

Mauvais emploi des statistiques en pharmacologie expérimentale

Sara Sadozai-Slama¹, Romain-Daniel Gosselin¹

Dans cette étude, les auteurs ont quantifié les articles qui ne respectaient pas les consignes statistiques parus dans six journaux en pharmacologie en 2014. Les résultats montrent que les pharmacologistes expérimentaux utilisent mal les biostatistiques et suggèrent une fréquence élevée de résultats faux positifs et faux négatifs dans les publications.

Les biostatistiques sont indissociables des sciences de la vie. Leur bon usage réduit les faux positifs et faux négatifs et assure la reproductibilité expérimentale. Pourtant, les biologistes comprennent et utilisent mal les concepts de base en biostatistiques, d'où une inquiétude croissante sur la reproductibilité des données. Dans cette étude, les auteurs ont quantifié les articles qui ne respectaient pas les consignes statistiques dans six périodiques en pharmacologie parus en 2014.

Il apparaît que des designs/analyses erronés sont fréquents, dont la non mention des limites d'intervalles de confiance 95% (i.e. seule mention des valeurs p) pour conclure (98% des articles), une faible puissance ou une taille d'échantillon inconnue (73%), des violations des prérequis paramétriques (86%) et une absence de correction pour comparaisons multiples (23%). D'autres erreurs incluent l'usage de

SEM (sommées des écarts à la moyenne) pour montrer la variabilité (72%), des valeurs p imprécises (86%), des barres d'erreurs inconnues (20%) et une absence de paragraphe statistique (2%).

De plus, le respect des consignes n'était pas corrélé à l'impact factor ($R^2_{\text{TOTAL}}=0.02$; $R^2_{\text{DESIGN/ANALYSE}}=0.03$; $R^2_{\text{PRESENTATION}}=0.00$). Ces résultats montrent que les pharmacologistes expérimentaux utilisent mal les biostatistiques et suggèrent une fréquence élevée de résultats faux positifs et faux négatifs dans les publications.

Méthodes

Six journaux ont été inclus: Neuropsychopharmacology, British Journal of Pharmacology, Biochemical Pharmacology, Neuropharmacology, BMC Pharmacology and Toxicology et European Journal of Pharmacology. Leurs archives ont été scannées et les 9 à 10 articles les plus récents ont été téléchargés (1^{er} avril 2014).

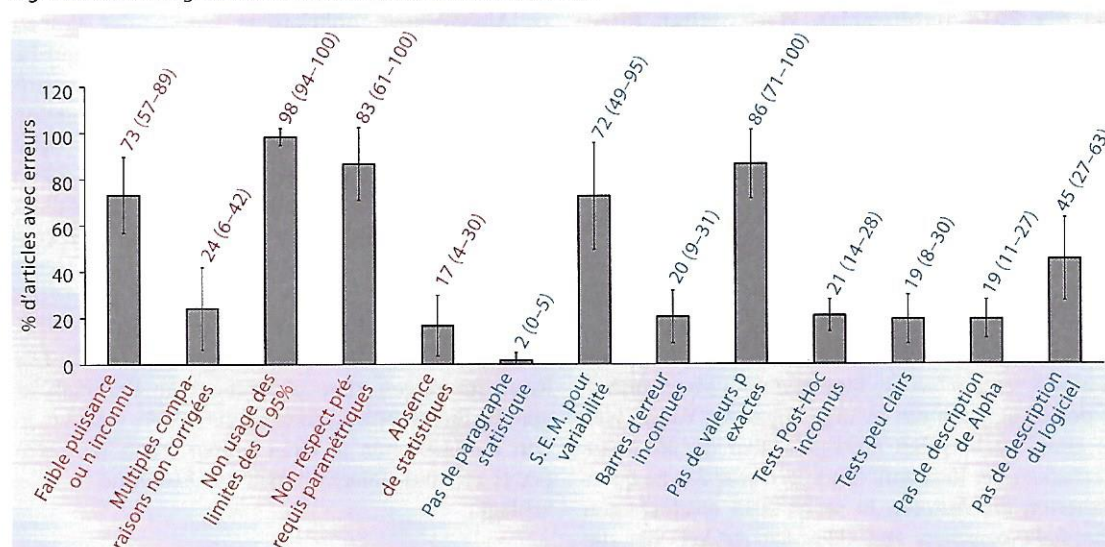
Les critères d'inclusion étaient:

- Articles, lettres ou communications courtes déjà publiés (exclusion quand sous presse ou épreuves non encore corrigées);
- Au moins une figure ou tableau avec une quantification;



Romain-Daniel Gosselin, de Biotelligences LLC, a reçu le Prix spécial lors du dernier Swiss Pharma Science Day pour son poster «Misuse of Statistics in Experimental Pharmacology – An Evaluation of Six Journals». Cet Award était sponsorisé par Vifor Pharma Fribourg.

Figure1: Pourcentages d'erreurs en fonction des différents critères



En rouge, les erreurs de design et d'analyse. En bleu, les erreurs de présentation des données ou de déclaration. Représentation en moyenne $\pm$ intervalle de confiance (C.I.) 95%. Taille d'échantillon: 6 dans toutes les séries.

¹ Biotelligences LLC

- Approche expérimentale sur animal, ou cellules humaines, in vitro/ex vivo ou in silico (les articles cliniques ou épidémiologiques ont donc été exclus);
- Figures supplémentaires non analysées;
- Les méthodes supplémentaires ont été analysées.

Les critères d'évaluation ont été adaptés de Curran-Everett and Benos D.J. 2004 J. Neurophysiol et Bailar J.C., Mosteller F. 1988 Ann Int Med. Chaque critère a été scoré 0 (respect du critère dans toutes les figures et tableaux) ou 1 (erreur détectée dans au moins une figure ou tableau). Les analyses n'ont pas été réalisées en aveugle vis-à-vis de l'identité des auteurs ou de leur affiliation.

Résultats et conclusions

Les consignes statistiques ne sont pas respectées et les erreurs sont fréquentes en pharmacologie expérimentale. Les faibles puissances, multiples comparaisons non corrigées et l'usage inapproprié de tests paramétriques suggèrent l'existence de fréquents faux négatifs et faux positifs. De plus, une mauvaise

présentation (tests mal décrits, tests, inconnus, mauvaises barres d'erreur, valeurs p exactes non données, tests post-hoc* inconnus ou même absence de paragraphe statistique) empêche la vérification des données et compromet la reproductibilité.

Des mesures simples peuvent être immédiatement prises pour améliorer la prise de conscience et la compréhension des concepts basiques en biostatistiques chez les chercheurs tels que:

- L'importance de la puissance;
- Les mises en garde sur les comparaisons multiples;
- Les grandes limites des valeurs p;
- Et les différences entre les barres d'erreurs.

Les éditeurs des journaux scientifiques devraient plus activement s'impliquer pour améliorer la qualité statistique des publications. ■

* Permet de comparer des groupes sans qu'une hypothèse sur la relation entre ces groupes ait été posée avant d'examiner les données.

Adresse de correspondance

Romain-Daniel Gosselin

E-Mail: rdg@biotelligences.com

Posterpreise vom 7. Swiss Pharma Science Day

Unkorrekte Statistiken in der experimentellen Pharmakologie

Sara Sadozai-Slama¹, Romain-Daniel Gosselin¹



Romain-Daniel Gosselin, Biotelligences LLC, hat am diesjährigen Swiss Pharma Science Day für sein Poster «Misuse of Statistics in Experimental Pharmacology – An Evaluation of Six Journals» den Spezialpreis erhalten. Vifor Pharma Fribourg sponserte diesen Award.

In dieser Studie quantifizierten die Autoren in sechs Pharmakologie-Zeitschriften die Artikel, die die 2014 publizierten statistischen Richtlinien nicht einhielten. Die Untersuchungen zeigen, dass die experimentellen Pharmakologen Biostatistiken falsch anwenden und dadurch eine hohe Frequenz falsch positiver und falsch negativer Resultate in den Publikationen generieren.

Biostatistiken sind in den Life Sciences essentiell. Ihre korrekte Anwendung merzt falsch positive und falsch negative Resultate aus und garantiert die experimentelle Reproduzierbarkeit. Dennoch verstehen und verwenden die Biologen die Grundkonzepte der Biostatistik nur in ungenügender Weise, was zu einer wachsenden Besorgnis über die Reproduzierbarkeit der Resultate führt. In dieser Studie quantifizierten die Autoren in sechs 2014 erschienenen pharmakologischen Periodika die Artikel, die die statistischen Richtlinien nicht respektierten.

Es zeigt sich, dass falsche Designs und Analysen häufig sind, darunter fehlende Erwähnungen der

Limiten des Konfidenzintervalls 95% (z.B. nur p-Werte) in der Konklusion (98% der Artikel), geringe Aussagekraft, unbekannte Grösse der Proben (73%) oder der Umgehungen der vorgegebenen Parameter (86%) und eine fehlende Korrektur für multiple Vergleiche (23%). Weitere Fehlerquellen umfassen die Verwendung der MSE (mittlere quadratische Abweichung) zum Illustrieren der Variabilität (72%), ungenaue p-Werte (86%), unbekannte Fehlerbalken (20%) und das Fehlen eines Statistikparagraphe (2%).

Zudem korrelierte das Einhalten der Richtlinien nicht mit dem Impact Factor (R^2 TOTAL = 0.02; R^2 DESIGN/ANALYSE = 0.03; R^2 PRÄSENTATION = 0.00). Die Resultate zeigen, dass experimentelle Pharmakologen die Biostatistik schlecht beherrschen, was sich in den Publikationen in einer hohen Frequenz falsch positiver und falsch negativer Resultate niederschlägt.

¹ Biotelligences LLC

Methoden

Es wurden sechs Journale untersucht: Neuropsychopharmacology, British Journal of Pharmacology, Biochemical Pharmacology, Neuropharmacology, BMC Pharmacology and Toxicology und European Journal of Pharmacology. Deren Archive wurden gescannt und die neun bis zehn neusten Artikel heruntergeladen (Stichdatum: 1. April 2014).

Einschlusskriterien:

- bereits publizierte Artikel, Briefe oder Kurzmitteilungen (Ausschluss, wenn im Druck oder Druckfahnen noch nicht korrigiert)
- mindestens eine Abbildung oder Tabelle mit einer Quantifizierung
- experimentelles Vorgehen am Tier oder an menschlichen Zellen, in vitro/ex vivo oder in silico (klinische und epidemiologische Artikel wurden ausgeschlossen)
- zusätzliche nicht analysierte Abbildungen
- Analyse zusätzlicher Methoden

Die Evaluationskriterien wurden adaptiert nach Curran-Everett and Benos D.J. 2004 J. Neurophysiol und Bailar J.C., Mosteller F. 1988 Ann Int Med. Jedes Kriterium wurde gescort mit 0 (Kriterium in allen Abbildungen und Tabellen berücksichtigt) oder 1 (Fehler in mindestens einer Abbildung oder Tabelle). Die Analysen wurden punkto Identität der Autoren und deren Verbindungen unverblindet durchgeführt.

wendung parametrischer Tests führen zu einem häufigen Vorkommen falsch positiver und falsch negativer Resultate. Zudem verhindert eine schlechte Präsentation (ungenügend beschriebene Untersuchungen, unbekannte Tests, inadäquate Fehlerbalken, nicht angegebene genaue p-Werte, unbekannte post-hoc-Tests* oder sogar fehlende statistische Paragraphen) die Verifikation der Daten und kompromittiert deren Reproduzierbarkeit.

Einfache Massnahmen zur Verbesserung der Bewusstwerdung über und des Verständnisses für die Basiskonzepte der Biostatistik bei den Forschern können sofort ergriffen werden:

- Betonung der Bedeutung der Aussagekraft
- Warnung vor multiplen Vergleichen
- Hinweis auf die starken Limitierungen durch die p-Werte
- Erklärungen zu den Unterschieden zwischen den Fehlerbalken

Die Herausgeber wissenschaftlicher Journale sollten sich aktiver an der Verbesserung der statistischen Qualität ihrer Publikationen beteiligen. ■

* Erlaubt einen Vergleich von Gruppen, ohne dass vor der Datenanalyse eine Hypothese über die Beziehung zwischen diesen Gruppen aufgestellt worden ist.

Resultate und Konklusionen

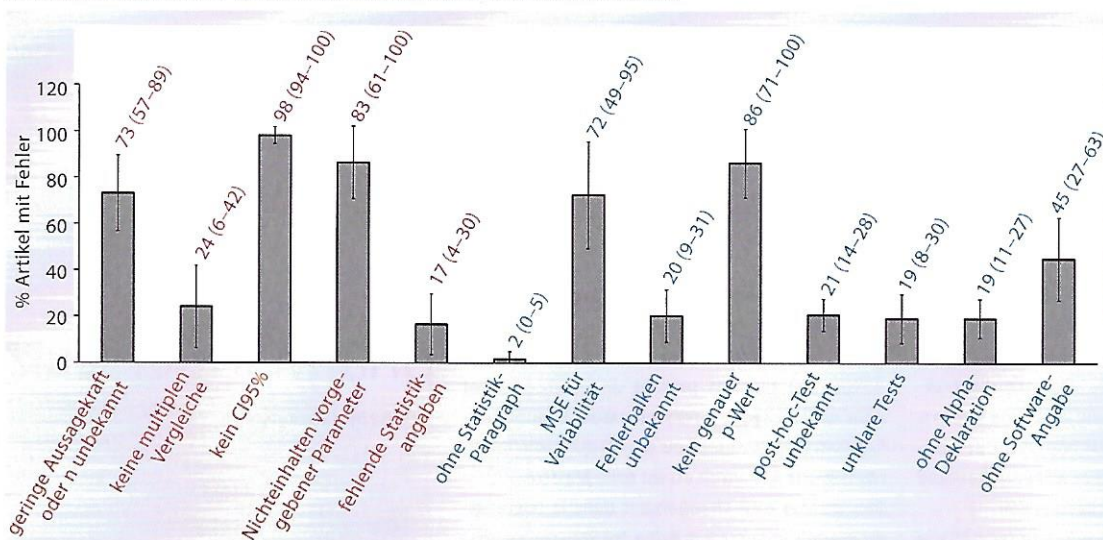
Die statistischen Richtlinien werden nicht respektiert; in der experimentellen Pharmakologie sind Fehler häufig. Eine geringe Aussagekraft, nicht korrigierte multiple Vergleiche und eine inadäquate Ver-

Korrespondenzadresse

Romain-Daniel Gosselin

E-Mail: rdg@biotelligences.com

Abbildung 1: Artikel mit Fehler in Funktion der verschiedenen Kriterien in %



Rot: Fehler im Design und in der Analyse. Blau: Fehler in der Präsentation der Daten oder der Deklaration

Darstellung als Mittelwert ± Konfidenzintervall CI95% (der reelle Mittelwert mit 95% Chancen, innerhalb dieses Intervalls zu liegen). Probengrösse: 8 in der ganzen Serie