β_N , β_M et β_T , et non ces paramètres eux-mêmes¹⁶ (Annexe 2). La méthode des densités n'entre pas dans ce cadre car elle est fondée sur une combinaison non linéaire de N et T ; elle ne cherche pas à mesurer un effet spécifique de N mais la part représentée par l'apport en nutriment N dans l'alimentation. De même que pour les trois autres méthodes d'ajustement sur l'apport calorique, les paramètres dans la méthode des densités peuvent néanmoins être interprétés en terme d'effets d'addition ou de substitution.

Effets d'addition et effets de substitution

Dans la méthode standard (modèle 1), le paramètre a_1 associé à l'apport en nutriment N représente l'effet de l'augmentation de 1 kcal de N en maintenant l'apport calorique total T constant. Ceci ne peut être réalisé que par la réduction simultanée de 1 kcal de l'apport calorique par les autres nutriments M . Par conséquent, a_1 représente l'effet de la *substitution* de 1 kcal de N à 1 kcal de M , à T constant¹⁴. C'est donc la différence entre les effets spécifiques de N et de M , à savoir $\beta_N - \beta_M$ selon les paramètres du modèle complet (modèle 5)¹⁶. Avec un raisonnement similaire, on montre que le paramètre a_2 associé à T représente l'effet de l'accroissement de 1 kcal de T à N constant, c'est-à-dire l'effet de l'*addition* de 1 kcal de M à T ¹⁴. En partant du modèle complet, un tel effet est défini comme la somme de l'effet spécifique et de l'effet calorique générique de M , c'est-à-dire $\beta_M + \beta_T$ ¹⁶.

Dans la méthode des résidus (modèle 3), c_1 représente l'effet de l'accroissement de 1 kcal d'apport résiduel $N' = N - (\hat{\alpha}_0 + \hat{\alpha}_1 T)$ lorsque T est constant, donc l'effet de l'accroissement de 1 kcal de N à T constant, ou encore l'effet de la *substitution* de 1 kcal de N à 1 kcal de M ¹⁴. C'est le même effet que $a_1 = \beta_N - \beta_M$ ¹⁶. Le coefficient c_2 mesure l'effet de l'accroissement de 1 kcal de T à N' constant. Pour que N' reste constant après accroissement de 1 kcal de T , il faut que N augmente d'une quantité égale à la pente de la régression de N sur T , c'est-à-dire $\hat{\alpha}_1$ kcal. Or, si T augmente de 1 kcal et N de $\hat{\alpha}_1$ kcal, alors M augmente aussi de $(1 - \hat{\alpha}_1)$ kcal.

Le coefficient c_2 décrit donc l'*addition simultanée* à l'apport calorique total de calories de N et de calories de M en quantités respectives $\hat{\alpha}_1$ et $(1 - \hat{\alpha}_1)$ ¹⁴. Il s'agit donc de la somme pondérée des trois effets, β_N , β_M et β_T , à savoir $\hat{\alpha}_1 \beta_N + (1 - \hat{\alpha}_1) \beta_M + \beta_T$ ¹⁶.

Dans la méthode de partition (modèle 4), d_1 représente l'effet de l'accroissement de 1 kcal de N à M constant, c'est-à-dire l'accroissement de T par l'ajout de 1 kcal de N . Le paramètre d_1 a donc une interprétation différente de a_1 et c_1 : d_1 mesure l'effet de l'*addition* de N à T ¹⁴. Avec la paramétrisation issue

du modèle complet, d_1 est égal à la somme de l'effet spécifique et de l'effet calorique générique de N , c'est-à-dire $\beta_N + \beta_T$ ¹⁶. Le coefficient d_2 représente l'effet de l'accroissement de 1 kcal de M quand N est constant, c'est-à-dire l'addition de M à T ¹⁴. Ainsi, d_2 et a_2 sont égaux et mesurent le même effet, $\beta_M + \beta_T$ ¹⁶.

L'interprétation des paramètres des modèles d'ajustement sur l'apport calorique fait donc apparaître des égalités entre paramètres des méthodes standard, des résidus et de partition, les trois modèles correspondants étant en réalité des paramétrisations différentes d'un même modèle⁷. L'ensemble des relations d'équivalence et les effets mesurés en fonction des paramètres des différents modèles est résumé dans le Tableau 1 et établi en Annexe 2 et Annexe 3.

TABLEAU I. — Effets mesurés par les modèles standard, des résidus et de partition de l'énergie : équivalence entre ces modèles et correspondance avec le modèle théorique complet pour des apports nutritionnels et caloriques sous forme continue.

Effet mesuré	Paramètres théoriques du modèle complet	Paramètres estimables		
		Modèle standard	Modèle des résidus	Modèle de partition
Substitution isocalorique de N à M (T constant)	$\beta_N - \beta_M$	a_1	c_1	$d_1 - d_2$
Addition de N à T (M constant)	$\beta_N + \beta_T$	$a_1 + a_2$	$(1 - \hat{\alpha}_1)c_1 + c_2$	d_1
Addition de M à T (N constant)	$\beta_M + \beta_T$	a_2	$c_2 - \hat{\alpha}_1 c_1$	d_2
Addition simultanée de N et M en quantités respectives $\hat{\alpha}_1$ et $(1 - \hat{\alpha}_1)$	$\hat{\alpha}_1 \beta_N + (1 - \hat{\alpha}_1) \beta_M + \beta_T$	$\hat{\alpha}_1 a_1 + a_2$	c_2	$\hat{\alpha}_1 d_1 + (1 - \hat{\alpha}_1) d_2$

N représente l'apport brut en un nutriment, M l'apport brut en sources d'énergie autres que N , $T = N + M$ l'apport calorique total.

β_N et β_M mesurent les effets spécifiques des apports en nutriment N et en autres nutriments M respectivement, β_T l'effet générique de l'apport calorique total.

a_1 , c_1 et d_1 sont les paramètres associés à l'apport en nutriment dans les modèles standard, des résidus et de partition, respectivement ; a_2 , c_2 sont les paramètres associés à l'apport calorique total dans les modèles standard et des résidus, respectivement, d_2 le paramètre associé à l'apport en autres nutriments dans le modèle de partition.

$\hat{\alpha}_1$ est la pente de la régression linéaire simple de l'apport en nutriment brut N sur l'apport calorique total T (méthode des résidus).

Enfin, dans la méthode des densités avec ajustement sur l'apport calorique total (modèle 2b), le paramètre b_1 associé à la fraction énergétique N/T représente l'effet de l'augmentation de l'apport en nutriment N de 1 % de l'apport calorique total T (si l'unité de densité choisie est le pourcentage d'énergie totale) quand T est constant, donc la diminution de l'apport en autres sources d'énergie M de 1 %. Ainsi, b_1 mesure aussi un effet de substitution de N à M , mais exprimé en pourcentage d'énergie au lieu d'apport énergétique en kcal. Le paramètre b_2 associé à l'apport calorique total T mesure l'effet de l'augmentation d'une unité de T à fraction énergétique N/T constante. Il faut donc augmenter l'apport en nutriment N d'une quantité égale à (N/T) kcal, ainsi que l'apport en autres macronutriments M d'une quantité égale à (M/T) kcal. L'interprétation de b_2 en termes d'addition simultanée de N et M ressemble donc à celle du coefficient c_2 dans la méthode des résidus mais les variations des apports respectifs en nutriments N et M ne sont pas les mêmes et, de plus, elles dépendent, dans la méthode des densités, de la composition initiale de l'alimentation pour chaque individu. Notons enfin qu'en l'absence d'ajustement sur l'apport calorique total (modèle 2a), la méthode des densités n'estime ni un effet d'addition ni un effet de substitution puisque la variation de N/T peut être due à la fois à la variation de N et à celle de T . Ceci rend le modèle 2a plus difficilement interprétable.

Au final, aucune des méthodes usuelles d'ajustement sur l'apport calorique ne permet d'estimer un effet spécifique d'un macronutriment particulier, ni l'effet générique de l'énergie, à moins d'hypothèses supplémentaires fortes, invérifiables à partir des données, telles que la nullité de l'un des trois effets considérés. Les effets mesurés en réalité peuvent cependant être interprétés en termes d'effets d'addition et de substitution.

Analyse et interprétation de l'exemple

Pour chacune des méthodes d'ajustement sur l'apport calorique, le risque de survenue d'adénome colorectal en fonction de l'apport protéique a été étudié à l'aide d'un modèle de Cox avec l'âge comme échelle de temps Tableau 2. Le test de l'hypothèse des risques proportionnels à partir des résidus de Schoenfeld²⁸ ne montre d'écart important pour aucune des covariables dans les différents modèles. Les risques relatifs sont donnés pour un incrément de 100 kcal/j pour les méthodes standard, des résidus et de partition, et pour un incrément de 5 % d'apport calorique total pour la méthode des densités, ce qui correspond à un apport absolu comparable étant donné l'apport calorique total moyen de 2 155 kcal/j observé dans la population d'étude. La pente de la régression de l'apport protéique brut sur l'apport calorique total servant au calcul de l'apport protéique résiduel est estimée à $\hat{\alpha} = 0,147$.

TABLEAU II. — Association entre l'apport en protéines (variables continues) et le risque d'adénome colorectal chez 5 212 femmes de la cohorte E3N-EPIC selon les différents modèles nutritionnels.

Modèle	Nutriment (incrément)	Paramètre	Écart-type	<i>p</i> , test de nullité	Risque relatif	Intervalle de confiance à 95 % du risque relatif
	Calories totales (100 kcal)	- 0,0016	0,0080	0,84	1,00	0,98-1,01
Standard						
(1)	Protéines, apport brut (100 kcal)	- 0,071	0,046	0,12	0,93	0,85-1,02
(2)	Protéines, apport brut (100 kcal)	- 0,206	0,083	0,013	0,81	0,69-0,96
	Calories totales (100 kcal)	+ 0,028	0,014	0,047	1,03	1,00-1,06
Densités						
(1)	Protéines, densité (5 %)	- 0,187	0,083	0,025	0,83	0,71-0,98
(2)	Protéines, densité (5 %)	- 0,197	0,085	0,020	0,82	0,70-0,97
	Calories totales (100 kcal)	- 0,005	0,008	0,51	1,00	0,98-1,01
Résidus						
(1)	Protéines, résidu (100 kcal)	- 0,206	0,083	0,013	0,81	0,69-0,96
(2)	Protéines, résidu (100 kcal)	- 0,206	0,083	0,013	0,81	0,69-0,96
	Calories totales (100 kcal)	- 0,002	0,008	0,82	1,00	0,98-1,01
Partition						
	Protéines, apport brut (100 kcal)	- 0,178	0,071	0,013	0,84	0,73-0,96
	Autres calories (100 kcal)	+ 0,028	0,014	0,047	1,03	1,00-1,06

(1) Modèle comprenant un terme pour l'apport en protéines, sous forme d'apport brut, de densité ou de résidu, mais pas de terme pour l'apport calorique.

(2) Modèle comprenant un terme pour l'apport en protéines, sous forme d'apport brut, de densité ou de résidu, et un terme pour l'apport calorique total.

Un premier modèle en fonction de l'apport calorique total seul semble indiquer que le risque de survenue d'adénome colorectal n'est pas lié à l'apport calorique total (test de nullité du paramètre $p = 0,84$). L'interprétation d'un tel modèle est néanmoins difficile car le coefficient associé à l'apport calorique total ne fournit pas une estimation de β_T seul Annexe 4. Parmi les déterminants de l'apport énergétique, la corpulence et l'activité physique présentent une association avec le risque d'adénome colorectal, celui-ci étant augmenté pour les femmes les plus corpulentes et pour les moins actives (résultats non montrés). L'estimation du risque associé à l'apport protéique pourrait être biaisée en l'absence d'ajustement sur l'apport calorique total puisque le risque relatif associé à un incrément de 100 kcal/j de l'apport protéique dans la méthode standard passe de 0,93 (intervalle de confiance à 95 % (IC 95 %) : 0,85-1,02) avant ajustement à 0,81 (IC 95 % : 0,69-0,96) après ajustement. En revanche, ce risque n'est pratiquement pas modifié dans la méthode des densités ni celle des résidus.

L'équivalence des méthodes standard, des résidus et de partition implique l'égalité des log-vraisemblances des modèles ajustés sur l'apport calorique (-3 346,2), l'égalité des paramètres $\hat{\alpha}_1$ et $\hat{\alpha}_1$ dans les modèles standard et des résidus (-0,206) et l'égalité des paramètres $\hat{\alpha}_2$ et $\hat{\alpha}_2$ dans les modèles standard et de partition (+0,028). Le paramètre $\hat{\alpha}_1$ dans le modèle de partition peut être obtenu par somme des paramètres du modèle standard ($\hat{\alpha}_1 + \hat{\alpha}_2 = -0,206 + 0,028 = -0,178$).

Ces résultats indiquent que la substitution de 100 kcal de protéines à 100 kcal de macronutriments non protéiques est associée à une réduction significative du risque d'adénome colorectal (RR = 0,81 pour les modèles standard et des résidus), de même que l'addition de 100 kcal de protéines à un apport constant en nutriments non protéiques (RR = 0,84 pour le modèle de partition). De plus, l'addition de 100 kcal de nutriments non protéiques à une quantité constante de protéines est associée à une augmentation significative du risque d'adénome colorectal (RR = 1,03 dans les modèles standard et de partition). Enfin, la méthode des densités indique une diminution de risque significative et du même ordre de grandeur que les méthodes standard et des résidus (RR = 0,84 pour une substitution de 5 % d'énergie par les protéines à 5 % d'énergie par les autres macronutriments).

Ces résultats ne permettent pas de conclure sur le sens des effets spécifiques des protéines et des autres macronutriments. De l'inégalité $\beta_N - \beta_M < 0$ découlent en effet trois scénarios possibles : un effet délétère des protéines et des autres macronutriments, moindre (voire nul) pour les protéines ($0 \leq \beta_N < \beta_M$), un effet protecteur (ou nul) des protéines et un effet délétère (ou nul) des autres macronutriments ($\beta_N \leq 0 \leq \beta_M$), un effet protecteur des protéines et des autres macronutriments, moindre (voire nul) pour les autres macronutriments ($\beta_N < \beta_M \leq 0$). Le modèle de partition amène l'inégalité suivante : $\beta_N < -\beta_T < \beta_M$, de sorte que la connaissance de l'effet générique de l'énergie pourrait permettre de conclure sur les effets spécifiques de N ou M . Pour cela, des résultats supplémentaires sur la relation du risque d'adénome avec les déterminants de l'apport calorique ou d'autres études sont nécessaires, les modèles étudiés ici étant insuffisants pour évaluer β_T .

Le chapitre suivant décrit comment une difficulté supplémentaire dans l'interprétation des résultats des méthodes d'ajustement sur l'apport calorique est introduite lors de l'utilisation de variables alimentaires catégorielles.

Effet de la catégorisation des variables alimentaires

L'équivalence des méthodes et l'interprétation de leurs paramètres ont été établies dans le cas où les variables alimentaires sont continues. Or, en pratique, ces données sont souvent analysées de manière catégorielle, à la fois pour faciliter l'interprétation des paramètres estimés, éviter l'hypothèse forte d'un effet de type linéaire et limiter l'influence des valeurs extrêmes. Les catégories peuvent être définies soit d'après des valeurs critiques de la variable d'exposition (par exemple : un verre, deux verres d'alcool par jour, etc.), soit de façon à obtenir des groupes comprenant le même nombre d'individus ou *quantiles*. Cependant, une fois les données discrétisées, l'équivalence des méthodes standard, des résidus et de partition n'est plus vérifiée. Ce point a été soulevé par Kushi *et al.*¹¹ au cours de l'analyse de la relation entre risque de cancer du sein et apport lipidique dans une étude de cohorte prospective américaine, brièvement étudié par Pike *et al.*¹² à partir d'un exemple fictif, puis examiné au moyen de simulations d'études cas-témoins par Brown *et al.*¹⁵. Ces auteurs ont ainsi constaté que, après catégorisation de l'apport en nutriment d'intérêt, d'une part, les méthodes des résidus et des densités produisent des résultats très similaires, d'autre part, la méthode standard produit des estimations de risque plus élevées que la méthode des résidus.

Exemple

Le Tableau 3 présente les risques relatifs obtenus après catégorisation des variables en quartiles. Un test de tendance a été réalisé en introduisant dans le modèle une variable index prenant les valeurs 1, 2, 3 ou 4 selon le quartile d'apport protéique auquel appartient le sujet : son degré de signification est celui associé au test bilatéral de nullité du paramètre de cette variable index. Dans ce tableau figurent les risques relatifs associés aux différents quartiles d'apport protéique sans ajustement sur l'apport calorique (1), avec ajustement sur l'apport calorique (total pour les méthodes standard, des densités et des résidus, des autres macronutriments pour la méthode de partition) en continu (2a) ou en quartiles (2b).

TABLEAU III. — Association entre l'apport en protéines (variables en quartiles) et le risque d'adénome colorectal chez 5 212 femmes de la cohorte E3N-EPIC selon les différents modèles nutritionnels.

Modèle	Nutriment	Risque relatif (intervalle de confiance à 95 %)				p, test de tendance ^a
		Q1	Q2	Q3	Q4	
	Calories totales	1 (réf.)	0,93 (0,73-1,18)	0,93 (0,73-1,18)	1,00 (0,79-1,28)	> 0,99
Standard						
	<i>Médiane^b (kcal/j)</i>	<i>251,6</i>	<i>324,4</i>	<i>383,7</i>	<i>472,3</i>	
(1)	Protéines, apport brut	1 (réf.)	0,94 (0,74-1,19)	0,95 (0,75-1,10)	0,78 (0,61-1,01)	0,078
(2a)	Protéines, apport brut	1 (réf.)	0,85 (0,66-1,10)	0,80 (0,60-1,07)	0,58 (0,40-0,86)	0,012
(2b)	Protéines, apport brut	1 (réf.)	0,90 (0,68-1,17)	0,82 (0,60-1,12)	0,58 (0,40-0,84)	0,006
	Calories totales	1 (réf.)	1,01 (0,77-1,33)	1,13 (0,83-1,55)	1,48 (1,03-2,13)	
Densités						
	<i>Médiane (%)</i>	<i>13,9</i>	<i>15,9</i>	<i>17,5</i>	<i>19,9</i>	
(1)	Protéines, densité	1 (réf.)	0,86 (0,68-1,08)	0,69 (0,54-0,88)	0,73 (0,57-0,93)	0,003
(2a)	Protéines, densité	1 (réf.)	0,86 (0,68-1,08)	0,69 (0,54-0,88)	0,72 (0,56-0,92)	0,002
(2b)	Protéines, densité	1 (réf.)	0,86 (0,68-1,09)	0,69 (0,54-0,89)	0,72 (0,57-0,92)	0,002
	Calories totales	1 (réf.)	0,91 (0,72-1,17)	0,90 (0,71-1,15)	0,95 (0,75-1,22)	
Résidus						
	<i>Médiane (kcal/j)</i>	<i>300,9</i>	<i>341,6</i>	<i>373,6</i>	<i>422,3</i>	
(1)	Protéines, résidu	1 (réf.)	0,86 (0,68-1,08)	0,74 (0,58-0,94)	0,70 (0,55-0,90)	0,002
(2a)	Protéines, résidu	1 (réf.)	0,86 (0,68-1,08)	0,74 (0,58-0,94)	0,70 (0,55-0,90)	0,002
(2b)	Protéines, résidu	1 (réf.)	0,86 (0,68-1,09)	0,74 (0,58-0,95)	0,70 (0,55-0,90)	0,002
	Calories totales	1 (réf.)	0,93 (0,73-1,19)	0,93 (0,73-1,18)	0,99 (0,78-1,27)	
Partition						
	<i>Médiane (kcal/j)</i>	<i>251,6</i>	<i>324,4</i>	<i>383,7</i>	<i>472,3</i>	
(2a)	Protéines, apport brut	1 (réf.)	0,86 (0,67-1,11)	0,82 (0,62-1,08)	0,61 (0,43-0,87)	0,010
(2b)	Protéines, apport brut	1 (réf.)	0,89 (0,68-1,15)	0,81 (0,61-1,08)	0,58 (0,41-0,82)	0,002
	Autres calories	1 (réf.)	1,00 (0,77-1,31)	1,14 (0,85-1,53)	1,53 (1,10-2,12)	

^a Test de nullité du paramètre associé à une variable prenant les valeurs 1, 2, 3, 4 selon le quartile d'apport protéique auquel appartient le sujet

^b Médiane d'apport protéique.

(1) = modèle sans ajustement sur l'apport calorique.

(2a) = modèle avec ajustement sur l'apport calorique total (ou des autres macronutriments pour la méthode de partition) en continu.

(2b) = modèle avec ajustement sur l'apport calorique total (ou des autres macronutriments pour la méthode de partition) en quartiles.

En accord avec le cas continu Tableau 2, les résultats dans le cas catégoriel indiquent l'absence d'association entre apport calorique total et risque d'adénome colorectal et des tendances significatives à la diminution du risque avec un apport protéique croissant, sauf dans la méthode standard avant ajustement sur l'apport calorique total. En revanche, contrairement au cas continu Tableau 2, les estimations de risque relatif ne sont plus équivalentes d'une méthode à l'autre : lorsque l'apport calorique est ajusté sous forme continue, la diminution de risque d'adénome colorectal associé à un apport protéique élevé est la plus marquée dans la méthode standard (RR = 0,58 pour le dernier quartile comparé au premier), suivie de près par la méthode de partition (RR = 0,61). Les méthodes des résidus et des densités se distinguent avec des réductions de risque, certes encore significatives, mais de moindre ampleur (RR = 0,70 et 0,72, respectivement). L'ajustement sur l'apport calorique total ou des autres sources d'énergie en quartiles plutôt qu'en continu ne modifie pas

l'estimation du risque relatif associé à l'apport protéique sauf de façon limitée dans la méthode de partition (RR = 0,58 contre 0,61 avec ajustement continu).

Notons que, en relatif, les incréments sont très différents d'une méthode à l'autre, avec une médiane d'apport multiplié par 1,9 dans le dernier quartile comparé au premier pour les méthodes standard et de partition, et par 1,4 pour les méthodes des résidus et des densités. Les effets mis en évidence pour l'apport protéique catégorisé sont en outre plus marqués qu'avec les variables continues. Par exemple, dans la méthode standard, le risque relatif contrastant le dernier quartile au premier est de 0,58 alors que, compte-tenu de la différence des médianes d'apport protéique brut entre quartiles extrêmes qui vaut $472,3 - 251,6 = 220,7$ kcal/j, il serait de 0,63 dans le modèle continu. De même, il est de 0,61 dans la méthode de partition au lieu de 0,68. Pour la méthode des densités, on observe un risque relatif de 0,72 alors qu'on attend un risque relatif de 0,80 pour un incrément de $19,9 - 13,9 = 6,0$ % d'énergie apportée par les protéines. Enfin, pour la méthode des résidus, le risque relatif est observé à 0,70 contre 0,78 attendu pour un incrément de $422,3 - 300,9 = 121,4$ kcal/j.

Discussion

Ajuster sur l'apport calorique s'avère nécessaire pour étudier l'association entre une pathologie et l'apport en un nutriment donné, afin d'éviter un biais de confusion potentiellement sévère. Nous avons passé en revue les quatre méthodes d'ajustement sur l'apport calorique décrites dans la littérature, ainsi que l'interprétation des paramètres estimés par chacun des modèles correspondants. Aucune de ces méthodes ne permet l'estimation d'un effet du nutriment *per se*, même si certains auteurs parlent de la mise en évidence d'un effet du nutriment *indépendamment de l'apport calorique*^{4,6}. La difficulté de l'ajustement sur l'apport calorique résulte de ce que l'apport en macronutriment contribue pour partie à l'apport calorique total. Cet aspect est souvent négligé dans les études épidémiologiques alors qu'il est déterminant dans l'interprétation des résultats des différents modèles d'ajustement. Nous reprenons en discussion les avantages et inconvénients des méthodes présentées et abordons des extensions possibles pour en enrichir l'application.

Intérêt pratique des effets d'addition et de substitution

L'interprétation des modèles peut guider le choix entre une méthode d'addition (partition) ou de substitution (standard, densités, résidus) selon la question scientifique posée. Dans une optique de supplémentation (ou de restriction) par exemple, mesurer un éventuel effet d'addition par la méthode de partition semble l'approche la plus pertinente. Si l'objectif de prévention primaire est la modification qualitative des habitudes alimentaires en vue d'une réduction du risque de maladie, une approche par substitution isocalorique (méthodes standard, des densités ou des résidus) se justifie. D'autres considérations peuvent également entrer en jeu, par exemple, pour un modèle de substitution isocalorique, la comparabilité avec des protocoles expérimentaux chez l'homme ou l'animal¹⁸, ou encore la faible variabilité de l'apport énergétique total à long terme (sauf changement de poids ou d'activité physique)¹⁸. Cependant, la prévalence grandissante de l'obésité qui pourrait résulter d'un apport énergétique croissant ou d'autres facteurs tels que la sédentarité^{19,29}, ainsi que la pratique de régimes restrictifs, tendent à remettre en cause ce dernier argument. Quelle que soit la méthode d'ajustement utilisée, il paraît utile d'ajuster aussi sur l'activité physique et la corpulence, si ces variables peuvent jouer un rôle sur le risque de maladie¹⁹.

Enfin, les méthodes de substitution ajustant sur l'apport calorique total seraient à privilégier pour limiter deux types de biais sur l'estimation du risque de maladie associé au nutriment étudié, l'un lié aux erreurs de mesure¹⁸, l'autre propre aux études rétrospectives (biais de mémoire ou d'anamnèse)^{13,18}. Les études de validation des apports nutritionnels (voir par exemple³⁰) montrent que des erreurs de mesure sont inhérentes à l'utilisation de questionnaires alimentaires, que ces erreurs sont corrélées entre elles et qu'elles induisent souvent une forte atténuation des relations observées entre risque de maladie et apport en nutriment, voire les font disparaître. L'ajustement sur l'apport calorique total peut corriger totalement ou en partie ce phénomène d'atténuation mais aussi l'aggraver^{19,31}. En cas d'absence d'association entre le risque de maladie et l'apport en nutriment d'intérêt, il peut même faire apparaître des associations à tort³¹. Le biais de mémoire dans les études rétrospectives survient quant à lui en cas de mémorisation différentielle chez les cas et les témoins. Si la sur- ou sous-déclaration (des cas relativement aux témoins) est homogène pour l'ensemble des aliments, ajuster sur l'apport calorique total permet de limiter le biais de mémoire³². En revanche, si les cas se rappellent préférentiellement la consommation de certains aliments, ajuster sur l'apport calorique total ne suffit pas à réduire ce biais et peut même l'aggraver²⁰. Notons que la méthode de partition, en ajustant sur l'apport des autres macronutriments, présente les mêmes avantages et limites en termes de correction des erreurs de mesure ou du biais de mémoire que les méthodes de substitution^{20,31}. Les deux arguments énoncés

pour les méthodes de substitution sont donc à la fois incertains et non spécifiques puisqu'ils pourraient s'appliquer à la méthode de partition.

Avantages et inconvénients de chacune des méthodes

Méthode standard

L'inconvénient fondamental de la méthode standard est d'introduire dans un même modèle deux variables très corrélées, l'apport brut en nutriment et l'apport calorique total, ce qui peut rendre l'interprétation difficile et empêcher la convergence des algorithmes d'estimation des paramètres. De plus, cette forte corrélation entraîne une surestimation de la variabilité de l'apport en nutriment puisque la variabilité observée (mesurée par la variance empirique de l'apport en nutriment) est fortement déterminée par la variabilité de l'apport calorique total. Par conséquent, lorsque l'apport en nutriment est catégorisé, la méthode standard estime le risque associé à un incrément d'apport nutritionnel plus important que dans la méthode des résidus ou des densités, mais en pratique difficile à atteindre de lui-même et sans variation des autres apports nutritionnels.

Méthode des résidus

La méthode des résidus a pour avantage d'éliminer totalement la corrélation entre apport nutritionnel et apport calorique total, facilitant ainsi l'estimation des paramètres et l'interprétation. Par ailleurs, elle apporte une estimation plus réaliste de la variabilité de l'apport en nutriment puisque la variance des résidus équivaut à la variabilité de l'apport brut *entre sujets ayant un même apport calorique total*. De ce fait, alors que le paramètre associé au nutriment d'intérêt est identique à celui de la méthode standard lorsque la variable est continue, en catégoriel au contraire, le paramètre estimé par la méthode des résidus pourrait mieux refléter l'effet de substitution du nutriment d'intérêt que la méthode standard. En outre, du fait de l'absence de corrélation entre apport calorique total et apport nutritionnel résiduel, la méthode des résidus est plus robuste à la confusion résiduelle que la méthode standard lorsque l'apport calorique total n'est pas correctement spécifié dans le modèle. L'utilisation de la méthode des résidus est donc préférable pour mesurer un effet de substitution à l'aide d'une variable catégorielle^{15,18,19}. Dans le cas continu ou catégoriel, certains auteurs ajoutent que la méthode des résidus a l'avantage supplémentaire que le paramètre associé à l'apport calorique total représente l'effet biologique de l'énergie^{9,18}. Tel n'est pas le cas cependant lorsque le nutriment d'intérêt contribue à l'apport calorique total, puisque nous avons montré que l'effet de l'énergie mesuré par la méthode des résidus est alors un effet composite de l'addition d'énergie par ce nutriment et par les autres sources d'énergie. La difficulté de l'interprétation du paramètre associé à l'apport calorique total dans la méthode des résidus peut donc constituer un inconvénient, bien que ce ne soit généralement pas l'effet étudié en priorité.

Méthode des densités

En pratique, les densités nutritionnelles sont souvent très corrélées aux apports résiduels^{6,33} et modérément corrélées à l'apport calorique total. Les paramètres d'un modèle des densités ont une interprétation très similaire à ceux d'un modèle des résidus, avec le même inconvénient d'un effet composite pour le paramètre de l'apport calorique total. Les densités nutritionnelles ont plusieurs avantages pratiques sur la méthode des résidus : on les calcule sans le recours à un modèle statistique, les nutritionnistes en sont familiers comme mesure de la composition de l'alimentation et elles sont utilisées dans les recommandations nationales en matière de nutrition³⁴. La méthode des densités permet donc de chiffrer directement l'augmentation ou la réduction du risque d'une pathologie donnée qui serait liée à la modification des habitudes alimentaires conformément aux recommandations nutritionnelles. Malgré des résultats souvent proches entre la méthode des densités et celle des résidus, la méthode des densités a fausement l'attrait de la simplicité. Sous un seul terme, la densité nutritionnelle N/T comprend en fait deux composantes, l'apport en nutriment N et l'inverse de l'apport calorique total T . En conséquence, si l'apport en nutriment considéré est nettement moins variable que l'apport calorique total, la valeur de la densité sera proche de l'inverse de ce dernier à une constante multiplicative près. En l'absence d'ajustement supplémentaire sur l'apport calorique total (modèle 2a), ceci peut donc faire apparaître à tort une association du risque de maladie avec la densité du nutriment d'intérêt dans le sens inverse de celle de l'apport calorique total. En cas d'ajustement supplémentaire sur l'apport calorique total (modèle 2b), introduire une variable et une autre variable corrélée à l'inverse de la première dans le même modèle peut entraîner des difficultés pour la convergence des algorithmes d'estimation et pour l'interprétation.

Méthode de partition

La méthode de partition étant la seule qui permette l'estimation directe de l'effet d'addition, l'utiliser en combinaison avec l'une des trois autres méthodes semble indispensable pour compléter l'interprétation des effets des nutriments et de l'apport calorique¹⁹. Il est certes possible, lorsque les apports nutritionnel et calorique sont codés en continu, de retrouver l'effet d'addition à partir des méthodes standard ou des résidus, du fait de l'équivalence entre les méthodes. Néanmoins, cette équivalence disparaissant avec la catégorisation, il est alors indispensable de conserver l'approche par partition. Il faut toutefois noter que, comme dans la méthode standard, les quantiles d'apport sont définis à partir de l'apport brut dont la variance surestime la variabilité réelle du nutriment. Par conséquent, l'écart entre les quantiles extrêmes peut être associé à un risque plus grand que dans les méthodes des résidus ou des densités mais correspondre, comme dans la méthode standard, à un incrément pratiquement irréalisable. En revanche, la corrélation de l'apport en nutriment considéré étant moins grande avec l'apport en autres macronutriments qu'avec l'apport calorique total, le risque est moindre de trouver des profils de covariables d'effectif nul ou faible lorsque l'on ajuste sur l'apport calorique en catégories, ce qui assure une plus grande précision des estimations qu'avec la méthode standard.

Extensions des méthodes et cas particuliers

Même lorsque l'intérêt d'une étude porte principalement sur un seul nutriment, une limite des méthodes présentées est d'identifier un effet des autres sources d'énergie dans leur ensemble sans savoir à quel(s) nutriment(s) l'attribuer spécifiquement. Par exemple, un effet nul des autres macronutriments pourrait être la somme d'effets de sens opposés et sembler nul à tort alors que substituer le nutriment d'intérêt à l'un ou à l'autre de ces autres macronutriments pris séparément pourrait faire apparaître des associations positives ou négatives. Plusieurs auteurs^{13,35} ont ainsi préconisé de ne pas se limiter aux quatre modèles d'ajustement sur l'apport calorique exposés jusqu'ici mais d'étudier tous les macronutriments simultanément. En pratique, ces méthodes de décomposition des sources d'énergie sont des extensions des méthodes d'ajustement sur l'apport calorique : soit on introduit toutes les sources d'énergie sous forme d'apports bruts sans inclure l'apport calorique total (extension de la méthode de partition), soit on introduit toutes les sources d'énergie sauf une sous forme d'apports bruts, résiduels ou de densités, ainsi que l'apport calorique total (extensions des méthodes standard, des résidus et des densités, respectivement). Les effets mesurés restent de type addition ou substitution mais il s'agit de l'addition d'un nutriment alors que l'apport de chaque autre source (et non leur somme) est constant, ou de la substitution d'un nutriment à un autre précisément identifié.

La décomposition des sources d'énergie peut être poursuivie pour chaque macronutriment en introduisant ses constituants dans les différents modèles, par exemple les acides gras saturés, monoinsaturés et polyinsaturés à la place des lipides (voir par exemple³⁶). Une autre extension pour la méthode des résidus consiste à emboîter le calcul des résidus⁷. Ainsi, pour étudier la relation entre apport en acides gras saturés et risque de maladie, on peut utiliser un modèle où l'on introduit l'apport calorique total, l'apport résiduel en lipides obtenu après régression sur l'apport calorique total et l'apport résiduel en acides gras saturés obtenu après régression sur l'apport résiduel en lipides (voir par exemple³⁷). L'effet mesuré par le paramètre associé à l'apport en acides gras saturés dans un tel modèle est alors un effet de substitution aux autres acides gras à apport lipidique constant. Dans tous ces modèles de décomposition, l'ajustement mutuel pose cependant des problèmes de multicollinéarité et peut conduire à diminuer la précision des estimations des paramètres des modèles, ainsi que la puissance statistique pour tester ces paramètres et conclure ou non à l'existence d'une association avec le risque de maladie étudiée.

L'apport en éthanol se distingue des autres apports en macronutriments du fait de sa plus faible corrélation avec l'apport calorique total. L'éthanol provient en effet d'un nombre limité d'aliments, des boissons alcoolisées pour la plupart, qui contiennent peu ou pas d'autres macronutriments. La question de l'ajustement sur l'apport calorique est donc moins cruciale pour l'apport en éthanol et, en pratique, les méthodes d'ajustement sur l'apport calorique ont rarement été utilisées pour mettre en évidence une association entre consommation d'alcool et risque de maladie. En revanche, il peut s'avérer utile d'ajuster sur la consommation d'alcool car celle-ci pourrait être un facteur de confusion de l'association entre risque de maladie et apport en autres nutriments³⁸. Ainsi, pour l'étude de la relation entre apport lipidique et risque de cancer du sein, plusieurs auteurs ont ajusté sur la consommation d'alcool en plus de l'apport calorique total (voir par exemple³⁶). Alternativement, on pourrait préconiser d'ajuster sur l'apport en éthanol et sur l'apport calorique *excepté l'alcool* dès lors que la consommation d'alcool pourrait être liée au risque de pathologie considérée. Dans les méthodes des résidus et des densités, les résidus sont alors issus de la régression de l'apport en nutriment sur l'apport calorique total alcool non compris et les densités nutritionnelles sont calculées en divisant les apports bruts par l'apport calorique total alcool non compris.

Notons que cette démarche est cohérente avec les recommandations nutritionnelles qui sont données pour la plupart en termes de pourcentage d'énergie alcool non compris³⁴.

Lorsque l'on s'intéresse à la relation entre risque de maladie et apports en nutriments non énergétiques (fibres, cholestérol, vitamines, minéraux, etc.), les arguments développés jusqu'ici pour les macronutriments perdent en partie leur sens¹⁷. Les apports en ces nutriments sont certes pour la plupart corrélés à l'apport calorique total mais dans une moindre mesure que les apports en macronutriments, du fait qu'ils ne contribuent pas à l'apport énergétique. Les recommandations nutritionnelles en ce qui concerne les nutriments non énergétiques sont de ce fait données en quantités absolues et non rapportées à l'énergie³⁴. Certains auteurs ont constaté néanmoins que l'apport calorique et l'apport en nutriment non énergétique (calcium par exemple) peuvent masquer leurs effets respectifs s'ils ne sont pas ajustés mutuellement³⁹. Techniquement, seules les méthodes standard, des résidus et des densités, sont pertinentes pour ajuster sur l'apport calorique. Dans le modèle standard, le paramètre associé à l'apport calorique total (a_2) reflète l'effet de l'énergie totale (générique et de l'ensemble des sources d'énergie) puisque le nutriment étudié n'apporte pas d'énergie. Pour cette raison, la méthode standard peut être utilisée de préférence aux autres quand le nutriment d'intérêt est faiblement corrélé à l'apport calorique total¹⁷. En cas de corrélation plus importante, mieux vaut recourir à la méthode des résidus, le paramètre associé à l'apport résiduel restant identique au paramètre associé à l'apport brut en nutriment dans la méthode standard dans le cas continu. La méthode des densités peut être adaptée en considérant la fraction énergétique, généralement exprimée en grammes ou autre unité standard pour 1 000 kcal d'apport calorique total.

Conclusion

En premier lieu, il revient à l'investigateur de décider lequel des effets décrits ci-dessus est le plus pertinent selon la question scientifique posée. Bien que les estimateurs des coefficients de régression se déduisent les uns des autres dans le cas d'un codage continu, il est plus facile d'utiliser le modèle dans lequel l'effet choisi est représenté par un paramètre unique (par exemple le modèle standard ou des résidus pour l'effet de substitution) : l'estimation du paramètre et de son écart-type est alors donnée directement par le logiciel statistique. En cas de codage en catégories, il n'est pas possible en revanche de déduire les estimations d'un modèle à un autre et un choix préalable s'impose donc.

Des trois méthodes qui mesurent un effet de substitution du nutriment d'intérêt à un autre, il vaut mieux éviter le recours à la méthode standard en cas de forte colinéarité entre le nutriment d'intérêt et l'apport calorique total. La méthode des densités pose des difficultés d'interprétation en cas de faible variation de l'apport en nutriment, surtout lorsque le risque de maladie est associé à l'apport calorique total. Dans les autres cas, cependant, on obtient directement avec cette méthode une estimation de la réduction du risque de maladie si l'on réduit (ou augmente) son apport nutritionnel d'un pourcentage de l'apport calorique total, ce qui est utile en pratique pour établir des recommandations nutritionnelles. La méthode des résidus constitue une alternative favorable à la méthode des densités du fait qu'elle élimine la corrélation entre apport en nutriment et apport calorique total. Elle présente néanmoins la limite, commune avec la méthode des densités, que le coefficient de l'apport calorique total est relativement peu pertinent. Enfin, la méthode de partition est la seule permettant l'estimation d'un effet d'addition du nutriment d'intérêt et peut s'avérer utile pour compléter l'interprétation issue des méthodes de substitution.

En l'absence de connaissance préalable sur l'effet impliqué, il nous semble intéressant de combiner les méthodes de substitution avec la méthode de partition. En terme de recommandations nutritionnelles, un coefficient positif associé à un nutriment dans les méthodes standard, des résidus ou des densités devrait alors encourager à substituer des calories d'un autre nutriment aux calories de ce nutriment ; un coefficient négatif au contraire devrait encourager à substituer des calories de ce nutriment à des calories des autres macronutriments. Un coefficient positif (négatif) associé à ce nutriment dans le modèle de partition devrait inciter à réduire (augmenter) l'apport de ce nutriment sans substitution par d'autres nutriments.

Les principes généraux des méthodes présentées dans cet article sont extrapolables à d'autres contextes que l'épidémiologie nutritionnelle. Ainsi, pour dissocier les effets respectifs de covariables fortement corrélées, telles que le poids et la taille, il serait possible de recourir aux méthodes standard, des résidus ou des densités ; la méthode de partition pourrait être utilisée en plus de ces méthodes pour dissocier les effets de covariables dont l'une est comprise dans l'autre (par exemple les nombres de cycles menstruels jusqu'à la première grossesse à terme et jusqu'à la ménopause).

Remerciements

Anne Thiébaud a bénéficié du soutien financier de la Fondation de France puis de la Ligue Nationale contre le Cancer. Emmanuelle Kesse a été soutenue par l'Association pour la Recherche sur le Cancer, puis par la Fondation pour la Recherche Médicale. L'étude E3N-EPIC est réalisée grâce à la collaboration de la Ligue Nationale contre le Cancer et de ses comités départementaux, du programme Europe contre le Cancer de l'Union Européenne, de la société 3M, de la Mutuelle Générale de l'Éducation Nationale, de l'Institut Gustave Roussy, de l'Institut National de la Santé et de la Recherche Médicale et de plusieurs Conseils généraux. Les auteurs sont également reconnaissants envers les femmes participant à l'étude et les médecins pour leur active collaboration.

Références

1. Willett WC. Implications of total energy intake for epidemiologic studies of breast and large-bowel cancer. *Am J Clin Nutr* 1987; 45: 354-60.
2. Willett WC. *Nutritional Epidemiology*. 2nd ed. New York, NY: Oxford University Press, 1998.
3. Lyon JL, Gardner JW, West DW, Mahoney AM. Methodological issues in epidemiological studies of diet and cancer. *Cancer Res* 1983; 43: 2392s-6s.
4. Willett W, Stampfer MJ. Total energy intake: implications for epidemiologic analyses. *Am J Epidemiol* 1986; 124: 17-27.
5. Howe GR, Miller AB, Jain M. Re: "Total energy intake: implications for epidemiologic analyses" (letter). *Am J Epidemiol* 1986; 124: 157-9.
6. Shekelle RB, Nichaman MZ, Raynor WJ, Jr. Re: "Total energy intake: implications for epidemiologic analyses" (letter). *Am J Epidemiol* 1987; 126: 980-3.
7. Pike MC, Bernstein L, Peters RK. Re: "Total energy intake: implications for epidemiologic analyses" (letter). *Am J Epidemiol* 1989; 129: 1312-3.
8. Howe GR. Re: "Total energy intake: implications for epidemiologic analyses" The first author replies (letter). *Am J Epidemiol* 1989; 129: 1314-5.
9. Willett WC. Total energy intake and nutrient composition: dietary recommendations for epidemiologists. *Int J Cancer* 1990; 46: 770-1.
10. Palmgren J, Kushi LH. Re: "Total energy intake: implications for epidemiologic analyses" (letter). *Am J Epidemiol* 1991; 133: 1291-3.
11. Kushi LH, Sellers TA, Potter JD, Nelson CL, Munger RG, Kaye SA, et al. Dietary fat and postmenopausal breast cancer. *J Natl Cancer Inst* 1992; 84: 1092-9.
12. Pike MC, Bernstein L, Peters RK. Re: "Dietary fat and postmenopausal breast cancer" (letter). *J Natl Cancer Inst* 1992; 84: 1666-7.
13. Pike MC, Peters RK, Bernstein L. Re: "Total energy intake: implications for epidemiologic analyses" (letter). *Am J Epidemiol* 1993; 137: 811-3.
14. Kipnis V, Freedman LS, Brown CC, Hartman A, Schatzkin A, Wacholder S. Interpretation of energy adjustment models for nutritional epidemiology. *Am J Epidemiol* 1993; 137: 1376-80.
15. Brown CC, Kipnis V, Freedman LS, Hartman AM, Schatzkin A, Wacholder S. Energy adjustment methods for nutritional epidemiology: the effect of categorization. *Am J Epidemiol* 1994; 139: 323-38.
16. Wacholder S, Schatzkin A, Freedman LS, Kipnis V, Hartman A, Brown CC. Can energy adjustment separate the effects of energy from those of specific macronutrients? *Am J Epidemiol* 1994; 140: 848-55.
17. Mackerras D. Energy adjustment: the concepts underlying the debate. *J Clin Epidemiol* 1996; 49: 957-62.
18. Willett WC, Howe GR, Kushi LH. Adjustment for total energy intake in epidemiologic studies. *Am J Clin Nutr* 1997; 65: 1220S-8S.
19. Freedman LS, Kipnis V, Brown CC, Schatzkin A, Wacholder S, Hartman AM. Comments on "Adjustment for total energy intake in epidemiologic studies". *Am J Clin Nutr* 1997; 65: 1229S-31S.
20. Bellach B, Kohlmeier L. Energy adjustment does not control for differential recall bias in nutritional epidemiology. *J Clin Epidemiol* 1998; 51: 393-8.
21. Clavel-Chapelon F, Jadand C, Goulard H, Guibout-Peigné C. E3N, étude de cohorte sur les facteurs de risque de cancer auprès de femmes adhérentes de la MGEN. Description du protocole, des principales caractéristiques et de la population. *Bull Cancer* 1996 ; 83 : 1008-13.
22. van Liere MJ, Lucas F, Clavel F, Slimani N, Villeminot S. Relative validity and reproducibility of a French dietary history questionnaire. *Int J Epidemiol* 1997; 26 (Suppl 1): S128-36.
23. Riboli E, Kaaks R. The EPIC Project: rationale and study design. *European Prospective Investigation into Cancer and Nutrition*. *Int J Epidemiol* 1997; 26 (Suppl 1): S6-14.
24. Lucas F, Niravong M, Villeminot S, Kaaks R, Clavel-Chapelon F. Estimation of food portion size using photographs: validity, strengths, weaknesses and recommendations. *J Human Nutr Dietetics* 1995; 8: 65-74.
25. Favier JC, Ireland-Ripert J, Toque C, Feinberg M. *Répertoire Général des Aliments. Table de composition (Composition tables)*. 2nd ed. Paris : INRA, CIQUAL-REGAL, Tec & Doc Lavoisier, 1995.
26. McCullagh P, Nelder JA. *Generalized linear models*. 2nd ed. London : Chapman & Hall Ltd., 1989.

27. Gail MH, Wieand S, Piantadosi S. Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. *Biometrika* 1984; 71: 431-44.
28. Schoenfeld D. Partial residuals for the proportional hazards regression model. *Biometrika* 1982; 69: 239-41.
29. IARC Working Group on the Evaluation of Cancer-Preventive Strategies. Weight control and physical activity. Lyon: International Agency for Research on Cancer, 2002.
30. Day NE, McKeown N, Wong MY, Welch A, Bingham S. Epidemiological assessment of diet: a comparison of a 7-day diary with a food frequency questionnaire using urinary markers of nitrogen, potassium and sodium. *Int J Epidemiol* 2001; 30: 309-17.
31. Kipnis V, Freedman LS, Brown CC, Hartman AM, Schatzkin A, Wacholder S. Effect of measurement error on energy-adjustment models in nutritional epidemiology. *Am J Epidemiol* 1997; 146: 842-55.
32. Trichopoulos D, Tzonou A, Katsouyanni K, Trichopoulou A. Diet and cancer: the role of case-control studies. *Ann Nutr Metab* 1991; 35 (Suppl 1): 89-92.
33. Mattisson I, Wirfält E, Gullberg B, Berglund G. Fat intake is more strongly associated with lifestyle factors than with socio-economic characteristics, regardless of energy adjustment approach. *Eur J Clin Nutr* 2001; 55: 452-61.
34. CNRS, CNERNA. Apports nutritionnels conseillés pour la population française. 3rd ed. Paris : Tec & Doc Lavoisier, 2001.
35. Decarli A, Favero A, La Vecchia C, Russo A, Ferraroni M, Negri E, et al. Macronutrients, energy intake, and breast cancer risk: implications from different models. *Epidemiology* 1997; 8: 425-8.
36. Wolk A, Bergström R, Hunter D, Willett W, Ljung H, Holmberg L, et al. A prospective study of association of monounsaturated fat and other types of fat with risk of breast cancer. *Arch Intern Med* 1998; 158: 41-5.
37. De Stefani E, Deneo-Pellegrini H, Mendilaharsu M, Ronco A. Essential fatty acids and breast cancer: a case-control study in Uruguay. *Int J Cancer* 1998; 76: 491-4.
38. Kesse E, Clavel-Chapelon F, Slimani N, van Liere M. Do eating habits differ according to alcohol consumption? Results of a study of the French cohort of the European Prospective Investigation into Cancer and Nutrition (E3N-EPIC). *Am J Clin Nutr* 2001; 74: 322-7.
39. Peters RK, Pike MC, Garabrant D, Mack TM. Diet and colon cancer in Los Angeles County, California. *Cancer Causes Control* 1992; 3: 457-73.

ANNEXE 1. — Variance de l'apport nutritionnel résiduel

La méthode des résidus (modèle 3) utilise le résidu de la régression linéaire de l'apport en nutriment considéré (N) sur l'apport calorique total (T). Le résidu ainsi obtenu, $N' = N - (\hat{\alpha}_0 + \hat{\alpha}_1 T)$, est de variance moindre que l'apport nutritionnel brut N . En effet, si l'on note $\text{Var}(N')$ la variance du résidu et $\text{Var}(N)$ la variance de l'apport brut, $\text{Var}(N')$ peut s'exprimer en fonction de $\text{Var}(N)$:

$$\text{Var}(N') = \text{Var}(N - \hat{\alpha}_0 - \hat{\alpha}_1 T) = \text{Var}(N) + \hat{\alpha}_1^2 \text{Var}(T) - 2\hat{\alpha}_1 \text{Cov}(N, T).$$

Or, dans le cas de la régression linéaire simple avec constante, l'estimateur du maximum de vraisemblance pour la pente est

$$\hat{\alpha}_1 = \frac{\text{Cov}(N, T)}{\text{Var}(T)} = r(N, T) \sqrt{\frac{\text{Var}(N)}{\text{Var}(T)}}$$

où $\text{Var}(T)$ désigne la variance de l'apport calorique total, $\text{Cov}(N, T)$ la covariance entre N et T et $r(N, T) = \frac{\text{Cov}(N, T)}{\sqrt{\text{Var}(N)\text{Var}(T)}}$ le coefficient de corrélation linéaire de Pearson entre N et T (compris entre -1 et $+1$).

D'où :

$$\begin{aligned} \text{Var}(N') &= \text{Var}(N) + r^2(N, T) \text{Var}(N) - 2r \sqrt{\frac{\text{Var}(N)}{\text{Var}(T)}} \times r(N, T) \sqrt{\text{Var}(N)\text{Var}(T)} \\ &= \text{Var}(N) + r^2(N, T) \text{Var}(N) - 2r^2(N, T) \text{Var}(N) = \text{Var}(N) \times (1 - r^2(N, T)) \end{aligned}$$

Étant donné que $r^2(N, T)$ est compris entre 0 et 1, on démontre ainsi que $\text{Var}(N') \leq \text{Var}(N)$.

ANNEXE 2. — Effets spécifiques des nutriments et effet générique de l'énergie

Partant du modèle 5 qui n'est pas identifiable,

$$g(D|N,M,T) = \beta_0 + \beta_N N + \beta_M M + \beta_T T,$$

on peut exprimer les paramètres des modèles standard, des résidus et de partition comme combinaisons linéaires de β_N , β_M et β_T .

Méthode standard. — La méthode standard (modèle 1) suppose que le risque de maladie ne dépend que du nutriment d'intérêt et des calories totales, soit : $g(D|N,T) = a_0 + a_1 N + a_2 T$. En éliminant le terme M dans l'équation du modèle 5, on obtient donc :

$$g(D|N,T) = \beta_0 + \beta_N N + \beta_M (T - N) + \beta_T T = \beta_0 + (\beta_N - \beta_M)N + (\beta_M + \beta_T)T$$

et l'on déduit les relations suivantes entre paramètres des modèles 1 et 5 :

$$a_1 = \beta_N - \beta_M \text{ et } a_2 = \beta_M + \beta_T.$$

Méthode des résidus. — La méthode des résidus (modèle 3) suppose également que le risque de maladie ne dépend que du nutriment d'intérêt et des calories totales, l'apport en nutriment étant exprimé sous forme de résidu, soit :

$$g(D|N,T) = c_0 + c_1 N' + c_2 T.$$

En faisant donc apparaître le résidu N' dans l'équation du modèle 5, on obtient :

$$\begin{aligned} g(D|N,T) &= \beta_0 + (\beta_N - \beta_M)N + (\beta_M + \beta_T)T \\ &= \beta_0 + (\beta_N - \beta_M)(N' + \hat{\alpha}_0 + \hat{\alpha}_1 T) + (\beta_M + \beta_T)T \\ &= \beta_0 + \hat{\alpha}_0 (\beta_N - \beta_M) + (\beta_N - \beta_M)N' + [\hat{\alpha}_1 \beta_N + (1 - \hat{\alpha}_1)\beta_M + \beta_T]T. \end{aligned}$$

Ainsi $c_1 = \beta_N - \beta_M$ et $c_2 = \hat{\alpha}_1 \beta_N + (1 - \hat{\alpha}_1)\beta_M + \beta_T$.

Méthode de partition. — Contrairement aux méthodes standard et des résidus, la méthode de partition (modèle 4) suppose que le risque de maladie dépend du nutriment d'intérêt et des autres sources d'énergie, mais pas de l'apport calorique total, soit :

$$g(D|N,M) = d_0 + d_1 N + d_2 M.$$

À partir de l'équation du modèle 5, on obtient :

$$g(D|N,M) = \beta_0 + \beta_N N + \beta_M M + \beta_T (N + M) = \beta_0 + (\beta_N + \beta_T)N + (\beta_M + \beta_T)M$$

Ainsi $d_1 = \beta_N + \beta_T$ et $d_2 = \beta_M + \beta_T$.

ANNEXE 3. — Équivalence mathématique des modèles nutritionnels

En développant l'équation du modèle 1 (méthode standard) de façon à faire apparaître l'apport résiduel en nutriment N' , on obtient :

$$g(D | N, T) = a_0 + a_1[N' + (\hat{\alpha}_0 + \hat{\alpha}_1 T)] + a_2 T = (a_0 + \hat{\alpha}_0 a_1) + a_1 N' + (\hat{\alpha}_1 a_1 + a_2) T.$$

On en déduit les relations d'équivalence suivantes entre paramètres des méthodes standard et des résidus : $c_1 = a_1$ et $c_2 = \hat{\alpha}_1 a_1 + a_2$ (ou, inversement, $a_1 = c_1$ et $a_2 = c_2 - \hat{\alpha}_1 c_1$).

En développant la même équation de façon à faire apparaître l'énergie apportée par les autres macronutriments M , on obtient :

$$g(D | N, T) = a_0 + a_1 N + a_2 (N + M) = a_0 + (a_1 + a_2) N + a_2 M,$$

ce qui permet d'établir les relations d'équivalence suivantes entre paramètres des méthodes standard et de partition :

$$d_1 = a_1 + a_2 \text{ et } d_2 = a_2 \text{ (inversement, } a_1 = d_1 - d_2 \text{ et } a_2 = d_2).$$

Les estimateurs équivalents ont même écart-type, comme démontré par Kipnis *et al.* dans le cas de la régression linéaire [14]. L'équivalence mathématique se traduit aussi par l'égalité des log-vraisemblances pour ces trois méthodes, standard, des résidus et de partition [7].

ANNEXE 4. — Interprétation du modèle incluant seulement l'apport calorique total

Pour déterminer l'effet de l'apport calorique total, on ajuste souvent un modèle avec l'apport calorique total seul, sous la forme :

$$g(D | T) = \beta_0 + \beta_T T, \text{ soit } g(D | T, N, M) = \beta_0 + \beta_N N + \beta_T T.$$

Par ailleurs, le modèle complet s'écrit :

$$\begin{aligned} g(D | T, N, M) &= \beta_0 + \beta_N N + \beta_M M + \beta_T T \\ &= \beta_0 + (\beta_N + \beta_T) N + (\beta_M + \beta_T) T. \end{aligned}$$

On en déduit la contrainte implicite que $\beta_T = \beta_N + \beta_T = \beta_M + \beta_T$, c'est-à-dire que $\beta_N = \beta_M$.

Ainsi le modèle incluant l'apport calorique total seul s'écrit :

$$g(D | T) = \beta_0 + \beta_T T = \beta_0 + (\beta_N + \beta_T) T.$$

Autrement dit, β_T mesure la somme de l'effet spécifique de N , supposé égal à celui de M , et de l'effet générique de T .