

Faut-il diminuer la valeur de p à $p < 0,005$?

Should we reduce the value of p to < 0.005 ?



**Pr Philippe
Gabriel Steg**



Rédacteur en chef; service de cardiologie,
hôpital Bichat, AH-HP; université Paris-
Diderot, INSERM U1148, Paris.

*“La démocratie est la pire forme de gouvernement,
à l’exception de toutes les autres.”*
Winston S. Churchill

Aujourd’hui, tous les médecins et toutes les personnes s’intéressant à la recherche et aux statistiques connaissent la valeur “magique” d’un $p < 0,05$. Celle-ci matérialise le fait que l’on a 5 chances sur 100 de rejeter à tort l’hypothèse nulle dans un test statistique compte tenu des données observées. Elle est utilisée presque universellement en recherche médicale, qu’elle soit clinique ou expérimentale. Pourtant, la recherche connaît un sérieux problème de reproductibilité et de nombreux statisticiens et méthodologistes s’inquiètent de ce que l’on nomme le “*p-hacking*”, c’est-à-dire la tendance qu’ont les chercheurs à rechercher à tout prix une valeur de p statistiquement significative dans leurs études, parfois par le biais de comparaisons multiples ou déterminées a posteriori. Un célèbre statisticien gréco-américain, le docteur John Ioannidis, a récemment suscité un débat dans la communauté scientifique en proposant d’abandonner la valeur seuil de $p < 0,05$ pour affirmer la significativité statistique d’un test et l’abaisser à $p < 0,005$ [1]. Il s’agit d’une proposition délibérément provocante dont le docteur Ioannidis est coutumier. Il avait déjà publié un éditorial célèbre suggérant que la grande majorité des résultats expérimentaux publiés était fausse [2]. Il est vrai que devant l’inflation des résultats positifs et parfois faussement positifs, une plus grande rigueur méthodologique et statistique serait bienvenue dans l’interprétation des résultats expérimentaux, surtout lorsqu’il n’est pas toujours possible de vérifier a posteriori que l’hypothèse testée est bien celle qui avait été émise avant l’expérimentation. Ainsi, dans le cadre des essais cliniques, l’enregistrement préalable des protocoles avant leur démarrage, devenu désormais obligatoire pour toute publication scientifique, minimise le risque de *p-hacking*, sans néanmoins empêcher les chercheurs de viser à tout prix l’optimisation de leurs résultats par l’étude de critères de jugement déterminés a posteriori ou la focalisation sur des sous-groupes où le résultat paraît particulièrement intéressant. Faut-il pour autant suivre le docteur Ioannidis ?

Abaisser le seuil de signification statistique permet d’adresser le risque statistique de première espèce, c’est-à-dire la crainte d’affirmer à tort un résultat comme positif, alors qu’il est en réalité négatif. Mais elle ne règle en rien le risque de deuxième espèce, c’est-à-dire celui de conclure à l’absence d’effet, alors qu’il en existe et

elle n'est pas plus gage de sérieux méthodologique qu'un p à 0,005. Aucune décision arbitraire sur les seuils de signification statistique ne permettra de s'affranchir des règles élémentaires qui consistent à poser une hypothèse a priori pour la tester et essayer de la valider expérimentalement. Il est probable aussi qu'il faudrait songer sérieusement à passer d'une interprétation binaire des études cliniques à une interprétation plus réaliste et vraisemblablement plus humble des données des essais cliniques : considérer un traitement comme inefficace et le rejeter parce que la valeur du test de p est à 0,06 est probablement aussi critiquable que de l'accepter aveuglément parce que la valeur du p est à 0,04. Dans la communauté des statisticiens et des méthodologistes, beaucoup plaident pour les méthodes bayésiennes [3]. Celles-ci permettent une interprétation plus nuancée qui prend en compte les probabilités a priori et le savoir existant, donnant une gamme de probabilités plus qu'un résultat binaire, et constituent certainement l'avenir des essais cliniques. Enfin, en dépit des problèmes liés à leur interprétation statistique, les essais cliniques prospectifs randomisés restent la moins mauvaise méthode pour évaluer rigoureusement les méthodes diagnostiques ou les traitements innovants. Malgré la tentation de se reposer sur des analyses observationnelles plus faciles et moins coûteuses, force est de constater que celles-ci sont pleines de pièges et ont conduit dans le passé à plusieurs reprises à des erreurs manifestes. Nous avons besoin d'accroître le nombre d'essais randomisés, pas de le diminuer. Or, augmenter la valeur du seuil statistique va obliger à augmenter la taille des essais cliniques et, par là même, rendre la conduite d'essais plus difficile, plus longue et plus coûteuse. En définitive, on peut craindre que cela aboutisse à une diminution du nombre des essais et, donc, des hypothèses qui seront testées. La recherche médicale a, de fait, besoin de plus de moyens pour conduire plus d'essais et obtenir plus de découvertes, plutôt que de solutions simples mais probablement simplistes fondées sur le changement des "règles du jeu" statistiques [4].

1. Ioannidis JPA. The proposal to lower p value thresholds to .005. *JAMA* 2018;319(14):1429-30.

2. Ioannidis JPA. Why most published research findings are false. *PLoS Med* 2005;2:e124.

3. Quintana M, Viele K, Lewis RJ. Bayesian analysis: using prior information to interpret the results of clinical trials. *JAMA* 2017;318:1605-6.

4. Kyriacou DN. The enduring evolution of the p value. *JAMA* 2016;315:1113-5.

L'auteur déclare avoir des liens d'intérêts :

- pour des bourses de recherche avec Bayer, Merck, Sanofi et Servier;
- pour des honoraires (orateur ou consultant), avec Amarin, Amgen, AstraZeneca, Bayer/Janssen, Boehringer-Ingelheim, Bristol-Myers-Squibb, Lilly, Merck, Novartis, Novo-Nordisk, Pfizer, Regeneron, Sanofi, Servier.

AVIS AUX LECTEURS

Les revues Edimark sont publiées en toute indépendance et sous l'unique et entière responsabilité du directeur de la publication et du rédacteur en chef. Le comité de rédaction est composé d'une dizaine de praticiens (chercheurs, hospitaliers, universitaires et libéraux), installés partout en France, qui représentent, dans leur diversité (lieu et mode d'exercice, domaine de prédilection, âge, etc.), la pluralité de la discipline. L'équipe se réunit 2 ou 3 fois par an pour débattre des sujets et des auteurs à publier. La qualité des textes est garantie par la sollicitation systématique d'une relecture scientifique en double aveugle, l'implication d'un service de rédaction/révision in situ et la validation des épreuves par les auteurs et les rédacteurs en chef.

Notre publication répond aux critères d'exigence de la presse :

- accréditation par la CPPAP (Commission paritaire des publications et agences de presse) réservée aux revues sur abonnements,
- adhésion au SPEPS (Syndicat de la presse et de l'édition des professions de santé),
- indexation dans la base de données internationale ICMJE (International Committee of Medical Journal Editors),
- déclaration publique de liens d'intérêts demandée à nos auteurs,
- identification claire et transparente des espaces publicitaires et des publi-rédactionnels en marge des articles scientifiques.