

Offre de stage de M2 en Biostatistique / Programmation R

Titre : Développement d'une alternative open-source sous R aux logiciels propriétaires de calcul de taille d'échantillon en oncologie

Année universitaire : 2025-2026

Encadrants : Dan CHALTIEL, Anne LOURDESSAMY

Référence : StageM2-Oncostat2025/SampleSize

Contexte du projet

Le calcul de taille d'échantillon est une étape clé dans la conception d'un essai clinique. Il permet de s'assurer que l'étude disposera de la puissance statistique nécessaire pour répondre à la question de recherche, tout en limitant le nombre de patients exposés au traitement expérimental. À Gustave Roussy, ce principe est essentiel car les essais évaluent des traitements anticancéreux à la toxicité élevée.

Dans la pratique, de nombreux services de biostatistique s'appuient sur des logiciels propriétaires tels que East ou nQuery. Ces outils sont validés et riches en fonctionnalités mais présentent plusieurs limites, leur coût est élevé, le code source est fermé et leur flexibilité est limitée pour des applications spécifiques.

Parallèlement, l'écosystème open-source R propose une grande variété de packages permettant le calcul de taille d'échantillon selon différents modèles et plans expérimentaux. Cependant, cet ensemble reste fragmenté : documentation inégale, interfaces hétérogènes, et absence d'évaluation systématique vis-à-vis des solutions propriétaires.

Dans un contexte où l'open-science et la reproductibilité sont devenues des valeurs scientifiques essentielles, il est crucial de renforcer l'usage d'outils libres, transparents et validés.

Ce stage s'inscrit dans cette dynamique : il vise à évaluer, comparer et consolider les solutions existantes sous R afin de proposer une alternative fiable, accessible et adaptée aux besoins de l'oncologie.

Mots-clés : R, biostatistiques, calcul de taille d'échantillon, oncologie, open science, GitHub, méthodologie, essais cliniques

Objectifs du stage

1. **Cartographier l'existant** : identifier et décrire les principaux packages R dédiés au calcul de taille d'échantillon, en particulier pour les méthodes utilisées en oncologie.
2. **Comparer les méthodes** : mettre en place une batterie de cas tests représentatifs d'essais cliniques en oncologie, comparer les résultats obtenus avec R à ceux produits par des logiciels commerciaux de référence (East, nQuery), identifier les divergences méthodologiques et les limites éventuelles.
3. **Harmoniser et intégrer les méthodes dans un outil R unifié** : rédiger un guide de référence (manuel utilisateur) présentant les fonctions R disponibles, leurs paramètres, hypothèses et exemples d'utilisation et développer une interface unifiée facilitant l'accès à ces fonctions. Selon l'avancement du stage, possibilité d'aller plus loin vers

la création d'un package R, d'une application interactive Shiny, ou d'une réflexion sur la data-visualisation des résultats.

4. **Valoriser et diffuser les résultats** : publier le code, les tests, et la documentation sur GitHub et rédiger un rapport de synthèse (possible participation à une communication scientifique).

Profil recherché

- Etudiant(e) de Master 2 ou dernière année d'école d'ingénieur dans l'un des domaines suivants : Biostatistique, Statistique appliquée, Informatique, Data Science, Santé publique, Bioinformatique.
- Appétence pour la programmation, idéalement en R
- Notions de statistiques
- Intérêt pour la recherche appliquée, la méthodologie clinique et l'open science
- Une première expérience avec GitHub serait un plus

Ce que ce stage vous apportera

- Compréhension des bases méthodologiques du calcul de taille d'échantillon.
- Pratique avancée de la programmation en R (code clair, reproductible et bien documenté)
- Capacité à évaluer et documenter de manière critique différents outils statistiques.
- Utilisation de GitHub pour le versioning, le partage, et la collaboration (open science).
- Possibilité de valorisation scientifique (publication, communication)

Environnement de travail

Le stage sera réalisé au sein du service de Biostatistiques et d'Epidémiologie de Gustave Roussy (Inserm, CESP U1018, Equipe Oncostat Gustave Roussy - B2M 39 rue Camille Desmoulins - 94805 Villejuif)

La durée prévue est de 6 mois.

Contact

Les candidat(e)s intéressé(e)s sont invité(e)s à envoyer leur CV et une lettre de motivation à : dan.chaltiel@gustaveroussy.fr et anne.lourdessamy@gustaveroussy.fr