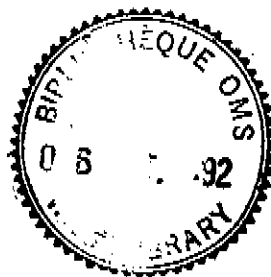


Distribution : générale
WHO/CDD/EDP/91.3
ORIGINAL : ANGLAIS

PROGRAMME DE LUTTE CONTRE LES MALADIES DIARRHEIQUES

**UNE SOURCE DE BIAIS DANS LES ETUDES SUR LES
ASSOCIATIONS BILATERALES : HYPOTHESE DE LA
MALNUTRITION ET DE LA DIARRHEE**

Simon N. Cousens



ORGANISATION MONDIALE DE LA SANTE

**UNE SOURCE DE BIAIS DANS LES ETUDES SUR LES
ASSOCIATIONS BIDIRECTIONNELLES : HYPOTHESE DE LA
MALNUTRITION ET DE LA DIARRHEE**

Simon N. Cousens¹

**¹ Section d'Epidémiologie Maternelle et Infantile
Département d'Epidémiologie et de Démographie
Ecole d'Hygiène et de Médecine Tropicale de Londres
Keppel Street, Londres WC1E 7HT, Royaume-Uni**

SOMMAIRE

	<u>Pages</u>
RESUME	2
INTRODUCTION	3
L'ASSOCIATION ENTRE LA DIARRHEE ET LA MALNUTRITION	3
LE PROBLEME DE LA CAUSALITE INVERSE	4
LE MODELE	4
DISCUSSION	11
REMERCIEMENTS	12
BIBLIOGRAPHIE	13

RESUME

On examine une source possible de biais dans les études de l'association entre deux variables A et B, qui peuvent être reliées entre elles par un lien causal réciproque ou bilatéral (A peut faire augmenter le risque de B et B peut faire augmenter le risque de A). A cet effet, un modèle probabilité simple qui porte en particulier sur l'association entre la malnutrition et la diarrhée chez le jeune enfant dans les pays en développement a été mis au point. On part de l'hypothèse que la diarrhée (A) contribue à un état nutritionnel médiocre (B) qui à son tour conduit à une aggravation de la diarrhée. Pour étudier l'influence d'un état nutritionnel médiocre sur l'incidence de la diarrhée on a comparé les taux de diarrhée chez les enfants qui sont sous-alimentés et chez ceux qui le sont convenablement. Le modèle adopté montre qu'en présence de taux de diarrhée hétérogènes, ces analyses peuvent donner des résultats biaisés, et, partant, conduire à surestimer l'importance d'un lien causal entre la malnutrition (B) et la diarrhée (A). Enfin, les résultats d'une simulation par ordinateur visant à évaluer quelle peut être l'ampleur de ce biais sont présentés.

INTRODUCTION

Pour étudier si un facteur A augmente le risque de survenue d'une affection B, les épidémiologistes procèdent souvent comme suit :

1. on choisit un groupe de sujets souffrant de A et un autre groupe qui n'en est pas atteint;
2. on suit ces deux groupes pendant un certain temps afin d'observer qui contracte B;
3. on compare le taux de B chez les sujets atteints de A et chez ceux qui n'en souffrent pas;
4. si l'on observe un taux plus élevé de B chez les sujets atteints de A (après avoir pris en compte les facteurs de confusion), on en déduit que A "cause" B.

Cette méthode est connue sous le nom d'étude de cohorte ou de suivi et l'un de ses avantages reconnus est que la suite d'événements entre l'exposition (A) et le résultat (B) peut être plus clairement établie (Hennekens & Buring, 1987). Cependant, déduire qu'"A cause B" peut poser problème. Rose & Shipley (1980), en parlant d'un lien apparent entre une faible cholestérolémie et un risque accru de cancer, soulignent que l'on peut faire cette constatation lorsqu'une affection préclinique entraîne à son début une faible cholestérolémie, c'est-à-dire qu'un état B non diagnostiqué peut causer un état A, lequel entraîne l'observation de B. On déduit que A cause B alors qu'en réalité c'est l'inverse qui est vrai.

Cette difficulté à distinguer entre cause et effet est particulièrement grande lorsqu'on essaie de déterminer le sens et l'importance d'un lien causal entre A et B alors que l'association peut être bilatérale. Bien souvent, le fait qu'A cause B conduira à masquer la question de savoir si B cause A, effet qui est souvent négligé. Dans le présent document est exposé un modèle probabilité simple qui permet de mettre en évidence cet effet de confusion. L'association entre les maladies diarrhéiques et la malnutrition chez le jeune enfant dans les pays en développement peut servir d'exemple. Le modèle montre que, dans l'hypothèse où la diarrhée (A) cause la malnutrition (B) mais où la malnutrition (B) ne cause pas la diarrhée (A), et où les taux de diarrhée varient selon les enfants, les techniques épidémiologiques classiques peuvent conduire à "observer" que la malnutrition (B) cause la diarrhée (A).

L'ASSOCIATION ENTRE LA DIARRHÉE ET LA MALNUTRITION

Dans les pays en développement, l'association entre un état nutritionnel médiocre déterminé par l'anthropométrie et la diarrhée chez le jeune enfant est bien établie (Tomkins & Watson, 1989). La nature exacte de cette association, cependant, reste assez obscure. On a émis l'idée que la relation était bilatérale, à savoir que les épisodes diarrhéiques ont une incidence néfaste sur l'état nutritionnel d'un enfant alors que les enfants dont l'état nutritionnel est médiocre risquent davantage de souffrir de diarrhée (Scrimshaw et al., 1968). On a suggéré des mécanismes biologiquement plausibles par lesquels la diarrhée pourrait conduire à la malnutrition, et un état nutritionnel entraîner un accroissement des taux de diarrhée (Chen, 1983; Chandra, 1986; Sirisinha et al., 1975).

Dans un certain nombre d'études épidémiologiques, on tente d'étudier l'association entre la diarrhée et la malnutrition. (Pour une bibliographie récente, se reporter à Tomkins & Watson, 1989.) Traditionnellement, ces études consistent à suivre un groupe d'enfants pendant un certain temps, à surveiller la morbidité et à faire des mesures anthropométriques à intervalles réguliers. Ensuite, afin d'examiner si la diarrhée

entraîne la malnutrition, on compare pendant la même période le taux de croissance à la morbidité par diarrhée. Les auteurs qui ont observé une association entre une augmentation des épisodes diarrhéiques et une diminution du rythme de prise de poids interprètent leurs résultats comme étant une preuve que la diarrhée est à l'origine de la malnutrition. Pour savoir si la malnutrition provoque la diarrhée, on classe les enfants selon qu'ils sont bien nourris ou pas d'après leurs données anthropométriques au début du suivi. Les deux groupes d'enfants sont ensuite comparés afin de repérer ceux qui ont souffert de diarrhée dans l'intervalle suivant. L'observation de taux de diarrhée plus élevés chez les enfants malnutris serait une preuve que la malnutrition est à l'origine de la diarrhée.

LE PROBLEME DE LA CAUSALITE INVERSE

Un certain nombre de ces études sont critiquables à plusieurs titres. D'abord, elles ne prennent pas en compte les variables parasites. Ensuite, on morcèle la période de suivi en plusieurs sous-intervalles de sorte que chaque enfant fait l'objet de plusieurs observations "indépendantes" aux fins de l'analyse. Enfin, on omet d'examiner l'effet parasite potentiel d'un lien causal agissant en sens opposé. C'est au dernier de ces problèmes que l'on s'intéresse dans le présent document. Le problème général est le suivant :

"Le fait qu'A cause B peut-il entraîner une association qui donne à penser que B cause A (ou qui surestime l'importance du lien causal entre B et A) ?"

Pour éviter de se perdre en abstractions, nous examinerons la question plus spécifique suivante :

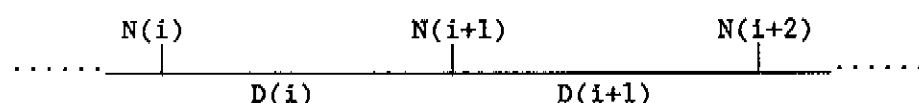
"Le fait que la diarrhée (A) cause la malnutrition (B) dans des circonstances particulières entraîne-t-il une association qui donne à penser que la malnutrition (B) cause la diarrhée (A) alors que ce n'est pas le cas ?"

Pour étudier cette question, nous avons établi un modèle probabilité simple.

LE MODELE

Pour élaborer ce modèle, nous avons suivi la méthode classique utilisée pour ce genre d'étude. Dans ces études, les enfants sont pesés et mesurés au début de la période d'observation. Chaque enfant est ensuite classé dans la catégorie "bien nourris" ou "malnutris" selon que son bilan anthropométrique se situe au-dessus ou au-dessous d'une valeur limite. Les enfants sont ensuite suivis pendant un certain temps et les cas de diarrhée enregistrés. Pour savoir si la malnutrition est à l'origine de la diarrhée, on compare les taux de diarrhée chez les enfants malnutris et chez ceux qui sont bien nourris. Les enfants pourront être à nouveau pesés et mesurés périodiquement tout au long de l'étude, chaque nouvelle pesée étant le point de départ d'une nouvelle période d'observation.

Prenons un enfant soumis à l'étude. On peut représenter l'évolution de son état de la manière suivante :



où $N(i)$ correspond au bilan anthropométrique de l'enfant au début de la i ème période d'observation et où $D(i)$ indique si l'enfant a souffert ou non d'un épisode diarrhéique pendant le i ème intervalle entre deux périodes d'observation. Pour simplifier, on admet qu'un enfant ne souffre jamais de plus d'un épisode de diarrhée pendant une quelconque période d'observation.

Posons $N(i) = 1$ dans le cas où l'enfant est malnutri au début du ième intervalle,
 $N(i) = 0$ dans les autres cas,
 $D(i) = 1$ dans le cas où l'enfant souffre de diarrhée durant le ième intervalle,
 $D(i) = 0$ dans les autres cas,

et supposons que :

- le bilan anthropométrique (B) de l'enfant au début du ième intervalle n'influe pas sur la probabilité qu'il souffre d'un épisode diarrhéique (A) pendant cet intervalle (autrement dit, B ne cause pas A);
- p corresponde à la probabilité que l'enfant souffre d'un épisode diarrhéique pendant un quelconque intervalle, qu'il ait ou non été atteint de diarrhée durant les précédents intervalles,
 c'est-à-dire : $P\{D(i) = 1 | p, n(i-1), d(i-1), d(i-2), \dots\} = p$;
- l'état nutritionnel de l'enfant à la fin du ième intervalle (début du (i + 1)ème intervalle, soit $N(i+1)$) dépende seulement de son état au début de ce même intervalle ($N(i)$) et du fait que l'enfant souffre ou non d'un épisode diarrhéique pendant le ième intervalle ($D(i)$), (autrement dit, A "cause" B).

Puis posons que : $P\{N(i+1)=1 | N(i)=1, D(i)=1\} = t_{11}$,

$$P\{N(i+1)=1 | N(i)=1, D(i)=0\} = t_{10},$$

$$P\{N(i+1)=1 | N(i)=0, D(i)=1\} = t_{01},$$

$$P\{N(i+1)=1 | N(i)=0, D(i)=0\} = t_{00}.$$

En termes mathématiques, on peut exprimer comme suit le fait qu'un état de malnutrition et une diarrhée à l'instant i accroît le risque de malnutrition à l'instant i+1 :

$$t_{11} > t_{10} \quad \text{et} \quad t_{01} > t_{00} \quad (1)$$

$$\text{et} \quad t_{11} > t_{01} \quad \text{et} \quad t_{10} > t_{00} \quad (2)$$

Les deux premières inégalités (1) traduisent le fait que la diarrhée prédispose à la malnutrition. Les deux suivantes indiquent qu'un enfant qui était déjà malnutri risque davantage de l'être à l'instant d'observation suivant qu'un enfant qui était bien nourri auparavant.

En résumé, on peut décrire ce modèle comme suit : la diarrhée prédispose à un état nutritionnel (anthropométrique) médiocre, mais un état nutritionnel (anthropométrique) médiocre ne prédispose pas à la diarrhée.

Conformément à ce modèle, l'état nutritionnel de l'enfant pendant la durée de l'étude ($N(1), N(2), \dots$) peut être considéré comme un processus de Markov puisque la distribution de $N(i+1), N(i)$ et p étant donnés, est indépendante de $N(i-1), N(i-2), \dots$. On peut représenter la matrice de transition de ce processus comme suit :

		N(i+1)	
		0	1
N(i)	0	(1-U)	U
	1	V	(1-V)

où $U = pt_{01} + (1-p)t_{00}$,

et $V = p(1-t_{11}) + (1-p)(1-t_{10})$.

Ici, U correspond à la probabilité que l'enfant, durant un quelconque intervalle, passe d'un état de bonne nutrition à celui de malnutrition. V correspond à la probabilité de voir le contraire se produire pendant un quelconque intervalle, l'enfant passant d'un état de malnutrition à un état d'alimentation correcte. On remarquera que les deux inégalités (1) impliquent que U est une fonction monotone croissante et V une fonction monotone décroissante de p. C'est-à-dire qu'à mesure que p augmente, U augmente et V diminue. Ensuite, selon un résultat bien connu (Cox & Miller, 1972), à la longue, la probabilité pour un enfant d'être malnutri au début de tout intervalle s'exprime comme suit :

$$R(p) = \frac{U}{U+V} .$$

Intuitivement, ce résultat indique que la probabilité à longue échéance pour un enfant d'être malnutri à un moment donné est proportionnelle à la probabilité de devenir malnutri pendant tout intervalle. De même, la probabilité qu'il soit bien nourri à un moment donné est proportionnelle à celle qu'il devienne bien nourri. La somme des probabilités est égale à 1 puisque l'enfant est soit bien nourri soit malnutri.

Puisque U est une fonction monotone croissante et V une fonction monotone décroissante de p, R(p) est une fonction monotone croissante de p. Autrement dit, à mesure que le risque de diarrhée ou son incidence (p) chez l'enfant augmente, la probabilité qu'il ou elle soit malnutri(e) à tout moment donné augmente également. C'est ce à quoi l'on s'attendrait intuitivement puisque, en définissant le modèle, nous avons spécifié qu'un épisode de diarrhée prédispose à la malnutrition.

A présent, examinons une population d'enfants chez qui le taux d'incidence de la diarrhée varie, c'est-à-dire qu'à différents enfants correspondent différentes valeurs de p. Désignons par la variable P aléatoire la probabilité que la diarrhée survienne pendant un intervalle donné et par f(P) la fonction de densité de la probabilité (pdf) de P. Dans les études où l'on étudie si la malnutrition prédispose à la diarrhée, il est courant d'estimer le risque relatif de diarrhée associé à une malnutrition préalable en comparant le risque de diarrhée chez les enfants malnutris au début de la période d'observation à ce même risque chez ceux qui ne le sont pas. Dans ces conditions, la valeur attendue de cet estimateur du risque relatif est :

$$E[RR] = \frac{P(D(i)=1 \mid N(i)=1)}{P(D(i)=1 \mid N(i)=0)}$$

$$\text{On a alors, } P(D(i)=1 \mid N(i)=1) = \frac{P(D(i)=1, N(i)=1)}{P(N(i)=1)}$$

$$\text{avec } P(D(i)=1, N(i)=1) = E[D(i) \times N(i)]$$

$$\begin{aligned}
 & - E_p[E[D \times N | P]] \\
 & - \int pR(p)f(p)dp \\
 & - E[P \times R(P)], \\
 \text{et} \quad P(N(i)=1) & - E[R(P)]. \\
 \text{De même, } P(D(i)=1 | N(i)=0) & - \frac{E[P \times (1-R(P))]}{1-E[R(P)]} \\
 \text{donc } E[RR] & - \frac{E[P \times R(P)] \times (1-E[R(P)])}{E[R(P)] \times E[P \times (1-R(P))]}
 \end{aligned}$$

Mais $R(p)$ est une fonction monotone croissante de p et, par conséquent, la covariance de $R(P)$ et P est supérieure à zéro. Cela implique que :

$$\begin{aligned}
 E[P \times R(P)] & > E[P] \times E[R(P)] \\
 \implies E[P \times R(P)] - E[P \times R(P)] \times E[R(P)] & > E[P] \times E[R(P)] - E[P \times R(P)] \times E[R(P)] \\
 \implies E[P \times R(P)] \times (1-E[R(P)]) & > E[R(P)] \times E[P \times (1-R(P))] \\
 \implies E[RR] & > 1
 \end{aligned}$$

Par conséquent, dans ce modèle où la malnutrition n'entraîne pas la diarrhée, une comparaison des taux de diarrhée chez les enfants malnutris et chez ceux qui sont convenablement nourris donne à penser que le risque ou le taux de diarrhée est supérieur chez les sujets malnutris. C'est en effet le cas, puisque la diarrhée augmente le risque de malnutrition et, par conséquent, les enfants qui présentent les plus forts taux de diarrhée sont les plus exposés au risque de malnutrition. Toutefois, comme on commence par mesurer l'état nutritionnel de chaque enfant au début de la période d'observation pour ensuite enregistrer l'incidence de la diarrhée (ainsi, la malnutrition paraît précéder la diarrhée), ces résultats peuvent être interprétés à tort dans le sens d'une malnutrition aboutissant à la diarrhée. De même, si certains enfants ont tendance à souffrir d'épisodes diarrhéiques prolongés ou aggravés qui augmentent le risque de malnutrition qu'ils courent, on pourra observer un lien entre la malnutrition et l'aggravation ou la prolongation de l'épisode diarrhéique en l'absence de toute relation de cause et effet. Il convient de noter que ce "biais" provient de la causalité inverse (la diarrhée cause la malnutrition) et de l'hétérogénéité des taux de diarrhée enregistrés dans la communauté. En l'absence d'hétérogénéité ($P(P=p)=1$), $E[P \times R(P)] = E[P] \times E[R(P)]$ et $E[RR] = 1$, il n'y a pas de biais.

Deux études menées récemment au Soudan (El Samani et al., 1988) et au Mexique (Sepulveda et al., 1988) ont consisté à faire des analyses qui prennent en compte l'épisode diarrhéique qui a précédé la période d'observation. Pour ces analyses, les périodes d'observation ont été divisées en deux catégories : celles qui ont été précédées d'un épisode diarrhéique, et celles qui ne l'ont pas été. On a ensuite refait des analyses similaires à celles que l'on vient d'exposer, mais en stratifiant les données en fonction des antécédents de diarrhée pendant la période antérieure. La technique utilisée par ces auteurs prête sans doute le flanc à la critique puisque les données concernant chaque enfant figurent plusieurs fois dans les analyses où elles sont considérées comme les résultats d'observations indépendantes, mais on peut penser qu'elle tend à séparer les enfants qui présentent des taux de diarrhée élevés des autres, et donc à réduire l'hétérogénéité de chaque sous-groupe. Par conséquent, il est intéressant d'examiner l'efficacité de cette stratégie qui permet de neutraliser le biais dû à la causalité inverse et à l'hétérogénéité des taux de diarrhée dans la population.

Périodes d'observation précédées d'un épisode diarrhéique

Désignons par $RR_{p=1}$ le risque relatif pour les périodes précédées par un épisode diarrhéique.

$$\begin{aligned}
 \text{On a : } E[RR_{p=1}] &= \frac{P(D(i)=1 \mid N(i)=1, D(i-1)=1)}{P(D(i)=1 \mid N(i)=0, D(i-1)=1)} \\
 &= \frac{P(D(i)=1, N(i)=1, D(i-1)=1)}{P(D(i)=1, N(i)=0, D(i-1)=1)} \times \frac{[1 - P(N(i)=1, D(i-1)=1)]}{P(N(i)=1, D(i-1)=1)} \\
 P(D(i)=1, N(i)=1, D(i-1)=1) &= \int p^2 (R(p) \times t_{11} + (1-R(p)) \times t_{01}) \times f(p) dp \\
 &= t_{01} E[P^2] + (t_{11} - t_{01}) E[P^2 R(P)] \\
 &= r \text{ (disons)} \\
 P(D(i)=1, N(i)=0, D(i-1)=1) &= E[P^2] - r \\
 \text{et } P(N(i)=1, D(i-1)=1) &= t_{01} E[P] + (t_{11} - t_{01}) E[R(P)] \\
 &= s \text{ (disons)} \\
 \text{Donc, } E[RR_{p=1}] &= \frac{r}{E[P^2] - r} \times \frac{(EP - s)}{s}
 \end{aligned}$$

Périodes d'observation non précédées d'un épisode diarrhéique

Désignons par $RR_{p=0}$ le risque relatif pour les périodes qui n'ont pas été précédées d'un épisode diarrhéique. On a :

$$\begin{aligned}
 E[RR_{p=0}] &= \frac{P(D)(i)=1 \mid N(i)=1, D(i-1)=0)}{P(D)(i)=1 \mid N(i)=0, D(i-1)=0)} \\
 &= \frac{P(D)(i)=1, N(i)=1, D(i-1)=0)}{P(D)(i)=1, N(i)=0, D(i-1)=0)} \times \frac{[1 - P(N(i)=1, D(i-1)=0)]}{P(N(i)=1, D(i-1)=0)} \\
 P(D(i)=1, N(i)=1, D(i-1)=0) &= \int p(1-p) (R(p) \times t_{10} + (1-R(p)) \times t_{00}) \times f(p) dp \\
 &= t_{00} E[P(1-P)] + (t_{10} - t_{00}) E[P(1-P)R(P)] \\
 &= r \text{ (disons)} \\
 P(D(i)=1, N(i)=0, D(i-1)=0) &= E[P(1-P)] - r \\
 \text{et } P(N(i)=1, D(i-1)=0) &= t_{00} E[1-P] + (t_{10} - t_{00}) E[(1-P)R(P)] \\
 &= s \text{ (disons)} \\
 \text{Donc, } E[RR_{p=0}] &= \frac{r}{E[P(1-P)] - r} \times \frac{(1 - EP - s)}{s}
 \end{aligned}$$

On peut montrer par le calcul que chacune des prévisions ci-dessus est supérieure à 1. Par conséquent, stratifier les périodes d'observation en fonction de la survenue d'un épisode diarrhéique pendant la période préalable n'élimine pas le problème du biais imputable à des taux de diarrhée hétérogènes.

Exemple

Pour illustrer le problème de la causalité inverse en cas de taux de diarrhée hétérogènes et se faire une idée de l'ampleur du biais qui en résulte, faisons une application numérique. On a choisi des paramètres qui paraissent "raisonnables", mais il faut souligner qu'ils ne proviennent pas de données concrètes.

Commençons par exprimer la probabilité qu'un enfant est malnutri au début d'une période d'observation, compte tenu de l'état de l'enfant au début de la période précédente et du fait qu'il a ou non souffert de diarrhée dans cet intervalle :

$$t_{11} = 0,99, \quad t_{10} = 0,7, \quad t_{01} = 0,3, \quad t_{00} = 0,05.$$

Puis, adoptons la distribution présentée au tableau 1 du risque de diarrhée parmi la population pendant une seule période d'observation. On a choisi cette distribution pour disposer d'un éventail assez large de taux de diarrhée dans la population. Pour un enfant présentant un risque donné de diarrhée p , le risque de malnutrition au début de tout intervalle $[R(p)]$ peut alors être calculé (voir tableau 1). Dans cette distribution, les valeurs probables de P , de R et de $P \times R$ sont les suivantes :

$$\begin{aligned} E[P] &= 0,40, \\ E[R(P)] &= 0,45, \\ E[P \times R(P)] &= 0,21. \end{aligned}$$

TABLEAU 1. Distribution parmi la population du risque de diarrhée pendant une seule période d'observation et probabilité de malnutrition

Risque de diarrhée p	Proportion de la population $f(p)$	Probabilité de malnutrition $R(p)$
0,1	0,10	0,22
0,2	0,15	0,29
0,3	0,15	0,37
0,4	0,20	0,45
0,5	0,15	0,53
0,6	0,15	0,61
0,7	0,10	0,70
0,8	0,0	0,79
0,9	0,0	0,88

Dans cet exemple on remarquera que :

- (i) en moyenne, 40 % des enfants souffrent de diarrhée pendant une période d'observation;
- (ii) en moyenne, 45 % des enfants sont malnutris au début d'une période d'observation;

- (iii) l'état nutritionnel d'un enfant qui est malnutri au début d'une période d'observation et qui souffre d'un épisode diarrhéique pendant cette période a peu de chances de s'améliorer ($p=0,01$);
- (iv) un enfant bien nourri au début de la période d'observation et qui ne souffre pas d'un épisode diarrhéique risque peu d'être atteint de malnutrition ($p=0,05$);
- (v) un enfant malnutri qui ne souffre pas d'un épisode diarrhéique a 30 % de chances de voir son état nutritionnel s'améliorer;
- (vi) un enfant bien nourri qui souffre d'un épisode diarrhéique a 30 % de chances d'être atteint de malnutrition;
- (vii) la probabilité de malnutrition ($R(p)$) s'accroît à mesure qu'augmente p ;
- (viii) un état nutritionnel médiocre n'augmente pas le risque de diarrhée, qui ne dépend que de p .

Il est intéressant de remarquer que dans l'étude faite au Soudan (El Samani et al., 1988), un épisode diarrhéique s'est produit dans 36 % des périodes d'observation des enfants, alors que dans l'étude réalisée au Mexique (Sepulveda et al., 1988), la proportion était d'environ 53 %. Ces pourcentages sont comparables à celui de 40 % ($E(P)$) choisi pour le modèle. En outre, les taux globaux de prévalence de la malnutrition légère et de la malnutrition modérée dans les deux études étaient les suivants : Soudan, 66 %, Mexique, 36 %. Dans le modèle, la prévalence attendue de la malnutrition est de 45 % ($E(R(P))$), et donc située entre ces deux chiffres. Si l'on procède à une analyse approximative d'une étude menée dans ce type de contexte, sans tenir compte d'une éventuelle diarrhée survenue pendant la période ayant précédé l'analyse, alors la prévision de l'estimation du risque relatif que l'on obtiendra s'exprime comme suit :

$$E[RR] = \frac{0,21 \times (1-0,45)}{0,45 \times (0,40-0,21)}$$

$$= 1,30.$$

Ainsi, dans cette situation où la diarrhée prédispose à la malnutrition mais non l'inverse, une analyse "classique" des données peut donner l'impression que la malnutrition fait augmenter de 30 % le risque de diarrhée chez l'enfant.

Si les données sont classées en deux groupes, à savoir les périodes précédées par un épisode diarrhéique, et celles qui ne le sont pas, alors on peut exprimer comme suit les valeurs probables obtenues par une analyse de données stratifiées :

$$E[RR_{p=1}] = 1,15$$

et $E[RR_{p=0}] = 1,22.$

Le fait de prendre en compte des antécédents de diarrhée dans l'intervalle qui précède la période d'observation a donc réduit, mais non diminué, le biais dû à l'hétérogénéité du risque de diarrhée. Pour les périodes précédées par un épisode diarrhéique, la malnutrition paraît associée à une augmentation de 15 % du risque de diarrhée et pour celles qui ne l'ont pas été, cet état semble lié à une augmentation de 22 % de ce même risque.

Afin d'examiner la variabilité de ces estimations, des simulations sur ordinateur (écrites en langage BASIC) d'une étude épidémiologique semblable à celles qui ont été faites au Soudan et au Mexique ont été réalisées. Pour chaque simulation, un échantillon de 400 enfants a été suivi durant six périodes d'observation. Au début de chaque simulation, on a attribué à chaque enfant un risque de diarrhée (p_i) selon la distribution $f(P)$ donnée au tableau 1 en utilisant un générateur de nombres aléatoires. On a ensuite calculé la probabilité que l'enfant soit mal nourri [$R(p_i)$] et on s'est servi d'un autre nombre aléatoire, compris entre 0 et 1, pour déterminer l'état nutritionnel de l'enfant au début de l'étude. On a ensuite suivi l'enfant durant six périodes d'observation, la survenue de la diarrhée et de la malnutrition étant déterminée à l'aide de nombres aléatoires et appliquée aux probabilités définies ci-dessus. Trois cents simulations par ordinateur ont été faites; on trouvera au tableau 2 un résumé des résultats ainsi obtenus. Les valeurs moyennes des différents estimateurs du risque relatif sont très proches des valeurs probables théoriques obtenues ci-dessus. L'"intervalle de confiance 90%" indique la fourchette dans laquelle se trouvent 90 % des estimations.

TABEAU 2. Résultats de 300 simulations sur ordinateur portant sur une population de 400 enfants suivis durant 6 périodes d'observation

	Risque relatif "brut"	Risque relatif en cas de diarrhée dans l'intervalle précédent	Risque relatif si pas de diarrhée dans l'intervalle précédent
Moyenne	1,291	1,136	1,224
Minimum	1,084	0,924	0,949
Maximum	1,498	1,486	1,557
Ecart-type	0,081	0,101	0,104
Intervalle de confiance 90%	(1,16, 1,42)	(0,98, 1,31)	(1,04, 1,38)

DISCUSSION

Dans le présent document, nous avons présenté un modèle probabilité simple pour examiner le problème de la distinction entre cause et effet dans l'étude d'associations qui peuvent être bilatérales. Dans ce modèle, on montre comment, si A cause B et que le risque de A varie en fonction des sujets, les méthodes épidémiologiques courantes peuvent conduire à "observer" que B cause A. Il faut répéter qu'il s'agit d'un modèle hypothétique qui ne nous dit rien sur la nature réelle de l'association entre la diarrhée et la malnutrition. De nombreux chercheurs ont essayé d'élucider cette question. Le modèle se contente de montrer qu'en présence de taux de diarrhée hétérogènes chez la population étudiée, la manière habituelle d'envisager cette association ne permet pas de décider si un état nutritionnel médiocre précède la diarrhée ou vice versa. Si la diarrhée conduit à un état nutritionnel médiocre et que les taux de diarrhée dans la communauté sont hétérogènes, on peut avoir l'impression qu'un état nutritionnel médiocre fait augmenter la diarrhée, ce qui n'est pas le cas.

Il ne faut pas non plus oublier que ce modèle est d'une simplicité extrême et que, dans le cas de l'exemple de la diarrhée et de la malnutrition, il est critiquable sur un certain nombre de points; en effet, il part du principe que :

- l'état nutritionnel de la population est "stationnaire";

- étant donné un taux de diarrhée (p) d'un sujet, la survenue d'une diarrhée pendant une période d'observation donnée est indépendante des antécédents de diarrhée dudit sujet;
- le risque de malnutrition est homogène dans la population (mis à part l'effet d'hétérogénéité des taux de diarrhée).

Des recherches plus approfondies sont nécessaires pour pouvoir assouplir ces hypothèses et en adopter d'autres plus réalistes.

Dans le modèle utilisé, le degré de biais introduit dans l'estimation du risque relatif d'une association entre la malnutrition et la diarrhée dépend d'un certain nombre de paramètres : les diverses probabilités de transition entre les différents états nutritionnels en présence et en l'absence de diarrhée et la distribution des taux de diarrhée dans la population. Dans les simulations, on a utilisé les paramètres qui ont abouti à une prévalence et à un risque plausibles de malnutrition et de diarrhée. Si l'on compare les estimations du risque relatif obtenues à partir des études sur l'association entre la diarrhée et la malnutrition à celles qui découlent du modèle proposé, il ressort de ces dernières qu'une causalité inverse pourrait avoir beaucoup faussé les résultats de ces études. En choisissant d'autres valeurs pour les paramètres, on aurait pu faire diminuer ou augmenter le degré de biais introduit dans l'estimation du risque relatif. Il serait intéressant d'estimer certaines de ces valeurs à partir des groupes de données existants pour déterminer si ces estimations sont comparables aux paramètres utilisés dans les simulations. On pourrait alors appliquer ces estimations au modèle pour évaluer le degré de biais que l'on aurait peut-être atteint dans ces conditions. Ainsi, on saurait mieux si oui ou non le biais dû à une causalité inverse pose en pratique un sérieux problème dans les études sur la malnutrition et la diarrhée.

Le biais que nous avons observé est dû à la fois à la causalité inverse (A cause B) et à l'hétérogénéité du taux de A (diarrhée) dans la communauté. Si tous les sujets couraient autant de risques d'être atteints de A, ou que A ne causait pas B, alors l'estimation du risque relatif obtenue par une analyse "classique" ne serait pas entachée d'un biais. D'où l'idée d'envisager le problème, du moins dans le cas de la diarrhée et de la malnutrition, par des analyses intra-individuelles des données où chaque sujet est son propre témoin. Cette démarche a été adoptée par Cole (1989), qui a utilisé la "méthode régression intra-individuelle" pour étudier les effets de la maladie sur le rythme de prise de poids (est-ce que A cause B?). L'efficacité de différentes techniques qui permettent de faire des analyses "intra-individuelles" pour éliminer le biais dû à la causalité inverse, en particulier lorsque l'on ne peut pas partir du principe que les périodes d'observation sont indépendantes, reste à évaluer.

En résumé, on a examiné dans le présent document comment résoudre le problème des associations susceptibles d'être bilatérales par nature. On a montré qu'en présence d'un taux de maladie hétérogène, les méthodes épidémiologiques habituelles pourraient aboutir à des résultats biaisés. Il est nécessaire de mettre au point et d'évaluer d'autres manières d'envisager la conception et l'analyse d'études de telles associations.

REMERCIEMENTS

Le présent travail a reçu le soutien du Programme de lutte contre les maladies diarrhéiques de l'Organisation mondiale de la Santé et de l'Overseas Development Administration (Royaume-Uni). Notre gratitude va à Jean-Pierre Habicht, Betty Kirkwood, Chessa Lutter et José Martines pour les remarques et les suggestions constructives qu'ils ont apportées au sujet des premières versions du présent document.

BIBLIOGRAPHIE

- Chandra, R.K. (1986) Nutritional regulation of immunity and infection: from epidemiology to phenomenology to clinical practice. Journal of Pediatric Gastroenterology and Nutrition, 5:844-852
- Chen, L.C. (1983) Interactions of diarrhoea and malnutrition: mechanisms and interventions. In: Chen L.C. and Scrimshaw, N.S. (eds). Diarrhoea and malnutrition. New York, Plenum Press.
- Cole, T.J. (1989) Relating growth-infection interaction. Acta Paediatrica Scandinavica, Supplement, 350:14-20
- Cox, D.R. & Miller, H.D. (1972) The theory of stochastic processes. 2nd ed. London, Chapman and Hall, p.80
- El Samani, E.F.Z., Willett, W.C. & Ware, J.H. (1988) Association of malnutrition and diarrhoea in children aged under five years. A prospective follow-up study in a rural Sudanese community. American Journal of Epidemiology, 128:93-105
- Hennekens, C.H. & Buring, J.E. (1987) Epidemiology in medicine. Boston, Little, Brown and Company, p.153
- Rose, G. & Shipley, M.J. (1980) Plasma lipids and mortality: a source of error. Lancet, i:523-526
- Scrimshaw, N. S., Taylor, C. E. & Gordon, J. E. (1968) Interactions entre l'état nutritionnel et les infections. (Série de Monographies N° 57). Genève, Organisation mondiale de la Santé
- Sepulveda, J., Willett, W. & Munoz, A. (1988) Malnutrition and diarrhoea. A longitudinal study among urban Mexican children. American Journal of Epidemiology, 127:365-376
- Sirisinha, S., Suskind, R., Edelman, R. et al. (1975) Secretory and serum IgA in children with protein-calorie malnutrition. Pediatrics, 55:166-170
- Tomkins, A. and Watson, W. (1989) Malnutrition and infection: a review. Londres, Clinical Nutrition Unit, London School of Hygiene and Tropical Medicine.