

Stratégies d'analyse et risques en Epidémiologie Nutritionnelle

B. Maire^a

F. Delpuech^a

1 Introduction

Des millions d'enfants à travers le monde souffrent de ce que l'on appelle la **"malnutrition protéino-énergétique"**, terme forgé il y a de nombreuses années pour rendre compte des aspects complexes et de l'étiologie incertaine des formes que peut prendre la maladie. Cette malnutrition de l'enfance est probablement connue depuis les premiers temps de l'humanité, les conditions de vie et d'alimentation du jeune enfant n'ayant certainement pas toujours été idéales. Toutefois l'histoire médicale de cette pathologie en est assez récente; ainsi le kwashiorkor, la forme clinique la plus spectaculaire, a été porté à la connaissance du monde scientifique médical en 1933 par le Dr Cicely Williams, à la suite d'observations concernant un syndrome "nouveau" apparemment lié à des problèmes de carence alimentaire, au Ghana. La recherche médicale n'a cessé de se préoccuper depuis des aspects complexes du traitement et de la physiopathologie des cas de malnutrition sévère.

Mais l'attention des services de santé des pays en développement s'est porté bien vite aussi sur ce que l'on a appelé, par référence à un terme alors largement en vogue en épidémiologie, "la partie immergée de l'iceberg". En dehors des périodes de calamités ou de famine, la malnutrition clinique ne concerne qu'un faible pourcentage des jeunes enfants dans une communauté; mais un grand nombre présente un déficit physique révélateur d'une carence alimentaire sous-jacente. L'épidémiologie de la malnutrition s'est déplacée vers ces enfants, en essayant notamment de dénombrer les enfants malnourris à des degrés divers, avec l'aide de l'anthropométrie nutritionnelle.

2 Anthropométrie et Statistique

On appelle anthropométrie nutritionnelle le corpus de concepts et de techniques qui permettent d'estimer l'état nutritionnel à partir de données de croissance des jeunes enfants. La maintenance de la masse corporelle existante et l'accrétion de la masse tissulaire nouvelle sont dépendantes d'un apport adéquat de nutriments aux cellules; le déroulement de la croissance va donc être particulièrement sensible aux variations d'apport en ces nutriments.

^aORSTOM, Laboratoire de Nutrition Tropicale (LNT), D.E.S., UR 4F. B.P. 5045, 34032 Montpellier cedex.

La base de l'anthropométrie nutritionnelle est donc la suivante: la croissance se caractérise par des modifications rapides de la taille et des masses corporelles maigre et grasse; pour apprécier la qualité de cette croissance, il est nécessaire de comparer les données de poids et taille aux proportions attendues pour l'âge, le sexe et le groupe ethnique.

L'origine de cette discipline remonte probablement aux débuts de la révolution industrielle en Europe. Si on s'était bien intéressé auparavant à l'anthropométrie, c'était surtout, depuis la Renaissance, dans le but d'estimer les proportions corporelles idéales de "l'homme moyen". Mais vers 1830, en Angleterre tout particulièrement où l'embauche de jeunes enfants de moins de dix ans était courante dans certaines entreprises, l'opinion publique s'émut des conditions de vie sanitaires et alimentaires déplorables de ces enfants. Des commissions parlementaires commandèrent des investigations comprenant notamment des relevés de poids et de tailles qui mirent en évidence des mesures extrêmement faibles. Parallèlement, en France, la médecine sociale prenait aussi son essor; c'est ainsi que Villermé démontrait une relation étroite entre croissance en taille et pauvreté. On manquait cependant à l'époque de véritables mesures de référence.

Cette discipline nouvelle évolua parallèlement à l'essor de la statistique, celle-ci se nourrissant des données fournies par l'anthropométrie; vers la fin du 19^{ème} siècle de grands noms comme Galton, Pearson ou Boas ont contribué à l'émergence de concepts essentiels, comme la corrélation - le concept de régression vers la moyenne a été mis en évidence à partir de mesures de tailles de pères et de leurs fils - ou encore de techniques comme celle de la construction des "percentiles" par exemple. L'exploitation statistique des données de croissance fit émerger les premiers débats importants sur la part de la génétique par rapport à celle de l'environnement, et de l'alimentation en particulier. De véritables tables de référence du poids et de la taille en fonction de l'âge furent disponibles vers les années 1940, après qu'un certain nombre d'études longitudinales aient été menées à bien, aux Etats Unis essentiellement. Mais ce n'est que récemment que l'anthropométrie nutritionnelle, confinée par la suite aux cénacles étroits de l'anthropologie physique en est ressorti pour être largement utilisée dans des sphères de plus en plus vastes; car est apparue ces dernières années la nécessité d'apprécier systématiquement l'état nutritionnel des jeunes enfants d'âge préscolaire dans les pays en voie de développement pour alerter l'opinion et les dirigeants politiques sur les conditions de vie de ces enfants, comme pour apprécier les progrès accomplis.

Après un certain nombre de débats, l'OMS a progressivement conforté de 1976 à 1986 une approche standard de la mesure de l'état nutritionnel non pas seulement à partir d'un indice de poids ajusté pour l'âge et le sexe par rapport aux valeurs de référence, mais à partir de deux indices complémentaires: le poids pour la taille ou poids comparé au poids d'un enfant de même taille, soit un indice de maigreur ajusté pour la taille et le sexe, et la taille pour l'âge ou taille comparée à celle d'un enfant de même âge, soit un indice de retard de croissance en taille ajusté sur l'âge et le sexe.

En ce qui concerne le mode d'expression, les *percentiles* sont le mode le plus approprié; en pratique, pour une limite évidente lorsque la population étudiée a des valeurs plus basses que celles de la référence, il est plus commode d'utiliser d'autres modes d'expression. Le *pourcentage de la médiane*, longtemps utilisé, est de nos jours remplacé par un *score en écarts-types*. Ce mode de calcul est plus rigoureux au plan statistique, tient compte des variations des distributions selon

l'âge dans la référence et, avantage certain de présentation, permet d'exprimer les déficits de poids pour taille et taille pour âge de manière comparable. On confronte ainsi les valeurs observées à celles d'une population de référence composée d'enfants sains et bien nourris. Comme les valeurs de taille et de poids se répartissent généralement selon un mode proche d'une distribution de Gauss, on se réfère à des multiples de l'écart-type de la population de référence pour exprimer la position d'un individu par rapport à l'ensemble de référence pour l'indice considéré comme le montre la courbe de fréquence portée en Fig.1 .

3 Diagnostic de malnutrition

Si l'on sait mesurer l'état nutritionnel au travers des paramètres de la croissance du jeune enfant, il en va tout autrement lorsqu'il faut porter un jugement de valeur sur cet état nutritionnel: est-il bon ou mauvais ? Depuis plusieurs décennies, les chercheurs et les praticiens ont tenté d'apporter une réponse. Devant le manque de méthodes adaptées pour apprécier l'état nutritionnel en terme de "maladie" ou "non maladie", on s'est tourné vers l'anthropométrie nutritionnelle, qui, malgré ses limites évidentes, a le mérite de mesurer le phénomène en continu. On a tenté de lui appliquer une méthodologie statistique rigoureuse, suivant en cela l'évolution de la médecine moderne confrontée de plus en plus souvent à une appréciation en continu des phénomènes de santé.

Le raisonnement, classique pour tous tests diagnostiques, est le suivant: on identifie et compte les individus mal- ou bien- nourris en fonction d'un seuil spécifique pour différents indicateurs. Le seuil idéal pour un indicateur donné serait celui qui résulte en une séparation complète des deux groupes. Mais comme dans toute classification établie de cette façon, il faut tenir compte d'un certain nombre de sujets mal classés: tandis que certains individus bien-nourris peuvent être classés faussement comme mal-nourris (faux positifs), des individus mal-nourris peuvent être considérés comme bien-nourris (faux négatifs), car les distributions des caractéristiques biologiques d'individus "malades" et "non malades" se superposent toujours en partie.

En l'absence de critères externes de validité ou d'instruments de mesure directe de la réalité que l'on cherche à caractériser on dispose en général de trois approches pour élaborer des critères d'anormalité: l'approche normative, l'approche fonctionnelle, et l'approche réponse au traitement. Cette dernière correspond à une réponse sélective à un traitement qui vise à modifier un déterminant de la maladie chez des sujets choisis sur la base de résultats du test diagnostic. Elle est rarement utilisée dans des populations, et elle ne peut être utilisée sans risque avec n'importe quel indicateur. Ce sont donc les deux premières qui sont généralement utilisées.

3.1 L'approche normative:

Elle repose sur les propriétés statistiques de la distribution normale des valeurs d'une population de référence que l'on suppose a priori être en bonne santé. Ceci implique de comparer les valeurs observées d'un indicateur avec celles dérivées d'une population connue, en bonne santé, placée dans des conditions environnementales optimales, et de considérer comme "anormales" toutes les valeurs qui se placent en dehors de la distribution des valeurs dans cette population de référence.

En pratique, c'est cette approche qui a été retenue de tous temps, en l'absence de critères de risque fonctionnels universels. Le seuil fixé initialement au 3ème percentile, a par la suite été porté à -2 écarts-types, seuil équivalent en score d'écart-type (exclusion de 2,27% des individus) pour tous les indicateurs. L'OMS recommande cependant dans un souci de rigueur, de soustraire aux chiffres de prévalence ainsi obtenus les 2,27% de la population de référence, soit les faux positifs. Si cette méthode permet un diagnostic aisé, elle reste insatisfaisante pour le calcul de la prévalence; aussi, plus récemment, en poussant jusqu'au bout cette logique normative, on a proposé de donner un chiffre de prévalence correspondant à l'ensemble des individus en dehors de la population de référence aux différents seuils (Mora, 1989).

Soit une prévalence normalisée de la malnutrition du jeune enfant, définie comme une anomalie anthropométrique, en utilisant la formule simple :

$$PN = PM - FP + FN$$

où PM représente la prévalence mesurée, c'est à dire la proportion d'enfants de la population considérée qui est en dessous d'un certain seuil de la distribution normale de référence, FP le nombre de faux positifs, où la proportion d'individus normalement attendus en dessous de ce seuil dans la population de référence et FN le nombre de faux négatifs, où la proportion d'individus au dessus du seuil en excès par rapport à la population de référence.

Dans la plupart des populations, les trois indicateurs anthropométriques habituels (poids pour l'âge, taille pour l'âge, et poids pour la taille) ont une distribution plus ou moins symétrique sous forme d'une courbe en cloche. Dans les pays en voie de développement, la distribution des indicateurs anthropométriques est approximativement gaussienne, mais décalée vers la gauche par rapport à celle de la population de référence, avec une légère inflexion et un degré variable de recouvrement dépendant de la distance entre les deux distributions. A titre d'exemple la Figure 2 montre une courbe hypothétique avec une moyenne décalée à gauche d'un écart-type par rapport à celle de la distribution de référence et avec un écart-type de 1,0 également: leur intersection intervient à environ mi-distance entre leurs moyennes, ce qui est toujours le cas quand l'écart-type des deux distributions est du même ordre.

Si l'on ajuste seulement pour les faux positifs comme le suggère l'OMS, cela tend à sous-estimer la prévalence de la malnutrition. En ajustant à la fois pour les faux positifs et les faux négatifs, on obtient une estimation normalisée et plus précise de la prévalence. Calculer la proportion de faux positifs est une opération simple: c'est la proportion d'individus en dessous d'un seuil donné dans une distribution normale; le calcul se fait indépendamment de la distance entre les deux courbes. Au contraire le calcul de la proportion de faux négatifs est dépendant non seulement du seuil choisi mais aussi de la distance entre les deux courbes, puisqu'ils représentent la proportion comprise entre ce seuil et le point d'intersection des courbes. Le calcul mathématique de cette proportion fait appel à la fonction de distribution cumulative de la courbe normale de probabilité; la condition d'application en est que les deux distributions, observée et de référence, soient à peu près normales et que la distance soit exprimée en score d'écart-type.

On obtient ainsi une comparaison mathématique simple et globale des deux distributions, plus parlante qu'une moyenne et un écart-type exprimés en score de l'écart-type de référence, puisque cela se traduit bien par un chiffre de prévalence,

soit une proportion d'individus malnourris; on ne peut cependant pas identifier clairement ces derniers. Son utilisation se conçoit pour évaluer la situation nutritionnelle d'une population: elle fournit un chiffre utile en cas de plaidoyer auprès des décideurs; cette méthode ne permet pas cependant de procéder à un dépistage individuel ni de repérer les bénéficiaires potentiels d'un programme d'intervention.

3.2 L'approche fonctionnelle:

Le principe est de repérer une association significative (avec éventuellement des points de rupture) entre des indices et des changements en terme d'altération visible de la santé: morbidité, handicap, ou mortalité. Cette approche a été assez peu tentée pour les différents indices anthropométriques car la réponse est souvent liée à des conditions locales particulières, ce qui rend toute généralisation impossible. Elle peut se révéler cependant très utile pour préciser la signification des différents seuils définis par la méthode normative.

A titre d'exemple, nous présenterons ici des résultats obtenus au Sénégal, qui montre la liaison très forte qui existe entre mortalité et malnutrition chez les jeunes enfants.

L'étude a eu lieu à Niakhar, dans la région de Fatick, en 1983/1986. Cette région est assez représentative des régions soudano-sahéliennes tant par son climat et ses cultures que par les taux élevés de malnutrition et de mortalité qui la caractérisent. Elle est en outre sous surveillance démographique continue, condition indispensable pour avoir une connaissance précise de l'âge des enfants, contrôler l'exhaustivité de l'enquête et assurer un suivi des décès en tenant compte des mouvements migratoires assez importants. Cette région est en outre totalement homogène au point de vue ethnique, et les variations socio-économiques sont assez faibles, ce qui a permis d'éviter là encore des facteurs de modification signalés dans d'autres études.

L'étude a été réalisée de manière exhaustive sur les enfants de 0-5 ans d'une trentaine de villages. Les mesures anthropométriques ont été réalisées au début de l'étude puis deux fois par an sur deux ans.; ceci compte tenu des variations annuelles et saisonnières importantes de la mortalité et de l'alimentation. Au total 14.257 "périodes-enfants", soit des enfants examinés à un passage, ont été inclus dans l'analyse. On a calculé le risque brut de décès, soit la probabilité de décès au cours d'une période définie après une mesure de l'état nutritionnel et le risque relatif entre différentes catégories pour un indice donné avec la catégorie de référence, celle des valeurs au dessus de la moyenne. Une relation claire a été retrouvée pour l'ensemble des paramètres anthropométriques mesurés, aux seuils généralement indicateurs de malnutrition, avec le risque de décès dans les 6 mois qui suivent la mesure, au delà de six mois d'âge.

Une relation forte et différenciée étant ainsi établie pour les deux indicateurs de base, nous avons examiné celle-ci à différents niveaux de malnutrition. La Figure 3 montre la gradation du risque (risque relatif) pour une évolution de 0,5 écart-type en dessous de la médiane de référence. On voit que l'évolution est progressive pour les deux indicateurs, le risque relatif étant plus élevé dans le cas de la maigreur, évidemment que dans le cas du retard de taille. La relation est significative très tôt pour la maigreur, plus tardivement pour la taille pour l'âge (-3 ET).

Toutefois l'analyse des groupes d'âge les plus touchés par le retard de taille montre que la relation est significative bien avant le seuil classique de -2 ET;

une telle analyse permet de confirmer le bien fondé de l'utilisation du calcul de prévalence normalisée qui est proche de la réalité observée en termes fonctionnels.

La question qui nous préoccupait alors était de savoir si le risque de décès lié au retard de taille était réel ou simple *facteur de confusion* par rapport au risque lié à la maigreur. Nous avons procédé à une *régression logistique*, modèle particulièrement approprié lorsqu'on étudie la relation entre une maladie représentée par une variable en deux modalités (malade/non malade) et des facteurs de risque quantitatifs. ce modèle exprime la probabilité d'être malade, ici la probabilité de décéder D+, connaissant les valeurs des facteurs de risque xi (poids pour taille pt, taille pour âge ta) selon l'expression:

$$\Pr(D^+ | x_1(pt), x_2(ta)) = \frac{1}{1 + e^{-(a + \sum_i b_i x_i)}}$$

ou encore:

$$\text{Logit}(P) = a + \sum_i b_i x_i$$

Dans le cas particulier où on s'intéresse à la relation entre une exposition (ici l'indice poids pour taille) et la maladie (ici le risque de décès), le modèle à plusieurs variables permet de prendre en compte un ou plusieurs facteurs de confusion éventuels (ici l'indice taille pour âge). L'estimation des paramètres se fait par la méthode du maximum de vraisemblance, c'est à dire le choix de modèle fonction des estimateurs \hat{a} et \hat{b} qui rend l'observation de l'échantillon la plus probable.

L'analyse a été réalisée sur 12316 semestres-enfants, avec 292 décès. Après itérations successives, le modèle se stabilise sur les paramètres suivants :

Prédicteurs	b	p
Poids/taille	0.60793	<0.0001
Taille/Age	0.23698	<0.0001

Il apparaît ainsi que le retard de taille comporte en lui-même, indépendamment de la maigreur qui peut l'accompagner, un risque supplémentaire non nul de décès par rapport aux enfants de taille comprise dans les limites de la référence, résultat très important, bien qu'il n'y ait pas encore d'explication satisfaisante à ce phénomène. Une analyse plus poussée par tranches d'âges a largement confirmé la validité de ce résultat.

4 Développement d'outils de prédiction : L'approche risque en épidémiologie nutritionnelle

Dans les recherches sur les malnutritions, la nécessité d'une approche analytique fondée sur un recueil individuel d'informations et autorisant une inférence statistique n'a émergé que depuis une dizaine d'années: c'est l'application du concept de risque à la nutrition du jeune enfant proposée par l'Union Internationale des Sciences Nutritionnelle. Il s'agit d'un élargissement du concept de dépistage: identifier les facteurs de risque à un stade précoce pour classer les enfants et essayer de prévenir la malnutrition; ce n'est donc pas un diagnostic. Initialement dirigée

vers la malnutrition clinique aiguë, cette approche a entraîné l'émergence progressive des méthodes épidémiologiques dans le domaine des malnutritions: hiérarchisation des facteurs de risque et signification dans la communauté par l'évaluation des *risques relatifs* et des *risques attribuables*. Il s'agit sans conteste d'un progrès majeur car permettant une analyse et une interprétation dirigées vers l'action.

La mise en oeuvre de cette approche suppose une recherche épidémiologique préalable pour sélectionner les facteurs de risque et les combinaisons de ces facteurs, et proposer une prédiction efficace.

En matière de prévention de la malnutrition la prédiction de groupes à risque pourrait permettre un changement de politique en passant d'une détection précoce fondée sur la surveillance de la croissance et limitée au secteur de la santé à des efforts de prévention primaire basés, par exemple, sur des programmes sociaux.

5 Application à la prédiction des retards de taille : L'exemple des recherches menées au Congo

5.1 Le contexte de l'étude

La recherche d'outils de prédiction des malnutritions (ici le retard de croissance en taille) chez les jeunes enfants s'inscrit dans le cadre d'un vaste programme d'épidémiologie nutritionnelle développé au Congo depuis 1986.

Les données ont été recueillies sur des échantillons aléatoires représentatifs du milieu rural congolais (2429 enfants de 0 à 5 ans) et de Brazzaville (1816 enfants). Les indices nutritionnels ont été établis selon les recommandations internationales à partir de mesures anthropométriques réalisées selon des méthodes standardisées déjà décrites par ailleurs (Cornu, Delpuech, Simondon *et al.* 1990).

La variable dépendante est la taille en fonction de l'âge des enfants de plus de 24 mois et le retard de croissance en taille est défini pour une taille en fonction de l'âge inférieure à deux écarts-type en dessous de la moyenne de référence NCHS/CDC/OMS.

A Brazzaville 12,4% (IC 95%, 9,7-15,1) des enfants de plus de 24 mois sont dans ce cas, et au Congo rural 32,2% (IC 95%, 28,5-35,9). En milieu rural comme en milieu urbain, les prévalences sont les plus faibles en dessous de un an, d'où l'intérêt d'une prédiction permettant des efforts préventifs précoces.

Les variables considérées comme potentiellement utiles pour la prédiction portent sur diverses caractéristiques bio-sociales de l'enfant, de la mère, du chef de ménage, du ménage et sur plusieurs indicateurs socio-économiques: au total 46 variables ont été considérées à Brazzaville et 47 en milieu rural.

5.2 Première étape de l'analyse : Etude de chaque variable et mesure du risque simple

Chaque variable a été croisée avec la variable à prédire (présence ou absence du retard de taille après 24 mois); la stabilité des associations a été testée par χ^2 et le risque relatif estimé par calcul des odds-ratio. En épidémiologie, le risque relatif est le rapport du taux de maladie dans le groupe exposé au facteur de risque sur le taux de maladie dans le groupe non exposé au facteur de risque. Les valeurs supérieures à 1 traduisent un risque augmenté. Dans le cas de maladies à taux faibles, l'odds-ratio ou rapport de cotes est une bonne approximation du risque

relatif. A l'instar d'autres auteurs nous l'avons utilisé ici, en dépit de prévalences assez élevées, car il peut être entouré d'un intervalle de confiance (les IC ont été calculés par la méthode de Cornfield) et le modèle logistique utilisé plus loin pour la prédiction est basé sur l'odds-ratio.

Le tableau 1 fournit un exemple des sorties de cette première étape de l'analyse. 29 variables au Congo rural et 16 à Brazzaville sont significativement associées ($p < 0,05$) au retard de taille et ont un odds-ratio avec une limite inférieure de l'intervalle de confiance supérieure à 1 (ex. dans le tableau 1, le sexe, la parenté et le poids de naissance).

Tableau 1: Retard de croissance chez les enfants de plus de 24 mois au Congo rural : distribution selon les caractéristiques de l'enfant et associations statistiques.

Facteur	n	%	retard de croissance %	P	Odds ratio non ajusté	IC 95 %
Sexe de l'enfant						
Fille	680	50,6	29,4		1	
Garçon	663	49,4	35,1	=0,02	1,30	(1,03-1,65)
Parenté avec le chef de ménage						
Enfant	1095	81,6	30,9		1	
Petit-enfant	247	18,4	38,4	=0,02	1,38	(1,02-1,85)
Personne qui s'occupe de l'enfant						
Mère	1213	90,5	32,5		1	
Autre	127	9,5	28,3	=0,34	0,82	(0,54-1,25)
Poids de naissance						
3,5 kg et +	160	23,2	21,3		1	
3,0-3,4	260	37,7	28,1		1,45	(0,89-2,37)
2,5-2,9	154	22,4	33,8		1,89	(1,11-2,23)
<2,5	115	16,7	43,5	<0,001	2,85	(1,63-5,01)

5.3 Deuxième étape de l'analyse: Mesure du risque lié à plusieurs facteurs : le modèle logistique

Pour le développement d'instruments de prédiction, la méthode utilisée a été la régression logistique. En présence de plusieurs facteurs, le risque peut s'exprimer comme une fonction de ces facteurs de risque, les risques relatifs simples se multipliant; on peut relier le risque en présence de deux ou plusieurs facteurs au risque en présence de chacun d'entre eux mais, au-delà de quelques facteurs le manque d'effectifs impose de faire appel à un modèle (Lellouch 1976). La fonction logistique correspond à cet objectif et se révèle plus adaptée que la fonction linéaire discriminante lorsque l'hypothèse de normalité de plusieurs variables n'est pas satisfaite (Rothman 1986, Engelman 1988).

Le risque ou probabilité prédite de maladie en fonction de n facteurs s'exprime par :

$$R = \frac{e^{(a + \sum_i b_i x_i)}}{(1 + e^{(a + \sum_i b_i x_i)})}$$

Sa transformation logistique donne une fonction linéaire des variables indépendantes :

$$\text{Logit}(R) = \ln[R/(1 - R)] = a + \sum_i b_i x_i \quad (\text{Rumeau-Rouquette et al. 1985})$$

a est la constante et dépend de la fréquence de la maladie; les b_i sont les coefficients de régression estimant la liaison de chaque facteur avec la maladie, à niveau constant des autres facteurs inclus dans le modèle.

Or, pour une variable indépendante binaire x dont les valeurs sont codées 0/1 (facteur absent/présent), $e^{(b_i)}$ est une estimation de l'odds ratio lié à la variation de x_i après ajustement de tous les autres facteurs pris en compte dans la régression logistique. Rothman (1986) fait remarquer qu'il s'agit là d'un avantage bien plus important que les considérations théoriques (normalité) qui ont motivé le développement de la méthode. La fonction de risque varie entre 0 et 1 et les odds-ratio se multiplient.

Enfin, il est possible d'utiliser des variables à plusieurs classes et de calculer des odds-ratio ajustés pour chaque classe à condition de comparer à une classe de référence. C'est cette option que nous avons utilisée puisque les calculs des odds-ratio en risque simple avaient été réalisés de cette manière. Ainsi, la fonction logistique peut-être utilisée pour ajuster des liaisons sur plusieurs autres variables indépendantes et comparer des odds-ratio bruts et ajustés. Pour être valide cette option exige une recodification préalable des variables en 0/1. Dans le cas d'une variable à plus de deux classes, la catégorie de référence est toujours codée 0 et les autres catégories 0/1 selon la présence ou l'absence de la condition.

Les coefficients étant affectés de leur erreur-type, nous avons pu calculer l'intervalle de confiance à 95% de chaque odds-ratio ajusté.

Nous avons estimé les paramètres des fonctions logistiques par la méthode du maximum de vraisemblance qui ne suppose aucune hypothèse de distribution. L'adéquation au modèle logistique a été appréciée par le test de "CC Brown", une valeur faible de p indiquant une mauvaise adéquation; en complément, les effectifs prédits et observés sont comparés par le test de "Hosmer-Lemeshow", une faible valeur de p signalant également une mauvaise adéquation (Engelman 1988).

Pour la prédiction, nous avons utilisé des régressions logistiques pas à pas; elles sélectionnent les variables avec $p < 0,05$, qui sont alors prises en compte dans le modèle. Ce type d'analyse reconnaît les facteurs de risques associés qui fournissent une information redondante: à chaque pas, une variable entre dans le modèle; les variables qui sont déjà dans le modèle et dont p est $> 0,05$ (seuil que nous avons fixé, au même niveau que celui d'inclusion) sortent du modèle. Le processus se poursuit jusqu'à ce qu'aucune variable ne remplisse les conditions d'entrée et de sortie dans le modèle.

A titre d'exemple, pour le Congo rural 11 des 28 variables initiales sont finalement entrées dans le modèle logistique du retard de taille (Tableau 2). L'adéquation est bonne (CC Brown $p=0,65$ - Hosmer-Lemeshow $p=0,64$) et le risque varie de 0,0205 à 0,908. Les variables liées à la mère (taille, activité agricole et situation matrimoniale) entrent en tête dans le modèle.

Tableau 2: Retard de croissance au Congo rural : variables incluses dans le modèle logistique et estimation des odds-ratio correspondants

Variable	Odds-ratio ajusté	IC à 95%
1. Taille de la mère		
1,65 m et +	1	
1,60-1,64	2,10	1,13-3,90
1,55-1,59	2,78	1,55-4,98
1,50-1,54	4,54	2,48-8,31
< 1,50	6,03	3,10-11,75
2. Mère cultivatrice de Manioc		
Non	1	
Oui	2,19	1,55-3,08
3. Situation matrimoniale de la mère		
Mariage reconnu	1	
Union libre	2,12	1,41-3,18
Célib./Veuve/Divorcée	1,84	1,31-2,57
4. Murs		
En parpaings	1	
En terre	1,07	0,63-1,80
En torchis	0,78	0,44-1,38
En bois	1,87	1,03-3,39
5. Origine de l'eau		
Robinet/Pompe	1	
Puits	2,28	1,34-3,88
Source	1,42	0,88-2,30
Pluie/Rivière	1,24	0,73-2,11
6. Activité du chef de ménage		
Cadre	1	
Employé/Ouvrier	1,62	0,92-2,87
Commerçant	1,24	0,65-2,37
Agri./Chasseur/Pêch.	1,61	0,93-2,77
Retraité	1,20	0,54-2,69
Sans profession	3,07	1,56-6,03
7. Sexe de l'enfant		
Fille	1	
Garçon	1,48	1,14-1,92
8. Sols		
En ciment/carrelage	1	
En terre battue	1,69	1,14-2,51
9. Taille de la famille		
5-9 personnes	1	
4 et -	1,12	0,78-1,61
10-14	1,21	0,86-1,70
15 et +	2,52	1,44-4,43
10. Mère cultivatrice de légumes		
Non	1	
Oui	1,45	1,05-1,99
11. Poids de naissance		
3,5 kg et +	1	
3,0-3,4	1,31	0,77-2,21
2,5-2,9	1,58	0,90-2,76
< 2,5	2,47	1,36-4,47

Terme constant a= -4,285

Initialement cette méthode a été développée dans les études de mortalité et des facteurs de risque des maladies ischémiques. Les applications aux domaines de la mesure des facteurs de risque nutritionnel, de l'évaluation des interventions et de la prédiction des malnutritions sont très récentes.

5.4 Troisième étape de l'analyse: Etude des qualités discriminantes et prédictives des fonctions logistiques de risque:

- **Sensibilité, Spécificité, Valeurs prédictives**

Les capacités de discrimination des différents seuils de classification (seuils de risque des fonctions logistiques) ont été appréciées classiquement en utilisant les critères suivants: La *sensibilité* est la proportion d'enfants retardés en taille correctement classés à risque par le test; c'est la capacité du test à identifier les retards de taille. La *spécificité* est la proportion d'enfants non retardés en taille correctement classés à faible risque; c'est la capacité du test à correctement identifier les enfants qui n'ont pas de retard de taille.

- **Les fonctions de risque et les courbes ROC**

En pratique, un grand nombre d'indicateurs peuvent avoir plusieurs seuils ou même être des variables continues avec des possibilités de seuils différents très grandes (ex: une taille, un poids de naissance et, ici, une fonction de risque telle que celles fournies par les modèles logistiques). L'étude des relations Sensibilité/Spécificité a été conduite grâce à l'établissement et l'analyse de courbes ROC (Relative Operating Characteristic) qui sont des courbes caractéristiques de la performance d'un test: les probabilités de vraie positivité sont portées en fonction des probabilités de fausse positivité et tracées sous forme de courbes. Initialement l'analyse ROC a été développée dans le domaine des images radar et de l'imagerie médicale puis dans celui des tests psychométriques (Swets 1979). Les propositions d'application à l'épidémiologie sont assez récentes (Erdreich et Lee 1981). Elle constitue une méthode de choix pour faire une analyse synthétique à tous les degrés de sensibilité et spécificité. En raison de cette propriété, l'analyse ROC a été proposée pour comparer plusieurs modèles logistiques de prédiction et rechercher des divergences (Moise et al 1986), ce qui est, en l'occurrence, un des objectifs principaux de notre travail. Erdreich et Lee (1981) indiquent notamment qu'elle peut aider, en complément de la régression logistique, à l'identification de facteurs de risque et à la sélection de critères de décision.

La Figure 4 illustre l'intérêt de l'utilisation des courbes ROC: la prédiction est d'autant meilleure que la courbe s'éloigne de la diagonale, appelée encore ligne de chance, puisqu'elle représente en tous points la même proportion de positifs chez les "malades" que chez les "non malades".

La partie gauche des courbes est une zone de haute spécificité et la partie droite une zone de haute sensibilité. Pour les zones de forte spécificité, la prédiction est du même ordre à Brazzaville et au Congo rural. Au dessous d'une spécificité de 80%, la prédiction est toujours meilleure à Brazzaville. Il est également facile de préciser pour chaque fonction celle qui donne la probabilité la plus élevée d'une classification correcte (maximum de la somme sensibilité+spécificité): à Brazzaville ce maximum est atteint pour une sen-

sibilité de 84% et une spécificité de 64% contre respectivement 63% et 74% en milieu rural.

La figure 5 montre que les performances du modèle de prédiction du retard de taille sévère (taille en fonction de l'âge inférieure à -3 écarts-types) sont du même ordre à Brazzaville mais bien plus faibles en milieu rural. On peut ainsi comparer les performances de différents instruments de prédiction.

6 Les applications en Santé Publique

La construction de modèles logistiques établissant les risques suppose une recherche préalable utilisant matériels informatiques et logiciels statistiques adaptés. Mais il est ensuite facile de transformer le risque en un système de score à points, les points étant proportionnels aux coefficients de régression des variables dans le modèle. L'application par un système de santé ou par un programme social peut être ensuite assez simple.

7 Les recherches à entreprendre

Enfin dans le cadre de "cette approche risque" deux voies de recherche principales sont ouvertes:

- d'une part l'utilisation d'autres méthodes qui pourraient s'avérer plus performantes que les modèles logistiques, notamment les procédures CART (Classification et arbres de régression)
- d'autre part la validation des instruments de prédiction dans le cadre d'interventions: dans quelle mesure permettent-ils une meilleure prévention ?

Références bibliographiques

- Bouyer J. La régression logistique en épidémiologie; partie I. Rev Epidém. et Santé publique 1991; 39:79-87.
- Cornu A, Delpuech F, Simondon F, Tchibindat F, Olivola D, et al. Enquête nationale sur l'état nutritionnel des enfants d'âge préscolaire au Congo. Coll Etudes et Theses. Paris, ORSTOM, 1990.
- Engelman L. Stepwise logistic regression. in : DIXON WJ-BMDP Statistical Software vol 2. Los Angeles-London :UCLA Press, 1988; 941-69.
- Garenne M, Maire B, Fontaine O, Dieng K, Briend A. Risques de décès associés à différents états nutritionnels chez l'enfant d'âge préscolaire. Orstom, Dakar, 1987.

- Erdreich LS, Lee ET. Use of relative operating characteristic analysis in epidemiology. *Am J of Epidem* 1981; 1145:649-62.
- Lellouch J. Le risque : définitions et procédés de calcul. *Rev Epidem et Santé Publ* 1976;24:201-10.
- Moise A, Salamon R, Commenges D, Clément B. L'utilisation des courbes ROC, The use of receiver operating characteristic (ROC) curves. *Rev Epidem et Santé Publ* 1986;34:209-17.
- Rothman KJ. Multivariate analysis. chap 14. In: KJ Rothman ed. *Modern epidemiology*. Boston/Toronto : Little Brown and Cie, 1981:285-310.
- Rumeau-Rouquette C, Breart G, Padieu R. *Méthodes en épidémiologie*. 3e ed. Paris, Flammarion Médecines-Sciences, 1985.
- Swets JA. ROC analysis applied to evaluation of medical imaging techniques. *Investigative Radiology* 1979;14:109-21.
- Who Working Group. Use and interpretation of anthropometric indicators of nutritional status. *Bull WHO* 1986;64(6):929-941.

Figure 1: Fréquence d'individus en dessous de chaque seuil (en écart type) pour une distribution normale

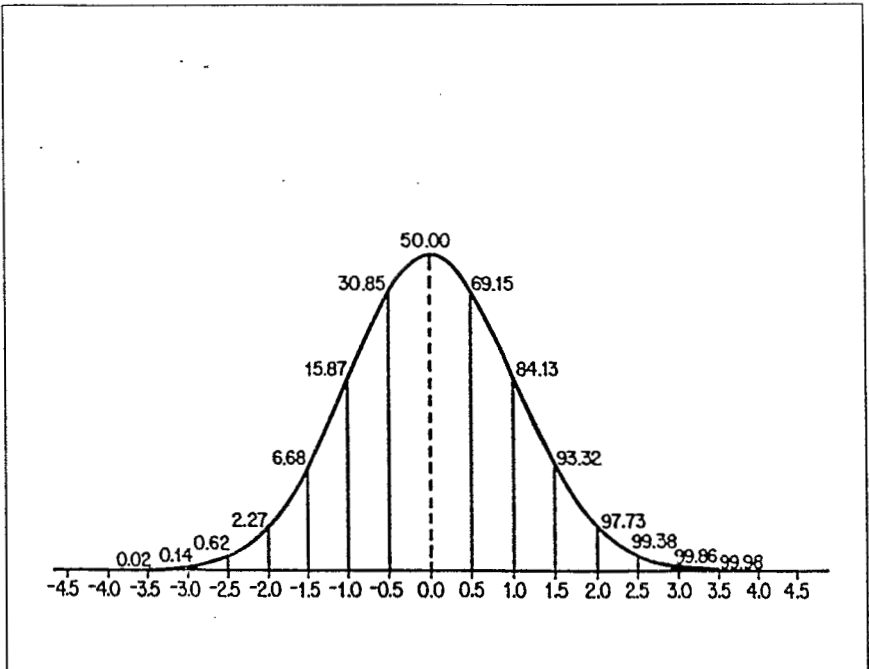


Figure 2: Prévalence normalisée

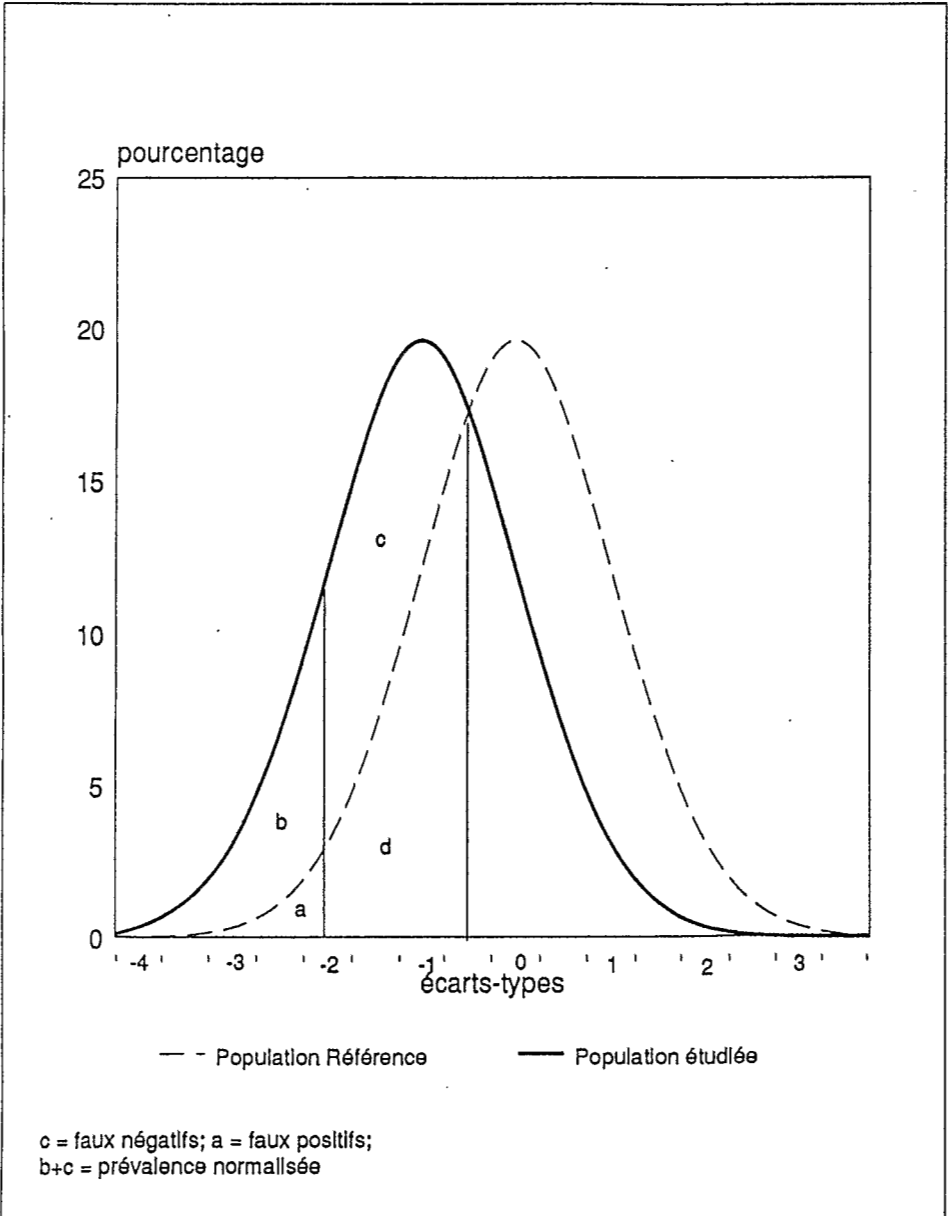


Figure 3: Risque relatif de décès (sur une période de 6 mois)

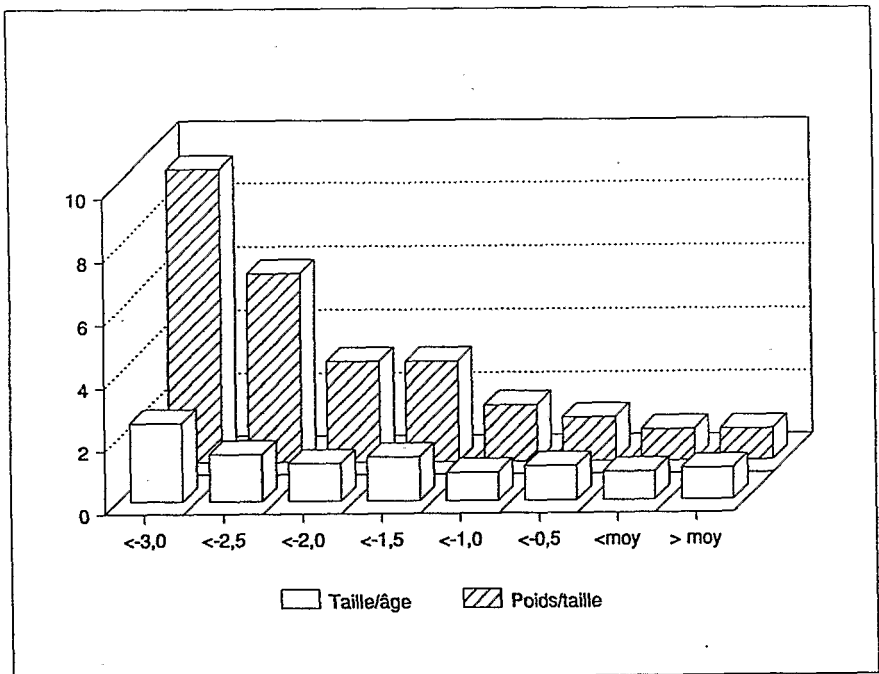


Figure 4: Courbe ROC pour 2 équations de régression logistique : retard de croissance à Brazzaville et au Congo rural

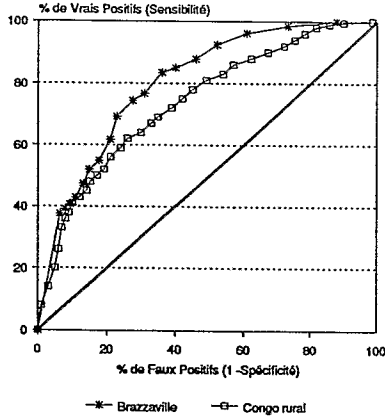


Figure 5: Courbe ROC pour 2 équations de régression logistique : retard de croissance sévère à Brazzaville et au Congo rural

