

Points importants à retenir en matière de rédaction scientifique liée aux analyses biostatistiques

Auteur : Stéphane Morisset

Note préalable : les points ci-dessous comportent les grandes lignes à verbaliser dans une publication scientifique basées sur des analyses biostatistiques et sont à adapter en fonction du journal dans lequel vous allez écrire, il faut bien prendre note du nombre de tables, graphiques (définition d'image, couleurs autorisées ou non) et nombre de caractères auxquels vous avez droits. Les noms des paragraphes (et sous-paragraphes) peuvent évidemment changer d'une publication à une autre.

Paragraphe 'Methods'

- Study design : Rappel du contexte et description de l'expérience/analyse réalisé, le 'design' de l'étude qui peut aussi être illustré sous la forme d'un schéma (stratégie de thérapie, durée, étapes...) ainsi que principaux objectifs de l'étude. C'est dans ce paragraphe que l'on indique préalablement la nature de l'étude (rétrospective, prospective, type d'essai clinique, méta-analyse,...), mono- ou multicentrique...
 - ⇒ *Importance de la nature de l'étude sur les possibles biais que celle-ci pourra rencontrer, une étude rétrospective risque d'avoir bien plus de biais qu'un essai clinique contrôlé par exemple. Le design de l'étude a aussi son importance car elle peut déjà indiquer à quelle période des mesures vont être effectuées par exemple. Cela donne une première idée globale de la difficulté et de la complexité des analyses biostats à effectuer.*
- Data collection : comment (dossier patient ? e-CRF ? base de données en *open access* ?...) et quelles sont les données qui ont été collectées (démographiques, cliniques, paramètres biologiques). D'autres détails ont leur importance, comme les outils ou machines spécifiques à la mesure de certaines variables (marque et modèle des machines,
 - ⇒ *Quels sont les « ingrédients » de cette étude ? Que va-t-on évaluer comme facteurs ? et de manière implicite, quelles variables seront absentes ?*
- Critères d'inclusion et d'exclusion si analyses sur la personne : quelle est la population ciblée ? Quel est l'échantillon qui va être étudié ?
 - ⇒ *Spécificités de la population étudiée : générale ? ou au contraire d'individus avec une pathologie méconnue par exemple ? Autrement dit, à quelle échelle se situe-t-on, quels individus sont représentés, concernés par ces résultats.*
- Objectifs ou 'endpoints' : On explicite un peu plus en détail les différents objectifs primaires et secondaires de l'étude : survie ? indicateur de l'état du patient ? constante biologique ?
 - ⇒ *Cela donne des indices capitaux sur la manière, sous quelle forme ces données 'réponses' sont entrées dans la base (de manière binaire ? numérique ? catégorielle ? avec des dates d'événements ou sur une période fixe ?)*

- Matériels et méthodes : écriture des tests utilisés dans un premier temps dans le descriptif et la comparaison des groupes d'individus. Ensuite le processus méthodologique, la stratégie d'analyse qui a été effectuée pour répondre à l'ensemble des objectifs primaires, expliquer le mieux possible comment les résultats ont été obtenus en toute transparence, en ne manquant aucun détail (ne pas oublier qu'un biostatisticien ou un reviewer zélé peut relever des manques qui peuvent nuire à la reproduction des résultats). Attention aussi à indiquer le seuil de significativité (même si c'est 5% dans 95% des études), car de cela découle le pourcentage des intervalles de confiance, et aussi des logiciels et packages utilisés, ainsi que leurs versions. Le détail des graphiques est rarement explicité dans ce paragraphe s'ils ne servent que d'illustration, excepté s'ils ont valeur de résultats (exemple : un « volcano plot » indiquant les différences d'expressions de gènes observés, analyses en composantes principales...).

⇒ *On déroule dans ce paragraphe l'ensemble du « mode d'emploi » de la stratégie d'analyse effectué sur le plan biostat, les différentes étapes qui permet d'aboutir et de comprendre les résultats qui vont être décrits, d'analyser les outcomes en fonction des données collectées et du degré de robustesse des résultats ou estimations obtenus. C'est probablement l'aide la plus précieuse qu'une publication puisse fournir et un indicateur fort de la rigueur sur le plan statistique de la publi, et par conséquent de la rigueur de l'équipe de scientifiques d'une manière générale. En effet, si des relecteurs observent que des tests ou analyses ont été réalisés de manière peu adéquate, la publication pourra être critiquée. Il est conseillé que le ou la biostatisticien.ne de l'étude relise ou écrive cette partie.*

Paragraphe 'Results'

- Dans un premier temps, une description littérale de la population ou des groupes d'individus sur la base des analyses descriptives effectuées. Elle peut être plus ou moins exhaustives, l'important étant d'évaluer la population cible et de certaines caractéristiques qui peuvent impacter sur les résultats. Note importante : il ne faut pas sous-estimer l'importance du report de données manquantes pour certaines variables ! Evaluer un facteur de risque dont il manque 60% des données est une faiblesse qu'il faut rendre compte. Un *flowchart* permet également de voir les étapes de sélection de la population analysée par rapport à celle initialement prévue dans le cadre de l'étude.
- Ensuite les résultats : pour chaque objectif analysé, il faut reporter les différences ou mesures observées, des estimations des régressions le cas échéant, avec toutes les statistiques (ainsi que les intervalles de confiance) qui permettent la compréhension des résultats. Et bien sûr les p-values respectives, avec mention des tests ou régressions utilisés.
- Les tables : doivent être lisibles et compréhensibles, en rapport avec les objectifs analysés, suffisamment documentés (mettre juste une variable et une p-value ne suffit pas, il faut reporter les calculs et estimations). Il est essentiel re-indiquer dans la légende le type d'analyses stats effectués pour obtenir cette table, avec pourquoi pas des détails supplémentaires qui n'ont pu être intégrés dans le paragraphe « matériels et méthodes ».
- Les graphiques : le choix du graphique et ce qui est représenté doit être forcément en lien avec les résultats observés, comme si c'était la version « visuelle » des résultats. Son

esthétique et la lisibilité sont tout aussi importants que les informations présentes sur le graphique. Les détails sur le type de graphique (boxplot, histogramme, courbe de survie, etc.) et ce qui y est représenté (moyennes, probabilités, évolution... ?) doivent être décrits en légende. Dans certains cas particuliers où les graphiques font office de résultats, quelques détails permettant leur compréhension ne sont pas de trop. Astuce : cette description doit se faire comme si on décrivait le type de graphique comme à une personne malvoyante (sans l'interprétation que vous en faites).

Paragraphe 'Discussion'

- Le regard 'biostatistiques' a son importance dans la façon de nuancer et de formuler les résultats. Par exemple, on peut avoir un résultat « statistiquement significatif » avec un effet observé très relatif (comme par exemple un coefficient de corrélation faible). Il y a aussi tout l'aspect sémantique, les tournures de phrases concernant les résultats (ne sont-elles pas trop grossières ou ambiguës ?). Mais surtout, est-ce que les résultats ont été bien compris, ou est-ce que les explications autour des résultats sont bien claires et correctes (d'autant plus si sur le plan biostat, d'autres analyses complémentaires ont été effectuées, et si oui lesquelles).
 - ⇒ *Outre le regard du clinicien permettant de comprendre le résultat sur le plan biologique et/ou clinique, et son impact, le regard biostatistique apporte une lecture secondaire et d'autres éléments, une relecture avec un autre œil critique qu'il ne faut pas sous-estimer. Le but ici est d'être le plus honnête possible, quitte à faire sa propre rétrospective.*

Une publication bien rédigée permettra d'évaluer la robustesse des résultats et des conclusions menées et ultimement, de juger la reproductibilité de ces résultats ou de la capacité à pouvoir reproduire une étude similaire. Voici quelques points-clefs à retenir avec le regard biostatistique :

- **Design de l'étude et data récoltées : cela peut donner des indications sur la difficulté de « nettoyer la base » (contrôle qualité), d'étudier avec le plus de précision possible l'outcome/endpoint/objectif analysé (comment est-il codifié dans la base de données ?), d'évaluer les potentiels biais si comparaison de cohortes/groupes d'individus. La reproduction des résultats sur d'autres cohortes ou leur re-vérification passe d'abord par cette étape ;**
- **Le paragraphe stats : en plus d'indiquer les stats, régressions et la marche à suivre, c'est également un gage du sérieux des analyses faites (ne serait-ce qu'afficher les tests et la façon dont sont décrites les données, est-ce que certaines hypothèses comme la normalité sont vérifiées par exemple ?). Si le paragraphe stats indique des analyses complexes mais avec très peu de détails sur le processus ou certaines méthodes spécifiques utilisées, cela peut être un « *red flag* » pour les relecteurs car on pourra se demander si des biais méthodologiques n'ont pas été introduits et dès lors si le résultat trouvé est vraiment tangible, ou pire, truqué.**
- **Importance du descriptif des données : le profil des patients doit être limpide et transparent (indication de données manquantes). D'une étude à l'autre, des résultats peuvent différer sur un même objectif simplement car le profil d'individus n'est pas le même.**

- **Utilité du paragraphe discussion : l'autocritique permet de nuancer les propos et résultats (peu importe si l'hypothèse se vérifie ou non !), avec les limites et points forts de la méthodologie appliquée. On peut aussi proposer d'autres pistes : que faire pour aller plus loin ? générer de nouvelles hypothèses ?**
- **Le point global le plus fondamental : faire en sorte que ses propres résultats puissent être très facilement reproduits ou régénérés, afin qu'ils puissent l'être ou challengés par d'autres équipes, et continuer ainsi à faire avancer la science !**