LES MAINS DANS LES PREUVES

La taille d'effet : ce que la valeur *p* ne vous dit pas

François Tharin
MSc, filière physiothérapie, Haute école de santé (HEdS), Haute École de santé Genève, HES-SO Haute École spécialisée de Suisse occidentale, Genève, Suisse. Université Savoie Mont Blanc, Laboratoire Interuniversitaire de Biologie de la Motricité, Chambéry, France

Mains Libres 2026 ; 3 : 233-236 | DOI : 10.55498/MAINSLIBRES.2026.02.3.233

En pratique clinique, il est fréquent de devoir choisir entre plusieurs approches thérapeutiques. Il n'est pas toujours évident de déterminer laquelle privilégier. Les résultats de la recherche constituent l'un des éléments qui permettent d'éclairer cette décision[1,2]. Cependant, une différence statistiquement significative suffit-elle à guider une décision ? La valeur p indique dans quelle mesure les données observées sont compatibles avec une hypothèse nulle, mais elle ne renseigne pas sur l'ampleur ni sur l'importance clinique de la différence observée[3]. Cet article vise à clarifier ce que les calculs de taille d'effet apportent à l'interprétation des résultats de recherche, au raisonnement clinique et à la prise de décision.

Pourquoi la valeur *p* ne suffit pas

Une différence entre deux groupes peut être statistiquement significative tout en étant trop faible pour être cliniquement utile[4,5]. Dans un grand échantillon, un effet minime peut atteindre le seuil de $p < 0,05$[3,5]. À titre d'exemple, une amélioration moyenne de 0,3 point sur une échelle de douleur de 0 à 10 pourrait devenir statistiquement significative dans un très grand échantillon, sans que cela suffise à établir sa pertinence clinique. Une conclusion limitée à « $p < 0,05$ » risque donc d'exagérer la portée clinique d'un résultat[5].

L'inverse est également possible. Une étude peut ne pas atteindre la significativité statistique parce qu'elle manque de puissance ou parce que l'incertitude autour de l'effet est trop importante[3]. Dans une analyse de la littérature en réadaptation, de nombreuses études présentaient une puissance insuffisante pour détecter des tailles d'effets habituellement observées dans ce domaine[6]. Un résultat avec $p > 0,05$ ne devrait donc pas être interprété automatiquement comme une preuve d'absence d'effet[3].

Il a récemment été rappelé, dans un article méthodologique consacré à la différence minimale cliniquement importante (MCID), que la significativité statistique ne garantit pas qu'une personne aille réellement mieux[7]. Il a également été souligné que la MCID constitue un indicateur utile de changement individuel, mais que son application à une différence moyenne entre groupes doit tenir compte de la méthode utilisée pour déterminer la MCID et du niveau d'analyse considéré[8–10]. Cette mise au point est importante, car elle montre que l'interprétation clinique d'un résultat dépend non seulement de sa valeur numérique, mais aussi du niveau d'analyse auquel cette valeur s'applique.

Cette réflexion ouvre une question complémentaire. Lorsqu'une étude compare deux interventions, il ne suffit pas de savoir si une différence est statistiquement significative. Il faut également savoir quelle est l'ampleur de cette différence, avec quelle incertitude elle est estimée, et dans quelle mesure elle pourrait influencer une décision clinique[3]. La taille d'effet occupe précisément cette place. Elle permet de passer d'une lecture centrée sur la détection statistique à une lecture centrée sur l'ampleur de l'effet[3].

Définir la taille d'effet : quantifier l'ampleur d'un phénomène

Une taille d'effet est une estimation quantitative de l'ampleur d'un phénomène observé[5]. Elle peut notamment exprimer une différence entre deux groupes, une association entre deux variables ou une différence de probabilité de survenue d'un événement[5]. La question à laquelle répond une taille d'effet n'est donc pas : « le résultat est-il significativement différent ? », mais « quelle est l'ampleur de l'effet, de la différence ou de l'association observés ? »[3]. Cette distinction est centrale, car deux résultats ayant la même valeur p peuvent correspondre à des effets d'ampleurs très différentes[5].

Les tailles d'effet peuvent être exprimées dans l'unité naturelle de l'outil de mesure, ce qui permet une interprétation directe du résultat[3,11]. À titre d'exemples illustratifs, une diminution de deux points sur une échelle numérique de douleur entre 0 et 10, un gain de 30 degrés de flexion du genou, ou une amélioration de 5 points au *Neck Disability Index*[12] sont des résultats directement interprétables. Lorsque l'unité de mesure est familière, cette forme d'expression facilite l'interprétation du résultat et son intégration dans le raisonnement clinique[11].

Il est également possible de standardiser la taille d'effet[3] : son calcul et son interprétation varient selon le type de taille d'effet considéré[3,5]. Une **taille d'effet de différence entre groupes** quantifie l'ampleur d'une différence entre deux groupes ou deux conditions[5]. Le *d* de Cohen et le *g* de Hedges expriment une différence standardisée en unités d'écart-type[5,13]. Cette expression sans unité facilite la comparaison de résultats obtenus avec des instruments différents lorsque ceux-ci évaluent un construit comparable[3]. Cette approche est particulièrement utile dans les méta-analyses, lorsque les résultats de plusieurs études doivent être combinés malgré l'hétérogénéité des instruments de mesure[3]. Le *g* de Hedges est une version corrigée

du *d* de Cohen qui réduit son biais d'estimation dans les petits échantillons[14].

Une **taille d'effet fondée sur une corrélation** quantifie la force d'une association entre deux variables, par exemple entre la douleur et la capacité fonctionnelle[5]. L'exemple le plus courant est le coefficient de corrélation de Pearson (r)[5]. Une valeur proche de 0 indique une association linéaire faible ou absente dans l'échantillon, sans exclure l'existence d'une relation non linéaire, tandis que des valeurs proches de -1 ou de +1 reflètent une association plus forte, respectivement négative ou positive[15]. Comme toute estimation, le coefficient r gagne à être interprété avec son intervalle de confiance[5].

Enfin, pour un critère de jugement binaire, les tailles d'effet peuvent être exprimées sous forme de différence absolue de risque, de risque relatif ou d'odds ratio[5]. Ces différents types de tailles d'effet répondent donc à des questions différentes et ne s'interprètent pas de la même manière[13].

Interpréter une taille d'effet

Quelle est l'ampleur de l'effet ?

Pour les **tailles d'effet de différence entre groupes**, les repères classiquement proposés pour interpréter le *d* de Cohen sont souvent résumés ainsi : environ 0,2 pour un effet faible, 0,5 pour un effet modéré et 0,8 pour un effet important[16]. Ces valeurs offrent un premier repère d'interprétation, mais ne doivent pas être utilisées comme des seuils universels[13]. Ainsi, des repères empiriques plus spécifiques à la physiothérapie ont été proposés pour les tailles d'effet de différence entre groupes exprimées par le *d* de Cohen ou le *g* de Hedges, avec des valeurs de 0,1 pour un petit effet, 0,4 pour un effet moyen et 0,8 pour un grand effet[13].

Pour les **tailles d'effet fondées sur une corrélation**, qui décrivent l'intensité d'une association entre deux variables plutôt qu'une différence entre deux groupes, les repères conventionnels sont souvent résumés par des valeurs de 0,1 pour un petit effet, 0,3 pour un effet moyen et 0,5 pour un grand effet[16]. Comme pour les différences standardisées, ces valeurs ne constituent pas des seuils universels[13]. En physiothérapie, des repères empiriques plus spécifiques ont récemment été proposés : 0,3 pour un petit effet, 0,5 pour un effet moyen et 0,6 pour un grand effet[13].

Dans quelle unité l'effet est-il exprimé ?

Une taille d'effet standardisée facilite les comparaisons statistiques, mais peut être moins directement interprétable sur le plan clinique[3,11]. Son interprétation peut être facilitée lorsqu'elle est rapportée dans l'unité naturelle de la variable étudiée, par exemple en points de douleur, en degrés d'amplitude articulaire ou en mètres parcourus[5,11].

Cette approche est illustrée dans une méta-analyse sur la physiothérapie après un accident vasculaire cérébral, dans laquelle les effets ont été rapportés sous forme standardisée et sous forme de différences exprimées dans leurs unités naturelles, notamment la vitesse de marche et la distance parcourue[17]. Pour qu'un effet soit comparé numériquement à un seuil clinique, les deux doivent notamment être exprimés dans la même unité.

Quelle est la précision de l'estimation ?

Une taille d'effet devrait être interprétée avec son intervalle de confiance, qui renseigne sur la précision de l'estimation[3]. L'effet ponctuel donne la meilleure estimation disponible à partir de l'échantillon observé ; l'intervalle de confiance renseigne sur l'incertitude autour de cette estimation et sur l'étendue des valeurs compatibles avec les données[3]. Un intervalle étroit traduit une estimation plus précise, tandis qu'un intervalle large traduit une plus grande incertitude quant à l'ampleur de l'effet[5].

L'intervalle de confiance inclut-il la valeur de non-effet ?

Pour une différence de moyennes, la valeur de non-effet est 0 : si l'intervalle de confiance inclut 0, les données restent compatibles avec une absence de différence entre les groupes[5]. Pour un ratio, comme un risque relatif ou un odds ratio, la valeur de non-effet est 1 : si l'intervalle de confiance inclut 1, les données restent compatibles avec une absence de différence relative entre les groupes[5]. Cette lecture est utile, mais elle ne suffit pas à juger l'importance clinique du résultat[5,11].

L'effet dépasse-t-il l'erreur de mesure de l'instrument ?

Le changement minimal détectable (MDC) renseigne sur l'ampleur du changement nécessaire pour dépasser l'erreur de mesure de l'instrument[8,18]. Il peut être défini au niveau de l'individu ou du groupe[18]. Au niveau individuel, un changement supérieur au MDC peut être distingué de l'erreur de mesure avec le niveau de confiance retenu, tandis qu'un changement inférieur ne permet pas de déterminer si un changement réel s'est produit[8,18]. Au niveau du groupe, le MDC est plus faible que chez un individu, car le seuil groupal diminue en fonction de la racine carrée de l'effectif[18]. Ainsi, lorsqu'une étude rapporte un changement moyen au niveau d'un groupe, son interprétation au regard de l'erreur de mesure nécessite de considérer un seuil adapté au niveau groupal plutôt que le MDC individuel[18,19]. Cette comparaison nécessite que le changement estimé et le MDC soient exprimés dans la même unité de mesure[18].

L'effet atteint-il un seuil d'importance clinique ?

Dépasser l'erreur de mesure ne signifie pas nécessairement que le changement observé est important sur le plan clinique[8,20]. Des seuils d'importance clinique, tels que la différence minimale importante ou la MCID, peuvent être déterminés à partir de changements intra-individuels et ne constituent donc pas automatiquement des seuils appropriés pour interpréter une différence moyenne entre groupes[8,10]. Avant d'utiliser un tel seuil pour interpréter un effet entre deux interventions, il faut donc vérifier qu'il est adapté à une différence entre groupes et non uniquement à un changement intra-individuel[10,21].

Il n'existe toutefois pas de méthode consensuelle unique pour déterminer le seuil d'importance clinique d'une différence entre groupes[10]. Le *smallest worthwhile effect* a été proposé à cette fin : il correspond au plus petit bénéfice d'une intervention jugé suffisamment important au regard de ses risques, coûts et contraintes, comparativement à une autre intervention[10]. Il peut notamment être déterminé à

partir des préférences des patientes et des patients par des méthodes de compromis entre bénéfices et inconvénients[10].

Lorsqu'un seuil d'importance clinique adapté à une différence entre groupes est disponible, l'intervalle de confiance de l'effet estimé peut être interprété par rapport à ce seuil[11,22]. Lorsque des valeurs plus élevées correspondent à un effet plus favorable, si l'ensemble de l'intervalle se situe en deçà du seuil, les valeurs compatibles avec les données ne l'atteignent pas[22]. Si l'intervalle comprend le seuil, les données restent compatibles avec des effets situés de part et d'autre de celui-ci[22]. Enfin, si l'ensemble de l'intervalle se situe au-delà du seuil, les valeurs compatibles avec les données atteignent ou dépassent ce seuil[22].

La valeur d'un seuil d'importance clinique peut varier selon la population étudiée, la méthode utilisée pour le déterminer et le contexte clinique, et ne constitue donc pas une propriété fixe et universelle d'un instrument de mesure[20,23]. Le fait que l'intervalle de confiance atteigne ou dépasse un tel seuil ne suffit donc pas, à lui seul, à établir la pertinence clinique globale d'une intervention[21].

Ces résultats peuvent-ils éclairer ma décision clinique ?

En consultant une taille d'effet afin d'argumenter une décision clinique, il faut considérer la crédibilité globale de l'étude, notamment son design, le risque de biais, la précision des estimations, la cohérence des résultats et la pertinence du critère de jugement[11]. L'applicabilité du résultat dépend ensuite de sa proximité avec la situation clinique considérée, notamment en ce qui concerne la population étudiée, l'intervention, le comparateur et le contexte de soins[2,11].

La taille d'effet permet notamment de quantifier l'ampleur de la différence observée entre les interventions comparées[11]. Son interprétation doit donc tenir compte du comparateur utilisé dans l'étude[11]. De même, un effet observé dans un contexte expérimental ne peut pas être automatiquement transposé à une pratique de terrain, où les contraintes de temps, les ressources disponibles, l'organisation du soin, les attentes de la personne et ses objectifs influencent la décision[2,24,25]. Une amélioration observée après une intervention ne peut ainsi pas être automatiquement interprétée comme l'effet propre du traitement[11,26]. Elle peut notamment refléter l'évolution naturelle, la régression vers la moyenne, les attentes de la personne ou les effets contextuels[11,26].

Ainsi interprétée, la taille d'effet fournit une estimation quantitative qui peut être mise en regard des bénéfices attendus, des risques et contraintes de l'intervention ainsi que des objectifs et préférences de la personne concernée[2,21]. Elle constitue ainsi un élément susceptible d'éclairer le choix entre plusieurs options thérapeutiques, sans se substituer au jugement clinique[2].

Conclusion

La taille d'effet ne remplace pas le jugement clinique, mais elle permet de mieux l'argumenter. Au-delà de la significativité statistique, elle aide à apprécier l'ampleur d'une différence et l'incertitude qui l'entoure. Mise en regard du comparateur, de la qualité des données et de la pertinence clinique du résultat, elle constitue ainsi une information utile pour éclairer le choix entre plusieurs options thérapeutiques.

Contact
François Tharin
francois.tharin@hesge.ch

Intelligence artificielle

La recherche documentaire a été menée conjointement sur les bases de données traditionnelles et via le moteur de recherche Consensus.app (Consensus, version web, consultée entre juin et juillet 2026) ; par ailleurs, ChatGPT (OpenAI, GPT-5.6 Sol, entre juillet et août 2026) a été utilisé pour optimiser la clarté et la lisibilité de la rédaction. Après avoir utilisé ces outils/services, l'auteur a revu et corrigé le contenu et assume l'entière responsabilité du contenu de la publication.

Références

1. Desarzens AS, Tharin F. Développer le raisonnement clinique pour préparer la relève de la physiothérapie à l'autonomie et à la complexité des soins : un article de synthèse. Mains Libres. 2025;42(4):294–303.

2. Veras M, Kairy D, Paquet N. What Is Evidence-Based Physiotherapy? Physiotherapy Canada. 2016 May;68(2):95–6.

3. Williams S, Carson R, Tóth K. Moving beyond *P* values in *The Journal of Physiology*: A primer on the value of effect sizes and confidence intervals. J Physiol. 2023 Dec;601(23):5131–3.

4. McGlothlin AE, Lewis RJ. Minimal Clinically Important Difference: Defining What Really Matters to Patients. JAMA. 2014 Oct 1;312(13):1342–3.

5. Schober P, Bossers SM, Schwarte LA. Statistical Significance Versus Clinical Importance of Observed Effect Sizes: What Do P Values and Confidence Intervals Really Represent? Anesth Analg. 2018 Mar;126(3):1068–72.

6. Kinney AR, Eakman AM, Graham JE. Novel Effect Size Interpretation Guidelines and an Evaluation of Statistical Power in Rehabilitation Research. Arch Phys Med Rehabil. 2020 Dec;101(12):2219–26.

7. Nicolo P. La différence minimale cliniquement importante : outil utile… ou piège méthodologique ? Mains Libres. 2026;41(1):79–80.

8. Grönkvist R, Vixner L, Äng B, Grimby-Ekman A. Measurement Error, Minimal Detectable Change, and Minimal Clinically Important Difference of the Short Form-36 Health Survey, Hospital Anxiety and Depression Scale, and Pain Numeric Rating Scale in Patients With Chronic Pain. J Pain. 2024 Sep;25(9):104559.

9. Malec JF, Ketchum JM. A Standard Method for Determining the Minimal Clinically Important Difference for Rehabilitation Measures. Arch Phys Med Rehabil. 2020 Jun;101(6):1090–4.

10. Willigenburg NW, Poolman RW. The difference between statistical significance and clinical relevance. The case of minimal important change, non-inferiority trials, and smallest worthwhile effect. Injury. 2023 Oct;54:110764.

11. Page P. Beyond statistical significance: clinical interpretation of rehabilitation research literature. Int J Sports Phys Ther. 2014 Oct;9(5):726–36.

12. Tharin F, Reinmann A, Bruyneel AV, Demoulin C. Neck Disability Index : mesure de l'impact fonctionnel des douleurs cervicales. Kinesither Rev. 2025 May;25(281):39–43.

13. Zieliński G. Effect Size Guidelines for Individual and Group Differences in Physiotherapy. Arch Phys Med Rehabil. 2025 Dec;106(12):1844–9.

14. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Front Psychol. 2013;4:863.

15. Altman DG. Practical statistics for medical research. Boca Raton, FL: Chapman & Hall/CRC; 1999. 611 p.

16. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. reprint. New York, NY: Psychology Press; 2009. 567 p.

17. Ferrarello F, Baccini M, Rinaldi LA, Cavallini MC, Mossello E, Masotti G, et al. Efficacy of physiotherapy interventions late after stroke: a meta-analysis. J Neurol Neurosurg Psychiatry. 2011 Feb 1;82(2):136–43.

18. Ohno S, Takahashi K, Inoue A, Takada K, Ishihara Y, Tanigawa M, et al. Smallest detectable change and test-retest reliability of a self-reported outcome measure: Results of the Center for Epidemiologic Studies Depression Scale, General Self-Efficacy Scale, and 12-item General Health Questionnaire. J Eval Clin Pract. 2017 Dec;23(6):1348–54.

19. Furlan L, Sterr A. The Applicability of Standard Error of Measurement and Minimal Detectable Change to Motor Learning Research—A Behavioral Study. Front Hum Neurosci. 2018 Mar 22;12:95.

20. Sedaghat AR. Understanding the Minimal Clinically Important Difference (MCID) of Patient-Reported Outcome Measures. Otolaryngol Head Neck Surg. 2019 Oct;161(4):551–60.

21. Innocenti T, Schleimer T, Salvioli S, Giagio S, Ostelo R, Chiarotto A. In trials of physiotherapy for chronic low back pain, clinical relevance is rarely interpreted, with great heterogeneity in the frameworks and thresholds used: a meta-research study. J Physiother. 2024 Jan;70(1):51–64.

22. De Cassai A. The ratio of confidence interval above the minimal clinically important difference. Anaesthesia. 2026 Mar;81(3):453.

23. Çelik D, Çoban Ö, Kılıçoğlu Ö. Minimal clinically important difference of commonly used hip-, knee-, foot-, and ankle-specific questionnaires: a systematic review. J Clin Epidemiol. 2019 Sep;113:44–57.

24. Gleadhill C, Bolsewicz K, Davidson SRE, Kamper SJ, Tutty A, Robson E, et al. Physiotherapists' opinions, barriers, and enablers to providing evidence-based care: a mixed-methods study. BMC Health Serv Res. 2022 Nov 21;22(1):1382.

25. Paci M, Faedda G, Ugolini A, Pellicciari L. Barriers to evidence-based practice implementation in physiotherapy: a systematic review and meta-analysis. Int J Qual Health Care. 2021 Jun 26;33(2):mzab093.

26. Ezzatvar Y, Dueñas L, Balasch-Bernat M, Lluch-Girbés E, Rossettini G. Which Portion of Physiotherapy Treatments' Effect Is Not Attributable to the Specific Effects in People With Musculoskeletal Pain? A Meta-Analysis of Randomized Placebo-Controlled Trials. J Orthop Sports Phys Ther. 2024 Jun;54(6):391–9.