



Statistiques et essais cliniques

François Kohler

Laboratoire SPI-EAO, Faculté de Médecine, Vandœuvre-les-Nancy

francois.kohler@univ-lorraine.fr

Introduction

« La statistique » ou « les statistiques » sont appliquées à de nombreux domaines comme les jeux, l'agriculture, la physique... et la santé. On parle de **Statistique** au singulier : méthode mathématique ou valeur calculée, et au pluriel : **Statistiques** pour des collections de données sur un thème : maladies à déclaration obligatoire, statistiques du chômage [1].

Envisageons trois types d'études fréquentes

Etudes visant à déterminer la fréquence des maladies

Elles ont comme objectif de déterminer l'**Incidence** : nombre de nouveaux malades dans une période donnée ou la **Prévalence** : nombre de malades présents un jour J (ou dans une période donnée), c'est-à-dire les malades qui ont débuté leur maladie avant le jour J plus ceux qui ont débuté leur maladie le jour J.

Etudes visant à déterminer les propriétés d'un signe ou d'un test diagnostique

Il n'y a pas de test parfait. Elles ont comme objectif de déterminer les caractéristiques intrinsèques du test : **Sensibilité** : fréquence des tests positifs chez les malades (vrai positif), et la **Spécificité** : fréquence des tests négatifs chez les sujets n'ayant pas la maladie (vrai négatif). Si la fréquence de la maladie est connue, ces études permettent d'aider à la décision médicale en connaissant la **Valeur prédictive positive** : probabilité qu'un sujet ayant un test positif soit atteint de la maladie et la **Valeur prédictive négative** : probabilité qu'un sujet ayant un test négatif ne soit pas atteint de la maladie.

Etudes permettant la recherche de facteur influant la maladie comme la recherche de facteur de risque ou l'évaluation d'un traitement

- ▶ **Etudes observationnelles** : études dans lesquelles on se contente d'analyser des données sans réaliser volontairement d'expérience ; on se contente d'observer sans intervenir. On distingue les **études prospectives** dont les **études exposés/non exposés** qui permettent dans chacun des groupes de connaître le risque de survenue de la maladie après un certain délai ou encore les **études de survie** qui non seulement prennent en compte la fréquence de survenue des événements mais leur dynamique de survenue dans le temps. Elles sont opposées aux **études rétrospectives** : comme les études **cas/témoins**.
- ▶ **Etudes expérimentales** : on réalise une expérience, l'exemple type est l'essai thérapeutique avec ses différents plans d'expériences [2]. Nous nous contenterons du cas le plus simple de la comparaison de 2 traitements administrés chacun à un groupe de sujets.

Des données observées à l'estimation : l'inférence statistique

Les études sont menées sur des échantillons représentatifs limités. D'une part, les paramètres statistiques : moyenne, écart type, amplitude, pourcentage... obtenus sur l'échantillon peuvent être ou non de bons estimateurs des paramètres de la population. Pour une variable quantitative, la moyenne de l'échantillon est le meilleur estimateur de la moyenne de la population mais l'écart type (SD) de l'échantillon n'est pas le meilleur estimateur de l'écart type de la population ; il faut apporter une correction en multipliant l'écart type de l'échantillon par $\sqrt{\frac{n}{n-1}}$ (n est l'effectif de l'échantillon).

Cette correction devient négligeable quand l'effectif est grand.

D'autre part, du fait du hasard, les paramètres vont fluctuer autour de la vraie valeur du paramètre de la population. Si nous avons une population de 3 individus dont les tailles sont 160, 170, 180 cm et que l'on tire au sort 2 individus, on peut obtenir les tailles 160 et 170, la moyenne de l'échantillon



est 165 alors que la vraie moyenne de la population est 170. Si on avait tiré un autre échantillon, on aurait eu une autre moyenne. La distribution des moyennes issues d'échantillons indépendants est caractérisée par des paramètres dont l'écart type des moyennes (SDM, ESM). Celui-ci peut être estimé à partir de l'écart type de l'échantillon par la formule $\frac{\text{Ecart type de l'échantillon}}{\sqrt{(n-1)}}$. On peut alors construire autour

de la moyenne observée dans l'échantillon un intervalle dans lequel la probabilité que la vraie moyenne de la population s'y trouve soit de x %. En général on prend x = 95, c'est-à-dire que l'on accepte un risque de 5 % que la vraie moyenne soit en dehors de l'intervalle. On parle d'intervalle de confiance à 95 %.

La même démarche est effectuée pour tous les paramètres, on peut créer un intervalle de confiance pour un pourcentage, pour un Odds Ratio...

Exemple : on a observé un pourcentage de 20 % sur 100 personnes. L'intervalle de confiance à 95 % a comme bornes 12 % et 28 %, Il y a 95 chances sur 100 que le vrai pourcentage soit compris entre 12 et 28 %.

Les études observationnelles

Les études cas/témoins et l'Odds Ratio

Elles sont utilisées quand le **phénomène est rare**. Dans un observatoire de gastroentérologie, on a observé, pour certains patients devant bénéficier d'une coloscopie, des hypoglycémies après l'examen. On soupçonne la préparation X. On sait que l'âge, le sexe et le poids sont des facteurs qui peuvent jouer un rôle dans l'apparition d'hypoglycémie (facteurs de confusion connus). On apparie chaque cas ayant eu une hypoglycémie à un témoin, de même sexe et de même âge, n'en ayant pas eu. Dans chacun des groupes, on regarde quelle préparation a été utilisée. Les résultats sont donnés dans le tableau ci-dessous :

	Hypoglycémie (Cas)	Pas d'hypoglycémie (Témoins)	Total
Préparation X	10	50	60
Autre préparation	5	50	55
Total	100	100	200

Du fait de la méthode, on a des effectifs identiques (100) de personnes ayant une hypoglycémie et de personnes n'en ayant pas. On ne peut pas avec ce type d'étude déterminer la fréquence des hypoglycémies ; par construction, elle est de 50 %. Pour répondre à la question « Y a-t-il plus de risque avec la préparation X ? », on utilise l'Odds Ratio (OR).

Odd en français se traduit par la cote, terme bien connu des turfistes. La cote de la préparation

X est le rapport $\frac{10}{60} / \frac{50}{60}$ soit 1/5 (1 contre 5 pour les turfistes).

Celle des autres préparations est de $\frac{5}{55} / \frac{50}{55}$ soit 1/10.

L'Odds Ratio (OR) est le rapport des cotes : OR $0,2/0,1 = 2$. Cela s'interprète de la manière suivante : on a 2 fois plus d'hypoglycémie avec la préparation X qu'avec une autre préparation.

Un OR > 1 signifie que le facteur est un **facteur de risque**, OR = 1 **le facteur n'a pas d'effet**, OR < 1 le facteur est un **facteur de protection**. L'intervalle de confiance permet de savoir si cela n'est pas le fait du hasard. **Si l'intervalle de confiance (95 %) de l'OR contient la valeur 1, on ne peut pas conclure** : La vraie valeur de l'OR peut être inférieure, égale ou supérieure à 1.



Les études exposés/non exposés

Ce sont des études prospectives dans lesquelles un groupe d'individus est suivi dans le temps. Certains sont exposés à un facteur de risque (consommateurs habituels de compléments alimentaires) et d'autres non. On mesure la fréquence de survenue de la maladie après un certain temps d'exposition, 5 ans par exemple dans le sous-groupe exposé et dans le sous-groupe non exposé. Ce type d'étude ne peut être réalisé que si la maladie est fréquente ; de plus, ces études sont longues et coûteuses. Elles permettent de connaître la fréquence de la maladie dans chaque groupe et dans la population. On étudie la survenue de cancer à 5 ans chez des personnes exposées à la consommation régulière de compléments alimentaires et chez des personnes n'en consommant pas. Les résultats sont présentés sous forme d'un tableau à 4 cases plus les totaux marginaux :

	Survenue d'un cancer	Pas de cancer	Total
Compléments alimentaires +	4	246	250
Compléments alimentaires -	196	9 554	9 750
Total	200	9 800	10 000

La fréquence globale des cancers dans ce groupe est de $200/10\,000 = 0,02$ soit 2 %. La fréquence de consommation de compléments alimentaires dans ce groupe est de $250/10\,000$ soit 2,5 %. Le risque de cancer dans le groupe consommant des compléments alimentaires (exposés) $R1 = 4/250 = 0,016$ soit 1,6%. Dans le groupe non exposé, le risque $R0$ est de $196/9\,750 = 0,02$ soit 2%. **Le risque relatif (RR) est le rapport $R1/R0 = 0,016/0,02 = 0,8$. L'interprétation est identique à celle de l'OR.** La consommation de compléments alimentaires semble être un facteur de protection contre la survenue de cancer mais pour conclure, il faut connaître l'intervalle de confiance à 95 % autour du RR (0,8) obtenu.

Si l'on avait calculé l'OR on aurait $OR = 0,79$. Quand la maladie est peu fréquente OR et RR sont très proches. Les études exposés/non exposés sont moins sujettes aux biais que les études cas/témoins et ont une meilleure valeur probante dans la médecine factuelle (evidence based medicine) mais sont plus difficiles et plus coûteuses que les études cas/témoins.

Les courbes de survie

Comme dans les études exposés/non exposés, on étudie la survenue d'un événement binaire : décès, survenue d'une métastase... mais en tenant compte de la dynamique temporelle de survenue. De plus, il existe des censures liées à l'interruption du suivi avant le décès de tous les patients ou survenant durant la période de suivi, induites par des patients perdus de vue. Les tests statistiques du type Logrank notamment, permettent de comparer des courbes de survie obtenues avec différents traitements. On utilise le **hazard ratio (HR)** pour quantifier l'effet du traitement. **HR est le rapport des risques instantanés dans le groupe traité ($h1$) divisé par le risque dans le groupe contrôle ($h0$).** En général, HR n'est pas constant mais est une fonction du temps. Si le risque instantané de décès dans le groupe contrôle est de 3 % et est constant dans le temps et que dans le groupe traité il est de 1,5 % et également constant dans le temps, le HR est $1,5/3 = 0,5$. **On associe au HR un intervalle de confiance. L'interprétation est identique à celle du RR :** le risque instantané de décès est seulement la moitié du risque (instantané) sans traitement. Dans la pratique, les risques instantanés sont généralement variables en fonction du temps et donc le HR est une fonction du temps. On est alors amené à réaliser une hypothèse de proportionnalité des risques : l'effet du traitement doit rester constant en fonction du temps [3,4].

Les études expérimentales

Il ne s'agit plus d'observer ce qui survient naturellement mais de réaliser une expérience en maîtrisant certains facteurs comme le traitement dont va bénéficier le patient. Ces études posent de nombreuses questions éthiques. L'exemple typique est **l'essai thérapeutique en groupes parallèles**. Dans un groupe de malades, on tire au sort deux sous-groupes : ceux qui vont recevoir A et ceux qui vont recevoir B. Ceci se fait en double aveugle : ni l'expérimentateur, ni le patient ne sauront quel traitement est administré ; de la même manière, les mesures se feront en aveugle (celui qui constate le décès ne connaît pas le traitement). Ces 2 conditions sont primordiales pour



l'imputation des différences observées au traitement. On veut comparer l'efficacité de 2 traitements A et B dans la maladie M. Le critère de jugement est le taux de mortalité globale à 2 ans ; il n'y a pas eu de perdu de vue. On a tiré au sort les patients bénéficiant de A et de B. Les résultats sont donnés dans le tableau ci-dessous :

	Décès à 2 ans	Vivants à 2 ans	Total
Traitement A	20	80	100
Traitement B	35	65	100
Total	55	145	200

On a 20 % de décès avec A contre 35 % avec B mais cette différence est-elle due au hasard ? Pour répondre à cette question, on va utiliser un test statistique qui va nous permettre de décider avec certains risques de se tromper.

Le tableau ci-dessous résume ce problème de décision :

		Réalité (inconnue)	
		Il y a une différence (un lien)	Il n'y a pas de différence (pas de lien)
Test Statistique	Rejet de H0 : la différence n'est pas due au hasard	Conforme à la réalité	Risque alpha : rejet de H0 alors qu'elle est vraie, on conclut à une différence significative à tort. ("p" est un risque alpha a posteriori si H0 est vraie)
	Acceptation de H0 : la différence est due au hasard	Risque Bêta : on conclut à l'égalité alors que c'est faux.	Conforme à la réalité

Au terme du test statistique, on pourra conclure que la différence n'est pas attribuable au hasard avec un risque dit de première espèce ou **risque alpha** ou conclure que la différence est due au hasard avec un risque dit de deuxième espèce ou **risque bêta**. La puissance est égale à 1-bêta ; elle dépend notamment du type de test statistique utilisé (les tests paramétriques sont plus puissants que les tests non paramétriques), de la variabilité du phénomène, du nombre de sujets... La planification de l'essai avec calcul du nombre de sujets nécessaire permet de maîtriser *a priori* le risque bêta, donc la puissance, et d'accepter en toute connaissance de cause H0. Sans cela, le statisticien utilisera la formule consacrée : « Il n'a pas été possible de mettre en évidence une différence significative ».

La démarche du test statistique est la suivante et s'apparente à une démonstration par l'absurde :

1. On pose une hypothèse appelée hypothèse nulle (H0) que l'on accepte dans le but de la rejeter. Celle-ci peut être formulée par « La différence observée entre 20 et 35 % est due au hasard ». Dans le cadre de la planification de l'essai qui permettra de maîtriser le risque bêta, la formulation devient « La différence est inférieure ou égale à un delta fixé a priori » : par exemple, le traitement A sera jugé différent du traitement B si la différence du pourcentage de décès est d'au moins 5 %. Ceci permettra de calculer un nombre de sujets nécessaires en tenant compte de cette différence et des risques alpha et bêta consentis, donc maîtrisés.
2. On pose l'hypothèse alternative (H1). « La différence n'est pas due au hasard ». Le taux de décès avec A est soit supérieur soit inférieur à celui de B sans qu'a priori on sache avant l'expérience le sens de l'inégalité ; le test est alors dit bilatéral. Si du fait de connaissances antérieures, a priori, le sens est imposé, on parle de test unilatéral : par exemple, la taille



des enfants entre 2 et 3 ans ne peut que rester identique (H_0) ou augmenter ; on ne teste donc pas une éventuelle diminution de taille entre 2 et 3 ans. H_1 devient « il existe une augmentation de la taille » et non pas « il existe une différence de taille entre 2 et 3 ans ».

3. En fonction du problème : comparaison de pourcentage, comparaison de moyenne... et des conditions d'application, on choisit un test statistique : test du χ^2 (Khi²), test de Student... qui permet de calculer une statistique dont on connaît la loi théorique de probabilité sous H_0 . Le résultat est le « p » qui est la probabilité d'avoir observé cette statistique si l'hypothèse nulle est vraie.

On décide :

► **Si « p » est inférieur au risque alpha** consenti avant d'avoir réalisé l'expérience, on rejette l'hypothèse nulle et on accepte l'hypothèse alternative. **La différence est significative.** La différence n'est pas due au hasard mais a-t-elle pour autant une importance clinique ? Le risque alpha consenti est habituellement de 5 % mais devrait être pris en fonction du problème médical.

► **Si « p » est supérieur au risque alpha** consenti, on n'a pas été en mesure de mettre en évidence une différence significative. On serait tenté de conclure que « la différence observée est due au hasard » **mais on ne connaît pas le risque bêta, sauf si l'on a calculé un nombre de sujets nécessaires**, et conclure devient impossible sans au moins calculer un risque bêta (ou une puissance) a posteriori.

Pour simplifier, dans un essai thérapeutique classique, le risque alpha (« p ») correspond au risque de conclure à l'efficacité du traitement alors qu'il ne l'est pas ; le risque bêta est celui de conclure à l'absence d'effet secondaire alors qu'il y en a. On pourrait penser que plus « p » est petit, mieux c'est. Ce serait n'avoir pas compris les termes de la décision qui doivent être posés a priori et non en fonction du résultat. En effet, devant le choix entre H_0 et H_1 , l'expérimentateur doit prendre une décision par rapport aux conséquences d'un résultat du test non conforme à la réalité : Quel risque acceptez-vous de prendre pour traverser la route avec comme conséquence, un éventuel accident grave : 10 %, 5 %, 1 pour mille ? Sachant que si vous y arrivez vivant, il y a un cadeau de 100 euros qui vous attend ? Et si le cadeau est de 1 million d'euros, prendriez-vous le même risque ? Ce n'est pas après avoir traversé la route et en connaissant le résultat que vous prenez votre décision.

Dans l'exemple, on pose H_0 : la différence entre 20 et 35 % est due au hasard. H_1 est que cette différence n'est pas due au hasard. Le risque alpha consenti est 5 % ; on n'a pas planifié l'expérience, le risque bêta est non maîtrisé. Un test statistique adéquat, ici le χ^2 donne $\chi^2 = 5.6$, Degré de liberté = 1, $p < 0,02$. Si l'hypothèse nulle est vraie, la probabilité d'obtenir ce résultat est inférieure à 2 %, plus faible que le risque alpha consenti 5 %. On rejette l'hypothèse nulle. On conclut : la différence n'est pas due au hasard ; la différence est significative ; il existe un lien entre décès et traitement. Mais le traitement en est-il la cause et la différence observée a-t-elle une importance clinique ? Du fait du dispositif expérimental (essai randomisé en double aveugle), on peut raisonnablement attribuer cette différence au traitement. Des études ultérieures le confirmeront, certaines donneront la même conclusion, d'autres non. **Une méta analyse, technique statistique particulière [5] permettra de prendre en compte l'ensemble de ces différentes études** et d'en tirer des conclusions générales. D'autres plans expérimentaux peuvent être réalisés comme le cross over, ils font l'objet d'ouvrages spécialisés [2].

Interprétation des résultats

Les méthodes statistiques permettent uniquement de savoir si le lien ou la différence, observés ne sont pas dus au hasard. En revanche, elles ne permettent pas d'attribuer la causalité du phénomène ni de généraliser les résultats et ne disent rien sur l'importance en terme de santé sur le phénomène mis en évidence. Trois grands pièges sont à prendre en compte :

Les facteurs de confusion

Le traitement A est préféré au traitement B en cas d'antécédent d'infarctus. Une fréquence plus importante d'insuffisance cardiaque est observée avec le traitement A qu'avec le traitement B. Après un infarctus du myocarde, l'insuffisance cardiaque est plus fréquente que dans la population. Est-ce que le traitement est la cause de l'insuffisance cardiaque ou est-ce l'antécédent d'infarctus ? Les facteurs de confusion sont très fréquents et plus ou moins maîtrisés dans les études observationnelles. Ils sont mieux pris en compte dans les essais randomisés en double aveugle bien conduits ou par les analyses multivariées.



Les biais de sélection

Quand la population étudiée est différente de la population cible, on a un biais de sélection qui empêche de généraliser les résultats : l'essai a été mené sur des femmes diabétiques insulinodépendantes dans un service de CHU ; on ne peut pas étendre les résultats aux hommes ni aux femmes suivis en médecine de ville.

Les biais de classement

Un exemple : un test diagnostique est positif chez les malades dans 80 % des cas, il y a 20 % de faux négatifs. Ce test est négatif chez les sujets sains dans 90 % des cas (10 % de faux positifs). La fréquence de la maladie dans la population est de 10 %. Nous pouvons résumer ces données dans le tableau à 4 cases suivant, établi sur 1 000 sujets :

	Vrais Malades	Sujets Sains	Total
Sujets avec un test +	80	90	170
Sujets avec un test -	20	810	830
Total	100	900	1 000

Sur les 170 sujets qui seront positifs au test et donc considérés comme malades, seuls 80 le seront réellement. Cela va fausser nos résultats vis-à-vis de la recherche de facteurs influant la maladie ; les valeurs des statistiques comme l'OR, le RR... seront alors faussées. Si la fréquence de la maladie est 20 %, quelles sont les conséquences ? (240 sujets seront positifs au test dont 160 malades et 80 sujets sains).

Références

1. Kohler F. *Éléments de biostatistiques : Exercices corrigés, QCM et rappels de cours à l'usage des étudiants en médecine et en biologie*. Ellipses, 2011.
2. Dagnelie P. *Principes d'expérimentation: planification des expériences et analyse de leurs résultats*. Les Presses agronomiques de Gembloux, 2003.
3. Cucherat M. *les courbes de survie*. <http://www.spc.univ-lyon1.fr/polycop/courbes%20de%20survie.htm>. Consulté le 10 juillet 2012.
4. Hill C, Com-Nougé C, Kramar A, Moreau T, O'Quigley J, Senoussi R, et al. *Analyse statistique des données de survie*. Flammarion, 1990 : p 31-55 et p 97-99.
5. Cucherat M; Boissel JP; Leizorovicz A. *La méta-analyse des essais thérapeutiques*. 1 ed. Paris: Masson, 1997 : p 1-9.

Réalisé avec le soutien des Laboratoires Janssen.