

Le biostatisticien : un acteur clé de la recherche en santé

Dans le domaine de la santé et des sciences du vivant, les données occupent une place centrale. Chaque étude clinique ou projet de recherche génère une quantité considérable d'informations qui doivent être analysées de manière rigoureuse afin de produire des résultats robustes et pertinents.

C'est dans ce contexte qu'intervient le biostatisticien, expert de l'analyse des données et partenaire essentiel des chercheurs, cliniciens et industriels de la santé.

➤ **Les biostatistiques : un outil essentiel d'aide à la décision**

Les biostatistiques regroupent un ensemble de méthodes permettant de représenter, analyser et interpréter les données.

Elles englobent notamment :

- Les statistiques descriptives et l'inférence statistique
- Les modèles de régression et les analyses multivariées
- Les analyses de survie

Elles constituent un véritable outil d'aide à la décision permettant de transformer des données brutes en informations exploitables afin d'évaluer l'efficacité et la sécurité des traitements, et d'aider les autorités réglementaires (telles que l'EMA, la FDA ou l'ANSM)¹ à prendre des décisions éclairées.

Grâce à ces méthodes, il est possible d'évaluer l'efficacité d'un traitement, de comparer différentes approches thérapeutiques ou encore d'identifier des facteurs de risque. Les résultats obtenus contribuent ainsi à l'avancement des connaissances scientifiques et au développement de nouveaux médicaments.

➤ **Du protocole aux résultats : les multiples facettes du métier**

Le biostatisticien intervient à toutes les étapes du cycle de vie d'une étude et apporte son expertise méthodologique dès sa conception.

Il contribue notamment à :

- Définir le type d'étude (essai contrôlé randomisé, étude observationnelle, etc.)
- Élaborer le plan expérimental (groupes parallèles, cross-over, etc.)
- Déterminer la taille d'échantillon nécessaire pour atteindre les objectifs scientifiques.

Il participe également à la détection et à la réduction des biais potentiels, notamment les biais de sélection et les facteurs de confusion. Il contribue par ailleurs à la définition des stratégies de gestion des données manquantes (imputation multiple, analyses de sensibilité, etc.), afin de garantir la fiabilité et la robustesse des résultats.

Le biostatisticien élabore ensuite le Plan d'Analyse Statistique (Statistical Analysis Plan, SAP), un document qui décrit de manière détaillée l'ensemble des méthodes statistiques qui seront appliquées lors de l'analyse des données. Rédigé avant le début des analyses, il garantit la transparence et la reproductibilité et la robustesse des résultats.

Dans le cadre des essais cliniques, il définit, réalise ou supervise les analyses statistiques prévues dans le SAP conformément aux recommandations de l'ICH² E9 et aux exigences réglementaires en vigueur. Il contribue également à la production et à la validation des TFLs³ utilisés pour l'interprétation des résultats et les soumissions réglementaires.

Au-delà de l'analyse statistique, le biostatisticien apporte une expertise méthodologique essentielle à l'interprétation critique des résultats. Son rôle consiste également à communiquer les résultats de manière claire et adaptée aux chercheurs, cliniciens et décideurs afin de faciliter leur prise de décision.

➤ **La boîte à outils du biostatisticien**

Pour mener à bien ses missions, le biostatisticien s'appuie sur plusieurs outils et standards reconnus dans le secteur de la santé et des sciences du vivant.

Dans l'industrie pharmaceutique et la recherche clinique, certains jouent un rôle central pour garantir la qualité, la traçabilité et la conformité réglementaire des données et des analyses destinées aux autorités de santé. Parmi eux :

- **CDISC⁴**, qui constitue un ensemble de standards internationaux destinés à l'harmonisation, à l'organisation et à l'échange des données cliniques ainsi qu'aux soumissions aux autorités réglementaires notamment à travers :
 - **CDASH⁵** pour la conception standardisée des eCRF⁶
 - **SDTM⁷** pour la structuration des données selon des domaines standardisés
 - **ADaM⁸** pour la création de jeux de données optimisés pour l'analyse statistique
- **SAS^{®9}**, qui demeure une référence grâce à sa robustesse, sa traçabilité et ses procédures standardisées répondant aux exigences réglementaires internationales

Parallèlement, les solutions open source occupent une place croissante dans l'environnement de travail du biostatisticien.

Le langage de programmation **R**, associé à l'environnement de développement **RStudio**, est particulièrement apprécié pour sa flexibilité, sa richesse fonctionnelle et son vaste écosystème de packages spécialisés.

Grâce à **Shiny**, un framework de l'environnement R, les biostatisticiens peuvent également développer des applications web interactives destinées à la visualisation, à l'exploration et au partage des résultats.

Leur caractère open source en fait des outils largement utilisés dans la recherche académique, les biotechnologies et, de plus en plus, dans l'industrie pharmaceutique et la recherche clinique.

➤ **Un métier au cœur de l'innovation médicale**

Ainsi, grâce à son expertise méthodologique et à sa maîtrise des outils statistiques, le biostatisticien joue un rôle essentiel dans la recherche en santé. Son travail contribue au développement de nouveaux traitements, à éclairer les décisions médicales et à garantir la qualité des résultats auprès des autorités de santé.

En combinant expertise scientifique, analyse de données et maîtrise des exigences réglementaires, le biostatisticien contribue pleinement au développement de l'innovation médicale. Cette dynamique est d'autant plus marquée avec l'essor de l'intelligence artificielle, du machine learning et des données de vie réelle (RWE¹⁰), qui permettent d'exploiter des volumes croissants de données et de développer des modèles prédictifs toujours plus sophistiqués.

Contrairement à une idée reçue, l'intelligence artificielle ne remplace pas le biostatisticien. Elle renforce au contraire son rôle. Les modèles prédictifs doivent être validés, leurs performances correctement évaluées, leurs biais identifiés et leur utilisation justifiée dans un cadre réglementaire exigeant

¹ EMA : European Medicines Agency

FDA : Food and Drug Administration

ANSM : Agence nationale de sécurité du médicament et des produits de santé

² ICH : International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use

³ TFLs : Tables, Figures and Listings

⁴ CDISC : Clinical Data Interchange Standards Consortium

⁵ CDASH : Clinical Data Acquisition Standards Harmonization

⁶ eCRF : electronic Case Report Form

⁷ SDTM : Study Data Tabulation Model

⁸ ADAM : Analysis Data Model

⁹ SAS : Statistical Analysis System

¹⁰ RWE : Real-World Data