

Quand rien ne va mal, peut-on dire que tout va bien?

A propos de l'interprétation des numérateurs nuls

James A. Hanley, PhD, Abby Lippman-Hand, PhD
Department of Epidemiology and Health and Centre for Human Genetics, McGill University, Montreal

Les patients demandent souvent à leur médecin de donner un ordre de grandeur du taux de risque inhérent à un acte médical particulier ou d'évaluer la probabilité de survenue d'une éventualité touchant à leur santé. Pour cela, le praticien fait appel autant à sa propre expérience, qu'aux données de la littérature; l'objectif étant de situer l'estimation dans une fourchette assez étroite.

Lorsque l'éventualité ou l'acte en question sont relativement rares, une estimation précise, bien que souhaitable, peut être difficile, voire impossible.

De plus, quand les données disponibles proviennent d'une seule étude, au cours de laquelle les événements en question ne sont pas survenus (c'est-à-dire une étude dans laquelle un numérateur zéro est rapporté), tirer des conclusions concernant le risque paraît particulièrement problématique. La façon très différente, aussi bien quantitative que qualitative, dont est perçue la survenue de « non-événement » comparée à la survenue d'un ou plusieurs événements, justifie que nous examinions les implications psychologiques et statistiques qui influencent l'interprétation d'un numérateur nul. Il importe en particulier de souligner (1) qu'un numérateur nul ne signifie pas nécessairement « absence de risque », (2) qu'un numérateur nul n'exclut pas que l'on puisse procéder à des inférences statistiques concernant l'importance du risque, (3) que les principes de raisonnement statistique s'appliquant aux numérateurs non nuls s'appliquent également aux numérateurs nuls. En fait, il existe une règle simple et rapide qui permet de calculer le risque maximal à long terme associé à l'observation d'effets nuls dans une population de n'importe quelle taille.

Exemples

Les études rapportant un numérateur nul sont assez fréquentes dans la littérature. Nos exemples sont tirés de numéros récents de revues médicales de premier plan (dont le JAMA). Ils ont été choisis de façon arbitraire, uniquement dans un but didactique. Nous avons donc réduit les développements à l'essentiel. Le problème proposé au lecteur est, pour chaque exemple, de savoir quelles inférences statistiques peuvent être tirées des données fournies.

1^{er} exemple. — Sur 14 garçons ayant été suivis pendant une durée moyenne de 5 ans et demi, après chimiothérapie pour leucémie, aucun n'a présenté d'anomalie de la fonction testiculaire (le taux d'anomalie a donc été de 0/14) [1]. Quelle est la probabilité de dysfonction testiculaire compatible avec ces résultats?

2^e exemple. — Une série de 112 nouveau-nés viables dont les mères avaient été vaccinées contre la rubéole, a été étudiée pour déterminer les risques de l'exposition fœtale au vaccin [2]. Aucun des nouveau-nés (0/112) n'a présenté une des anomalies congénitales associées à la rubéole congénitale. Quelle est la probabilité maximale d'anomalies compatibles avec le fait qu'aucun des 112 enfants pris en compte dans une seule étude n'a présenté d'anomalies?

3^e exemple. — Nous adapterons le dernier exemple pour expliquer les inférences statistiques qui peuvent être faites lorsqu'il existe un numérateur nul. Dans une étude portant sur les frères et sœurs de 167 nourrissons porteurs d'une malformation (dysraphie) trachéo-œsophagienne (MTO), aucun enfant n'a présenté d'anomalie du tube neural [3]. (Nous n'aborderons pas ici le problème de la définition exacte du dénominateur et supposons dans un but didactique que 167

frères et sœurs ont été étudiés, et que le taux d'anomalies du canal a été de 0/167). Le fait est-il suffisant pour affirmer que le risque d'anomalie de fermeture du tube neural dans la fratrie d'un enfant porteur d'un MTO n'est pas supérieur au risque d'anomalie dans la population générale (1,5 pour 1 000 naissances dans la région étudiée)? Ou plutôt, ce taux de 0/167 est-il compatible avec une probabilité telle que l'on doive systématiquement faire un examen de dépistage prénatal lors des grossesses suivantes chez les mères ayant donné naissance à un enfant porteur d'une anomalie du tube neural?

Malgré leur contenu différent, ces trois exemples ont la même structure : dans ces trois études, aucun événement indésirable n'est survenu, rien n'est arrivé. Seuls les dénominateurs (la taille des échantillons de population) diffèrent. Afin d'illustrer le mode d'inférence statistique utilisé lorsqu'il existe un numérateur nul, nous allons transposer les chiffres du 3^e exemple à l'étude du niveau de risque associé à l'utilisation d'un produit de contraste par les radiologues (nous choisissons cet exemple imaginaire afin d'éviter les controverses à propos des détails spécifiques d'une étude connue).

Supposons qu'il ait été montré que le produit de contraste courant utilisé par les radiologues depuis de longues années est responsable de réaction indésirable chez environ 15 pour 10 000 patients exposés. C'est dire que le risque connu pour cet ancien produit est de 1,5/1 000 (risque du même ordre que celui de survenue d'une anomalie du tube neural dans la population de la 3^e étude). Supposons ensuite qu'un nouvel agent de contraste soit introduit. Peu de temps après, est publié un rapport sur son utilisation chez 167 patients : aucun patient n'a

présenté la réaction en question. Ces données nous permettent-elles d'inférer qu'à long terme le produit nouveau entraînera un risque moindre, un risque identique, ou un risque supérieur à celui présenté par l'ancien produit?

Commençons par supposer que le risque à long terme avec le nouveau produit est identique à celui encouru avec le produit précédent. Dans cette hypothèse, chaque patient a 0,0015 chances de présenter une réaction, et inversement $1 - 0,0015 = 0,9985$ chance de ne pas en présenter. La probabilité pour que les 167 patients échappent aux effets secondaires (c'est-à-dire pour que le taux de réaction soit de 0/167) est de $0,9985^{167}$, soit 78 %. Il n'est donc pas surprenant de ne pas constater d'effet secondaire, dans ce groupe. En fait, si le nouveau produit présente le même taux de risque que le précédent, ce qui serait surprenant, c'est que quelqu'un présente une réaction.

Montrons-nous plus pessimistes et supposons que le taux de complications à long terme avec ce nouveau produit est beaucoup plus élevé qu'avec le précédent. Nous pouvons alors fixer arbitrairement un second niveau-test pour le risque, par exemple 1 pour 100. Là encore, ne pas observer de réaction chez 167 patients ne serait pas si surprenant que cela : en effet, il y aurait alors $0,99^{167}$ ou 19 % de chances qu'aucun de ces 167 patients ne présente de réaction à ce nouveau produit, même si le risque réel était de 1 %.

Pour finir, nous pouvons choisir de fixer un niveau de risque encore plus élevé, par exemple 4 %. Si c'est bien le risque vrai, il y aura alors 0,1 % de chances de ne constater aucune réaction dans un groupe de 167 patients. Nous serions alors très surpris d'obtenir un résultat aussi éloigné du taux « vrai ».

En fait, chaque fois que nous fixons arbitrairement une valeur pour le niveau du risque à long terme, la valeur choisie par nous dépend de ce que nous pourrions appeler notre « envie d'être surpris » par les résultats obtenus. C'est là que nous décidons que ce n'est pas uniquement une question de (mal)chance si un résultat qui ne « devait pas se produire (sur la base de ce que nous avons posé comme étant « le vrai ») semble néanmoins s'être produit ».

Plutôt si l'observation est vraiment trop surprenante par rapport à la valeur de risque choisie, nous préférons décider que le risque vrai doit être plus faible.

Par convention, on a adopté le niveau de probabilité de 5 % comme limite de crédibilité; ainsi la survenue d'un événement qui a au moins 5 % de chances de se produire est considérée comme n'étant pas « si surprenante que cela » (« cela arrive »). Pour n'importe quelle valeur inférieure, la survenue de l'événement est surprenante. Le tableau 1 montre une série de niveaux-tests choisis pour le risque à long terme et pour chacun d'eux, la probabilité de ne voir survenir aucune réaction dans une population de 167 patients. Comme ce tableau permet de le constater, notre « seuil d'étonnement » de 5 % est atteint quand le niveau « vrai » du risque à long terme est d'environ 2 %. En d'autres termes, les résultats « collent » avec ou « ne sont pas étonnamment éloignés » d'un risque à long terme de 2 % ou plus. Cet « éventail plausible de possibilités où situer le vrai » qui va de zéro au maximum compatible avec notre seuil d'étonnement de 5 % est appelé ordinairement : « intervalle de confiance à 95 % ».

Inférences statistiques : La règle du trois

Tous nos exemples portent sur un nombre fini de sujets auxquels l'événement particulier n'est pas arrivé. Lorsqu'on procède à des inférences statistiques à partir de ces observations rapportant un numérateur nul, on veut pouvoir affirmer avec un certain degré de confiance que le niveau de risque vrai à long terme (de la chimiothérapie, de l'immunisation contre la rubéole, ou d'avoir un frère porteur d'une MTO) se situe entre 0 et une limite supérieure.

Pour pouvoir nous fier à 95 % à notre fourchette d'estimation, une simple règle (d'origine inconnue) peut être appliquée [4]. Cette « règle du trois » postule que si aucun des n patients ne présente la réaction en question, nous pouvons être sûrs à 95 % que la probabilité de cet événement est au plus 3 sur n (nous écrivons : $3/n$). En d'autres termes, la limite supérieure de l'intervalle de

confiance à 95 % pour un niveau de risque de $0/n$ est approximativement $3/n$. (Cette approximation est remarquablement satisfaisante : quand n est supérieure à 30, les résultats obtenus par la règle du trois sont conformes à ceux du calcul exact à 1 % près. Au-dessous de 30, la règle du trois donne un niveau de risque légèrement surestimé mais alors le risque maximal est tellement élevé [26 % si $n = 10$] que la différence avec le chiffre donné par la règle du trois [$3/10 = 30\%$] ne vaut pas la peine qu'on se tracasse). La dérivation de la règle du trois se déroule comme suit :

Pour trouver la probabilité maximale avec laquelle un résultat de $0/n$ est compatible (« compatible » cela veut dire : possède un taux de crédibilité d'au moins 5 %), il faut soit procéder comme nous l'avons fait par la méthode des essais et erreurs, soit résoudre l'équation suivante :

$$(1 - \text{risque maximal})^n = \sqrt{0,05}$$

ce qui peut s'écrire en prenant la racine-nième de chaque membre de l'égalité :

$$1 - \text{risque maximal} = \sqrt[n]{0,05} \quad (1)$$

Le lecteur qui utilise une calculatrice pour évaluer $\sqrt[n]{0,05}$ ou $0,05^{1/n}$ pour $n > 30$ trouvera un résultat très proche de $(n - 3)/n$ ou $1 - (3/n)$; on peut donc récrire l'équation (1) comme suit :

$$1 - \text{risque maximal} = 1 - (3/n)$$

$$\text{ou risque maximal} = 3/n$$

Le lecteur qui voudrait comprendre « d'où vient » ce $3/n$ peut récrire $0,05^{1/n}$ comme une série infinie à savoir :

$$0,05^{1/n} = 1 + \{(\ln 0,05)/n\} + \{(\ln 0,05)^2/2 n^2\} + \dots \quad (2)$$

où $\ln 0,05$ représente le logarithme naturel de 0,05; sachant $\ln 0,05$ égal à $-2,9957 \approx -3$ (en arrondissant les deux dernières décimales) et étant donné $n \geq 30$, on peut considérer que dans le membre à droite de l'égalité (2) le terme contenant la fraction de n^2 est négligeable; dont $0,05^{1/n}$ est très proche de $1 - 3/n$. Pour un intervalle de confiance à 99 %, la règle correspondante serait une règle du 4,6 ($\ln = 0,01 = -4,6051\dots$) pour un intervalle de 99,9 %, une règle du 6,9.

Certains lecteurs se demanderont pourquoi on ne calcule pas immédiatement la probabilité de $0/n$ à partir de la loi de distribution de Poisson : on peut, bien sûr, le faire en résolvant l'équation : $\exp(-n \cdot \text{risque maximal}) = 0,05$. Comme on pouvait s'y attendre, la réponse est la même : $3/n$.

En appliquant cette règle à nos exemples, en limitant les inférences statistiques aux populations auxquelles appartiennent nos patients, nous

pouvons avec un intervalle de confiance à 95 % formuler les conclusions suivantes :

1^{er} exemple. — Le risque maximal d'anomalie de la fonction testiculaire après chimiothérapie, n'est pas supérieur à 21 % ($3/n = 3/14 = 21\%$). Ici, la limite supérieure de confiance « juste » basée sur le résultat de 0/14 est 19 %. Il est intéressant de noter que cette « règle des 14 non-survenues consécutives » est couramment utilisée en recherche sur le traitement des cancers [5], pour éliminer les produits qui n'auront probablement pas d'activité chez au moins un malade sur cinq.

2^e exemple. — La fréquence des malformations chez les enfants de femmes vaccinées contre la rubéole pendant la grossesse, ne dépasse pas 2,7 %.

3^e exemple. — La fréquence des anomalies du tube neural n'est pas supérieure à 1,8 % chez les frères et sœurs de patients avec MTO. Remarquons que cet éventail de 0 à 1,8 % comprend le niveau de risque de 1,5 pour 1 000 trouvé dans la population générale, aussi bien que le risque existant pour les parents au second degré de patients porteurs d'anomalies du tube neural.

Le tableau 2 résume la règle du trois et détermine le risque « vrai » qu'on peut déduire des non-événements dans des séries de différentes tailles.

Commentaires

Bien que les exemples choisis reflètent nos préoccupations propres, on peut rencontrer des numérateurs nuls dans des contextes très divers : un nouveau test diagnostique jamais pris en défaut, une technique chirurgicale n'ayant jamais entraîné de complication, un essai de vaccin pour lequel aucun effet secondaire majeur n'a été décrit, une clientèle d'ophtalmologiste dans laquelle aucun patient glaucomateux n'a jamais eu moins de 23 ans, une compagnie aérienne n'ayant jamais eu d'accident. C'est pourquoi il est important de comprendre les limites des inférences statistiques qui peuvent être faites à partir de telles observations.

Il est intéressant de noter que si, du point de vue statistique, la survenue de 0 événement dans n observations peut être interprétée

comme n'importe quel binôme, elle semble souvent avoir un impact qualitatif de loin supérieur à sa réelle signification quantitative. On peut trouver dans la littérature certains indices (parfois contradictoires) qui mettent sur la voie d'une explication : (1). On a tendance à oublier la taille des populations-dénominateurs sur lesquelles les taux sont fondés [6]. Par exemple, on accorde foi de la même façon, à un taux de 10 %, qu'il soit observé sur 20 ou sur 200 cas. Ce même mécanisme joue vraisemblablement pour un taux de 0 % (2). L'attention tend à se focaliser sur les numérateurs. Les parents qui ont bénéficié de l'assistance d'un conseiller génétique, par exemple, envisagent généralement le risque sous un mode binaire : quelque chose va ou ne va pas leur arriver, peu importe le risque réel [7]. Un « un » au numérateur ne s'efface jamais, quelle que soit la taille du dénominateur. Pour eux ce qui compte, ce n'est pas le niveau du risque mais la notion de survenue toujours possible. Peut-être les numérateurs nuls véhiculent-ils le même poids de signification en ce sens qu'ils suggèrent (à tort) que l'événement est impossible (3). Quand on a affaire à un risque théorique qui ne s'est jamais réalisé (lorsqu'une chose possible n'est pas encore arrivée), les gens ont tendance à croire que la chose n'arrivera pas dans le futur. En l'absence de preuves susceptibles de démontrer la variabilité des réactions possibles, on sous-estime le risque réel. (Le marquis de Laplace, conseiller et statisticien des joueurs royaux d'Europe et de Russie, à qui on rappelait vers 1800, que jamais au cours des 5 000 années précédentes, le soleil n'avait manqué de se lever, (apparemment, on ne se fiait pas aux données antérieures), et à qui on demandait de calculer la probabilité du lever du soleil le lendemain, répondit que la formule approchant $n : 1$, où n était le nombre de jours de la série pouvait fournir la base d'un *bonne* pari. Mais pour un parieur moins confiant en la nature, la formule $n/3$ aurait presque été plus sûre).

Il importe que soit reformulée dans une nouvelle optique la notion de numérateur nul et il faut encourager ceux qui publient des observations de ce type à prendre en compte le niveau de risque maximal compati-

Tableau 1. — Probabilité pour qu'un échantillon de 167 sujets ne présente « aucune complication »

Si le taux de complication à long terme est	Alors la probabilité (%) d'observer zéro complication chez 167 sujets est
1 sur 10 000	98
1 sur 1 000	85
1,5 sur 1 000	78
1 sur 200	43
1 sur 100	19
1 sur 56	5
1 sur 25	0,1

Tableau 2. — Risque à long terme inféré à partir de 0/n (avec intervalle de confiance à 95 %, c'est-à-dire une limite de crédibilité de 5 %)

Taux de	Inférence d'un taux de risque à long terme (%) plus élevé que
0/10	26*(30)**
0/20	14(15)
0/30	10(10)
0/50	6(3)
0/100	3(3)
0/1 000	0,3(0,3)

* Dérivé « exactement » en résolvant : $(1 - \text{taux maximal})^n = 0,05$.

** Dérivé de la règle du trois.

ble avec leurs résultats. Ici, la notion d'intervalle de confiance est utile car elle permet de traduire les résultats obtenus sur un échantillon, non plus par un chiffre unique, mais par une fourchette où se situe vraisemblablement le taux de risque caractéristique de la population. Dans la mesure où un intervalle de confiance peut être déterminé facilement à partir d'un numérateur nul, en utilisant la règle du trois, nous espérons que les auteurs assez chanceux pour pouvoir signaler : « aucun problème pour l'instant » seront enfin à même de calculer à quoi leurs populations de patients à venir pourront s'attendre, pour le meilleur et pour le pire.

Références

- Blatt J., Poplack D.G., Sherins R.J. : Testicular function in boys after chemotherapy for acute lymphoblastic leukemia. *N. Engl. J. Med.* 1981; 304 : 1121-1124.
- Preblud S.R., Stetler H.C., Frank J.A., et al. : Fetal risk associated with rubella vaccine. *JAMA* 1981; 246 : 1413-1417.
- Baird P., MacDonald E.G. : Siblings of children with tracheoesophageal dysphasia. *Can. Med. Assoc. J.* 1981; 125 : 1083-1084.
- Rumke C.L. : Implications of the statement : No side effects were observed. *N. Engl. J. Med.* 1975; 292 : 372-373.
- Holland J.F., Frei E. (eds) : *Cancer Medicine*, ed 2. Philadelphia, Lea & Febiger, 1982, p. 535.
- Tversky A., Kahneman D. : Judgment under uncertainty : Heuristics and biases. *Science* 1974; 185 : 1124-1131.
- Lippman-Hand A., Fraser F.C. : Genetic counseling : The postcounseling period : I. Parents' perceptions of uncertainty. *Am. J. Med. Genet.* 1979; 4 : 51-71.