

# Accordons-nous trop d'importance à la signification statistique?

par Lauren Bresee

Si vous repensez à votre cours d'introduction aux statistiques à l'université, vous vous souvenez peut-être du thème de la vérification des hypothèses, où l'hypothèse nulle précise qu'il n'existe aucune différence entre les groupes d'étude et que l'hypothèse alternative précise quant à elle qu'il en existe une<sup>1</sup>. Aujourd'hui, en tant que pharmacien praticien, vous menez un essai randomisé contrôlé (ERC) pour évaluer si une intervention que vous avez développée aide les patients à mieux respecter leur pharmacothérapie après leur sortie de l'hôpital. Les participants à l'étude seront répartis de manière aléatoire dans un groupe d'intervention ou dans un groupe recevant des soins habituels; le respect de la pharmacothérapie 90 jours après la sortie de l'hôpital constitue votre principal critère de jugement. Votre hypothèse nulle veut qu'il n'y a aucune différence d'observance entre les patients qui reçoivent votre intervention et ceux qui reçoivent des soins habituels 90 jours après leur sortie de l'hôpital; votre hypothèse alternative précise qu'il existe une différence d'observance entre votre groupe d'intervention et les groupes témoins. Pour évaluer votre critère principal, vous menez un test statistique par inférence pour décider si vous allez accepter ou rejeter votre hypothèse nulle. Selon ce que rapporte communément la littérature médicale, vous avez décidé d'utiliser une valeur seuil de probabilité ( $p$ ) de 0,05 pour déterminer si votre groupe d'intervention est statistiquement différent de votre groupe témoin, c'est-à-dire si la valeur  $p$  associée à votre test statistique est inférieure à 0,05, vous rejeterez l'hypothèse nulle et conclurez que la différence entre votre groupe d'intervention et le groupe témoin est significative sur le plan statistique.

Le concept de signification statistique et l'utilisation d'une valeur seuil  $p$  (et les intervalles de confiance de 95 % correspondants) pour établir la signification statistique suscitent depuis longtemps la controverse au sein de la communauté scientifique. Si vous êtes féru de statistiques comme moi, vous avez peut-être remarqué quelques publications récentes sur l'utilisation de la signification statistique dans la recherche. Dans le premier article, un éditorial publié en mars 2019 intitulé *Scientists rise up against statistical significance*, Amrhein et collab.<sup>2</sup> appellent à ne plus utiliser la signification statistique pour déterminer s'il existe une différence entre les groupes, parce que le concept est

fréquemment appliqué de façon dichotomique au lieu d'être évalué sur un continuum. Les auteurs précisent : « Soyons clairs concernant ce qui doit prendre fin : on ne devrait jamais conclure qu'il n'existe "pas de différence" ou "pas d'association" uniquement parce qu'une valeur  $P$  est supérieure à un seuil de 0,05 ou, similairement, parce qu'un intervalle de confiance comprend le zéro »<sup>2</sup>. Au lieu de cela, les auteurs proposent d'utiliser des intervalles de confiance, comme des « intervalles de compatibilité », c'est-à-dire que votre estimation ponctuelle et les limites de confiance sont les plus compatibles avec vos données à cause du modèle statistique que vous avez utilisé pour calculer vos résultats<sup>2</sup>. Pour revenir à l'ERC décrit ci-dessus, vous menez votre test statistique par inférence, et votre résultat donne un risque relatif de 2,0, avec un intervalle de confiance de 95 % de 1,5 à 2,5 et une valeur  $p$  bien en dessous de 0,001. En vertu de la déclaration du test statistique traditionnel, vous pourriez conclure que les personnes de votre groupe d'intervention étaient deux fois plus portées que celles de votre groupe témoin à respecter leur pharmacothérapie 90 jours après leur sortie de l'hôpital et que cette différence est statistiquement significative, car l'intervalle de confiance n'englobe pas la mesure d'équivalence de 1. Cependant, si vous deviez utiliser la proposition émise par Amrhein et collab.<sup>2</sup>, vous indiqueriez au lieu de cela que les valeurs du risque relatif, de l'intervalle de confiance de 95 % et de la valeur  $p$  la plus compatible avec vos données indiquent que les personnes ayant reçu votre intervention étaient deux fois plus portées à respecter leur prise de médicaments 90 jours après leur sortie de l'hôpital et que la différence de risque entre votre groupe de traitement et votre groupe témoin montrait que ce dernier était de 1,5 à 2,5 fois plus enclin à respecter le traitement, étant donné les suppositions du test statistique.

En réplique à l'éditorial d'Amrhein et collab.<sup>2</sup>, Ioannidis a publié un éditorial le mois suivant dans le *JAMA*, intitulé *The importance of predefined rules and prespecified statistical analyses : do not abandon significance*<sup>3</sup>. Il souligne dans son article que les décisions prises en médecine sont le plus souvent dichotomiques et qu'on met davantage l'accent sur les « affirmations inappropriées indiquant n'avoir trouvé aucune différence statistique, plutôt que sur les énoncés non fondés

indiquant une différence et un refus sans que ce rejet soit justifié», particulièrement quand les chercheurs ne développent pas ou ne respectent pas les règles des tests statistiques spécifiées au préalable<sup>3</sup>. Au lieu de bannir le concept de signification statistique, Ioannidis soulignait que les chercheurs devaient se concentrer sur le respect des règles des tests statistiques et faire en sorte que la pertinence clinique s'applique à la prise de décision<sup>3</sup>. Il est évident que ces deux éditoriaux portent surtout sur la mauvaise utilisation fréquente de la signification statistique et des valeurs  $p$ .

Qu'est-ce que la valeur  $p$  et que n'est-elle pas? L'American Statistical Association (ASA) la définit comme « la probabilité selon un modèle statistique spécifié, qu'un résumé statistique des données (p. ex., l'écart moyen de l'échantillon entre deux groupes de comparaison) serait égal à sa valeur obtenue par l'observation ou plus extrême que cette valeur »<sup>4</sup>. Le modèle statistique spécifié et les données utilisées dans le test statistique sont les éléments clés de cette définition sur lesquels se focaliser. Si on utilise un seuil de valeur  $p$  de 0,05 pour déterminer la signification statistique, cela signifie que la probabilité que le résultat obtenu soit dû à la chance est inférieure à 5 %, spécifiquement au modèle statistique et aux données utilisées pour le test. Dans le but de réduire la probabilité d'une mauvaise interprétation des valeurs  $p$ , le conseil d'administration de l'ASA a publié une déclaration sur la bonne et la mauvaise manière d'utiliser les valeurs  $p$ <sup>4</sup>. Bien qu'elles puissent être utilisées pour déterminer s'il existe des preuves réfutant l'hypothèse nulle, comme nous l'avons mentionné ci-dessus, de telles conclusions sont surtout applicables aux données utilisées et aux suppositions faites pour calculer la valeur  $p$ <sup>4</sup>. La valeur  $p$  ne décrit pas la puissance de la taille d'effet ou la précision de votre résultat (c'est-à-dire une valeur  $p$  plus faible ne reflète pas une plus grande taille d'effet ou une estimation plus précise) et ne représente pas la probabilité que l'hypothèse générale de l'étude soit vraie ou due à la chance lorsqu'elle est appliquée à la population examinée<sup>4</sup>.

Une controverse associée à la signification statistique et à l'utilisation des valeurs  $p$  existera vraisemblablement toujours. Il y a toutefois des concordances fondamentales entre l'éditorial d'Amrhein et collab.<sup>2</sup>, celui d'Ioannidis<sup>3</sup> et la déclaration de l'ASA<sup>4</sup> qui, si elles sont respectées, aideront à désamorcer la controverse. Il nous faut tout d'abord nous assurer que le plan statistique de chaque étude soit spécifié au préalable, transparent et applicable aux données de l'étude<sup>2-4</sup>. Deuxièmement, il faut que les résultats de tous les tests statistiques menés, y compris les

estimations ponctuelles, les mesures de précision (comme les intervalles de confiance) et les valeurs  $p$  soient rapportés. De tels rapports permettent d'être sûrs que nous ne reportons pas certains résultats de manière sélective et de tenir compte de l'évaluation de l'erreur de type 1, c'est-à-dire de trouver un résultat significatif d'un point de vue statistique là où il n'en existe aucun, et cela à cause de la multiplicité des tests statistiques<sup>2-4</sup>. Troisièmement, il convient d'être réaliste en interprétant les résultats de toute étude et d'éviter de leur donner trop ou trop peu d'importance<sup>2-4</sup>. Enfin, les décisions cliniques ne devraient jamais se baser uniquement sur les résultats statistiques. Il faut prendre en considération d'autres facteurs, y compris la validité de l'étude, la cohérence de ses résultats par rapport aux informations disponibles et la possibilité de les généraliser à l'ensemble de la population étudiée<sup>2-4</sup>. L'observation de ces recommandations favorisera une utilisation appropriée des tests statistiques dans la recherche, pour que nous puissions prendre les meilleures décisions cliniques possibles pour nos patients.

[Traduction par l'éditeur]

#### Références

1. Gaddis GM, Gaddis ML. Introduction to biostatistics: part 3, sensitivity, specificity, predictive value, and hypothesis testing. *Ann Emerg Med.* 1990; 19(5):591-7.
2. Amrhein V, Greenland S, McShane B. Scientists rise up against statistical significance [éditorial]. *Nature.* 2019;567(7748):305-7.
3. Ioannidis JPA. The importance of predefined rules and prespecified statistical analyses: do not abandon significance [éditorial]. *JAMA.* 2019;321(21): 2067-8.
4. Wasserstein RL. ASA statement on statistical significance and  $P$ -values. *Am Stat.* 2016;70(2):131-3.

---

**Lauren Bresee**, B. Sc. Pharm., ACPR, M. Sc., Ph. D., est conseillère scientifique auprès de l'Agence canadienne des médicaments et des technologies de la santé (ACMTS) à Ottawa (Ontario); professeure agrégée adjointe au Département des sciences de la santé communautaire, Faculté de médecine, Université de Calgary (Alberta); et membre de l'Institut O'Brien de la santé publique de l'Université de Calgary. Elle est également rédactrice adjointe du *Journal canadien de la pharmacie hospitalière*.

**Conflits d'intérêts** : Aucune déclaration.

#### Adresse de correspondance :

D<sup>re</sup> Lauren Bresee  
Agence canadienne des médicaments et des technologies  
de la santé (ACMTS)  
865 avenue Carling, bureau 600  
Ottawa ON K1S 5S8

**Courriel** : LaurenB@cadth.ca