

Consignes pour l'envoi de vos données aux biostatisticiens

1. Fichier de données (de type Excel)

| ID | GROUPE_TRAITEMENT | TYPE_EXAM | SEXE | AGE_ANNEES | BMI | ATCD_HTA | ATCD_CANCER | ATCD_CHIR | SDRA_48 |
|----|-------------------|-----------|------|------------|-----|----------|-------------|-----------|---------|
| 1 | 0 | 1 | 0 | 66.3 | | 1 | 0 | 0 | 0 |
| 2 | 0 | 3 | 1 | 62.0 | 32 | 0 | 1 | 0 | 0 |
| 3 | 1 | 3 | 1 | 56.2 | 25 | 1 | 0 | 0 | 0 |
| 4 | 0 | 2 | 1 | 55.8 | 32 | 0 | 0 | 1 | 0 |
| 5 | 1 | | 0 | 55.5 | 29 | 1 | 1 | 1 | 1 |
| 6 | 1 | 2 | 1 | 44.9 | 26 | 0 | 0 | 1 | 0 |
| 7 | 0 | 2 | 0 | 66.1 | 20 | 0 | 0 | 1 | 1 |
| 8 | 1 | 1 | 0 | 85.7 | 43 | 1 | 1 | 0 | |

A. Noms des variables

Respectez ces 5 règles pour que nous puissions exploiter votre fichier :

- **Pas d'espace**,
- **Pas de caractère spécial** (/ , - , accents , % , < , ...) — seul l'underscore est accepté (_),
- **Maximum 30 caractères** — abrégé si nécessaire,
- **Pas de doublons** — chaque nom de variable doit être unique,
- **Commencer par une lettre** — jamais par un chiffre,

| <input checked="" type="checkbox"/> Non accepté | | <input checked="" type="checkbox"/> À utiliser |
|---|---|--|
| GROUPE DE TRAITEMENT | → | GROUPE_TRAITEMENT |
| AUTRES_MALADIES/INFECTIONS | → | AUTRES_MALADIES_INFECTIONS |
| AGE (années) | → | AGE_ANNEES ou AGE |
| PRECEDENTE_HOSPITALISATION_EN_REA | → | PREC_HOSPI_REA |
| 48_SDRA | → | SDRA_48 |

B. Codage des variables

Toutes les données doivent être chiffrées.

Variable catégorielle : Réponse unique (une seule valeur possible)

Un chiffre par modalité :

| <input checked="" type="checkbox"/> Non accepté | | <input checked="" type="checkbox"/> À utiliser |
|--|---|---|
| SEXE = HOMME SEXE = FEMME | → | SEXE = 0 SEXE = 1 |
| TYPE_EXAM = ECHO TYPE_EXAM = SCANNER TYPE_EXAM = IRM | → | TYPE_EXAM = 1 TYPE_EXAM = 2 TYPE_EXAM = 3 |

Variable catégorielle : Réponses multiples (plusieurs valeurs possibles)

Créez une variable binaire (0/1) par modalité :

| <input checked="" type="checkbox"/> Non accepté | | <input checked="" type="checkbox"/> À utiliser |
|---|---|--|
| ANTECEDENTS = HTA, CANCER, CHIR | → | ATCD_HTA = 1 ATCD_CANCER = 1 ATCD_CHIR = 1 |
| ANTECEDENTS = 2+3 | → | ATCD_HTA = 0 ATCD_CANCER = 1 ATCD_CHIR = 1 |
| Si aucun ANTECEDENTS | → | ATCD_HTA = 0 ATCD_CANCER = 0 ATCD_CHIR = 0 |

Variabes quantitatives

Utilisez le point (.) comme décimal.

| <input checked="" type="checkbox"/> Non accepté | | <input checked="" type="checkbox"/> À utiliser |
|---|---|---|
| DLCO_PCT = 88% | → | DLCO_PCT = 88 |
| AGE_GESTATIONNEL = 38+5 | → | AGE_GESTA = 38.7 (rapportez les jours en semaine 5/7=0.7) |
| CRP = <1 (limite de détection) | → | CRP = 1 OU CRP = 0 |

Dates

- Format JJ/MM/AAAA ou MM/AAAA (gardez le même format pour toute la colonne). **Ne laissez que les dates utiles au calcul de délais**

C. Données manquantes

Règle unique : une donnée manquante ou non applicable = une cellule vide

- Supprimez toute valeur texte : NA, na, ND, NC, Non applicable, Non déterminé, ...

2. Lexique

Fournissez un lexique des variables (sur Word ou feuille Excel)

- Une ligne par variable,
- Indiquez le nom de la variable, sa signification, le sens du codage

- ID : Identifiant du sujet
- AGE : Âge du patient (années)
- CORTICO : Dose de corticoïdes, mg/L
- TYPE_EXAM : 1 = ECHOGRAPHIE, 2 = SCANNER, 3 = IRM

3. Plan d'Analyse Statistique (PAS)

Indiquez

- Vos coordonnées (téléphone, mail et spécialité)
- Brève description du sujet
- Les questions abordées durant le rendez-vous avec les biostatisticiens, en y précisant les noms des variables à utiliser

Le descriptif de toutes les variables présentes dans la base, sur l'ensemble des sujets, sera systématiquement réalisé.

Exemple de formulation des questions dans le PAS :

1. Décrire toutes les variables selon le traitement reçu par le sujet (GROUPE_TRAITEMENT)
2. Comparer les variables suivantes selon le traitement reçu par le sujet (GROUPE_TRAITEMENT) :
 - a. SEXE
 - b. SDRA_48
3. Rechercher les facteurs de risque de positivité de la TEP (FDG=1), en ajustant sur l'hyperglycémie (HYPERGLY), parmi les variables suivantes :
 - a. AGE_ANNEES
 - b. SEXE
 - c. TYPE_EXAM
 - d. TAILLE_TUMEUR
 - e. Présence d'au moins un antécédent (Variable à dériver : ANTECEDENTS via la formule suivante : Si ATCD_HTA = 1 OU ATCD_CANCER = 1 OU ATCD_CHIR = 1 ALORS ANTECEDENTS = 1, SI ATCD_HTA = 0 ET ATCD_CANCER = 0 ET ATCD_CHIR = 0 ALORS ANTECEDENTS = 0)
4. Comparer la survie globale (DECES, DELAI_DECES_DN) selon le traitement (GROUPE_TRAITEMENT). DELAI_DECES_DN sera calculé entre date_chirurgie et date_deces ou date_chirurgie et date_dern_nouvelles.

⚠ Les questions non abordées lors du rendez-vous de consultation mais ajoutées au plan d'analyse seront traitées sous réserve de leur compréhension, faisabilité ou pertinence au regard des questions abordées lors du rendez-vous.

En résumé — checklist avant envoi

- Fichier anonymisé (sans IPP ni IEP)
- Noms de variables : sans espace, sans accents, maximum 30 caractères, commençant par une lettre
- Modalités codées en chiffres
- Réponses multiples : une colonne par modalité
- Données manquantes : cellules vides (pas de NA/ND/...)
- Lexique dans une feuille à part de vos données
- Plan d'analyse statistique qui reprend les noms des variables

4. Envoi des fichiers aux biostatisticiens

1. Envoyez vos fichiers via la plateforme CHU transfert (<https://chutransfert.chru-lille.fr>) et renseignez comme destinataire, l'adresse suivante : biostatistique@chu-lille.fr.
2. Envoyez par mail, à la même adresse, le mot de passe de chiffrement que vous aurez renseigné lors du dépôt sur la plateforme, nécessaire à la récupération de vos fichiers.

Pour plus d'informations sur les missions de notre équipe, vous pouvez consulter notre site internet : <http://seed.chu-lille.fr>.

Pour toute demande de rendez-vous pour un nouveau projet, merci de contacter notre secrétariat via secretariat-databiostat@chu-lille.fr.