# Plan d'analyses statistiques

Biomonitoring visant à déterminer les niveaux d'imprégnation de référence de la population wallonne aux substances chimiques présentes dans l'environnement – phase 3 (BMH-WAL 3)

**Institut Scientifique de Service Public (ISSeP)**

Mme Aline Jacques

Mme Ingrid Ruthy

Version : 2.0
Date : 26.04.2024

Biostatisticiennes : Pr. AF. Donneau, N. Dardenne, A. Diep

Centre Hospitalo-Universitaire de Biostatistique et Méthodes de Recherche (B-Stat)

**Table des matières**

**1 Introduction** ..... **4**
**2 Préparation des données** ..... **4**
**2.1 Imputation des données censurées à gauche** ..... **4**
**2.1.1 Imputation simple** ..... 4
**2.1.2 Imputation Multiple par Équation de Chaîne (MICE)** ..... 5
**2.1.3 Transformation en variable binaire** ..... 5
**2.2 Normalisation/Standardisation de la concentration urinaire/lipidique** ..... **5**
**2.3 Transformation de la valeur des biomarqueurs** ..... **6**
**2.4 Valeurs aberrantes** ..... **6**
**2.4.1 Méthode d'identification des valeurs extrêmes** ..... 6
**2.4.2 Traitement des valeurs aberrantes** ..... 6
**2.5 Traitement des données manquantes** ..... **6**
**2.6 Paramètres de somme** ..... **7**
**3 Statistiques descriptives** ..... **8**
**3.1 Caractéristiques des participants** ..... **8**
**3.2 Biomarqueurs** ..... **8**
**4 Modélisation** ..... **10**
**4.1 Régression linéaire multiple pour les biomarqueurs continus** ..... **10**
**4.1.1 Spécification du modèle** ..... 10
**4.1.2 Analyses simples et multiples** ..... 10
**4.1.3 Analyses diagnostiques du modèle** ..... 11
**4.1.4 Présentation de résultats** ..... 12
**4.2 Régression logistique pour les biomarqueurs binaires** ..... **12**
**4.2.1 Spécification du modèle** ..... 12
**4.2.2 Analyses simples et multiples** ..... 12
**4.2.3 Analyses diagnostiques du modèle** ..... 13
**4.2.4 Présentation de résultats** ..... 13

## Table des abréviations

| | |
|---|---|
| PAS | Plan d’analyses statistiques |
| LOQ | Limite de quantification |
| PARC | Partnership for the Assessment of Risks from Chemicals |
| ln | Logarithme naturel / népérien |
| MICE | Multiple imputation by Chained Equations |
| CRT | Créatinine |
| SG | Gravité spécifique |
| Q-Q plot | Graphique Quantile-quantile |
| Delta-beta | Dfbeta |
| IC | Intervalle de confiance |
| SD | Ecart-type |
| M | Moyenne arithmétique |
| GM | Moyenne géométrique |
| IQR | Intervalle interquartile |
| AIC | Critère d’information d'Akaike |
| AUC | Aire sous la courbe |
| HAP | Hydrocarbure aromatique polycyclique |
| PCB | Polychlorobiphényle |

# 1 Introduction

L'objectif du Programme de biosurveillance BMH-Wal est d'identifier les déterminants des expositions de 80 biomarqueurs (la liste des biomarqueurs est fournie dans le fichier BHW_resume_phases_et_substance.xlsx).

Un large éventail de facteurs personnels, sociaux, économiques et environnementaux qui peuvent influencer le niveau d'exposition aux biomarqueurs chez les individus est considéré (la liste complète de ces variables indépendantes se trouve dans le questionnaire envoyé le 24/08/2023). Plus spécifiquement, il s'agit de l'âge/de catégories d'âge, du sexe, du statut tabagique, des zones de résidence, du régime alimentaire, de la consommation alimentaire récente, etc.

L'impact de ces variables indépendantes sur les biomarqueurs, disponibles sous forme de variables continues (éventuellement transformées en logarithme naturel) et catégorielles, sera investigué à l'aide de régressions linéaires multiples et de régressions logistiques, respectivement. Dans cette version du plan d'analyse statistique (PAS), la stratégie d'application de régressions linéaires et logistiques multiples afin de répondre à cet objectif est présentée.

Le présent Plan d'Analyses Statistiques (PAS) est organisé en trois sections : préparation des données, résumé des données (statistiques descriptives) et modélisation.

# 2 Préparation des données

## 2.1 Imputation des données censurées à gauche

Les valeurs des biomarqueurs inférieures à la limite de quantification (LOQ) seront considérées comme des données censurées à gauche et seront imputées.

Sur la base de ce qui est rapporté, par exemple dans BMH-Wal rapport, les données des biomarqueurs présentent des distributions asymétriques et des valeurs positives. En fonction de la distribution des résidus des modèles de régressions (dans ces cas, certaines distributions sont plus applicables que d'autres comme Gamma, Weibull et log-normal) et du pourcentage de valeurs indétectables, une imputation simple ou multiple sera appliquée aux données sur base des recommandations proposées par PARC.

### 2.1.1 Imputation simple

L'imputation simple (log-normal) sera appliquée lorsque 1) au moins 30 % des valeurs d'un biomarqueur sont supérieures à la LOQ et 2) au moins 10 valeurs différentes sont observées au-dessus de la LOQ pour un biomarqueur spécifique par collecte de données. Ces conditions permettent de garantir une certaine variabilité des données.

Les données censurées à gauche (valeurs <LOQ) seront imputées à l'aide d'une technique d'imputation simple aléatoire, dépendant de la distribution de la variable. Par exemple, dans le cas où les données

suivraient une distribution log-normale, une régression log-normale tronquée utilisant des valeurs supérieures à LOQ sera ajustée. D'autres distributions seront investiguées si nécessaire afin de déterminer, au moyen d'un test statistique (par exemple, le test de Kolmogorov-Smirnov), la distribution la plus appropriée. Ensuite, des valeurs aléatoires comprises entre 0 et LOQ seront générées afin de garantir la variabilité des valeurs imputées.

Afin de garantir que les valeurs imputées n'influencent pas la variabilité des données, les données imputées et les données censurées à gauche seront comparées en termes de moyenne, médiane et distance de Kolmogorov-Smirnov pour garantir que les valeurs imputées n'influencent pas la variabilité des données.

### 2.1.2 Imputation Multiple par Équation de Chaîne (MICE)

Dans le cas où pour un biomarqueur la proportion de valeurs <LOQ est élevée (≥ 15%), la technique d'imputation multiple par équations en chaine (MICE) sera utilisée. Pour cela, les valeurs censurées à gauche du biomarqueur concerné seront imputées au moyen d'un ensemble de prédicteurs. Conformément à l'objectif du présent projet, cet ensemble des prédicteurs sera susceptible d'être les déterminants inclus dans le modèle de prédiction.

Le nombre d'imputations sera fixé à 20, ce qui est le minimum suggéré par Van Buuren (2008). Selon cette technique d'imputation, 20 ensembles de données imputées seront générés. Le modèle de prédiction permettant de répondre à l'objectif de l'étude sera appliqué sur chaque ensemble de données et les résultats obtenus seront ensuite regroupés afin d'obtenir les coefficients de régression et l'intervalle de confiance à 95 % correspondant.

Un 'stripplot' permettant de visualiser la relation entre le biomarqueur d'intérêt et un prédicteur donné (celui avec lequel le biomarqueur a une forte corrélation) sera utilisé pour examiner si les valeurs imputées s'inscrivent bien dans la plage de données attendues et ne sont pas des données aberrantes modifiant la relation initiale entre ces deux paramètres.

### 2.1.3 Transformation en variable binaire

Pour les biomarqueurs présentant plus de 70 % des observations inférieures à la LOQ, il est recommandé de diviser les observations en <LOQ versus ≥ LOQ.

### 2.2 Normalisation/Standardisation de la concentration urinaire/lipidique

Dans PARC, section 3.3.1.1 (p.45), on attire l'attention sur l'ajustement des différences inter-individuelles dans la dilution de l'échantillon urinaire par la créatinine (CRT), la gravité spécifique (SG) et l'osmolalité, etc. Dans le cadre de BMH-Wal, seule la créatinine a été dosée.

La normalisation à la créatinine ou aux lipides ne sera pas effectuée.

### 2.3 Transformation de la valeur des biomarqueurs

Lorsqu'une analyse de régression linéaire sera appliquée, l'examen de la normalité des résidus sera effectué. Cette étude de la distribution résiduelle sera effectuée à l'aide du test de normalité de Shapiro-Wilk et investiguée graphiquement à l'aide d'un histogramme, d'un graphique de densité ainsi que d'un graphique Q-Q (quantile-quantile). Lorsque la condition d'application liée à la normalité ne pourra être confirmée, une transformation des données du biomarqueurs pourra être appliquée. Dans ce cas, la transformation logarithmique naturelle/népérien sera investiguée afin d'obtenir des résultats interprétables et permettant la comparaison avec d'autres études.

### 2.4 Valeurs aberrantes

#### 2.4.1 Méthode d'identification des valeurs extrêmes

Afin d'identifier les valeurs extrêmes potentiellement aberrantes, des scores z exprimant l'écart par rapport à la valeur moyenne en termes d'écart-type, seront calculés. Les valeurs avec un score z supérieur à 3 ou inférieur à –3 seront considérées comme étant des valeurs aberrantes.

Si une transformation doit être appliquée sur les données d'un biomarqueur, la détection des données potentiellement aberrantes sera effectuée sur les données transformées.

#### 2.4.2 Traitement des valeurs aberrantes

Une fois détectées, il s'agira de vérifier si ces valeurs extrêmes et potentiellement aberrantes ont un impact sur les résultats. Pour cela, l'impact de la présence de ces valeurs sur les résultats des modèles de régression sera examiné au moyen d'une analyse de sensibilité.

Cette étape sera réalisée au moyen d'un graphique diagnostic des résidus par rapport à l'effet de levier. Par la suite, les delta-bêtas (DFbeta) – qui rapportent dans quelle mesure le coefficient du modèle de régression change lorsque l'observation est omise - et les DFBetas standardisés, seront utilisés pour définir les valeurs aberrantes influentes. Les DFBetas standardisés avec des valeurs absolues supérieures à 1 indiquent des valeurs qui influencent considérablement les paramètres du modèle (bien que Stevens (2002) suggère d'examiner les cas avec des valeurs absolues supérieures à 2).

Si les valeurs potentiellement aberrantes sont définies comme étant influentes, une transformation sera nécessaire (par exemple en utilisant la méthode Tukey) et l'analyse statistique des données transformées sera refaite. Si, après transformation, le modèle est toujours affecté par la présence de valeurs aberrantes influentes, ces valeurs aberrantes influentes seront supprimées du modèle.

### 2.5 Traitement des données manquantes

Selon la nature et la proportion des données manquantes (<15% ou >=15%), une technique d'imputation sera appliquée.

Conformément à PARC, les données manquantes de manière complètement aléatoire (MCAR – Missing completely at random), par exemple les échantillons de sang perdus ou endommagés, seront imputées.

Les données manquantes de manière aléatoire MAR (Missing at random - manquantes en raison d'une autre variable) et les données manquantes de manière non aléatoire MNAR (Missing not at random - manquantes en raison de la nature de la variable elle-même) seront imputées à l'aide de la méthode MICE tel que décrit à la section 2.1.2.

**2.6 Paramètres de somme**

Certains biomarqueurs sont des métabolites issus d'un seul et même composé parent. Pour ces biomarqueurs, il peut également être utile d'examiner la somme de ces biomarqueurs.

Pour ce faire, de nouvelles variables « somme » seront créées. Ces nouvelles variables ne seront calculées que si tous les biomarqueurs qui la composent sont dosés et si au moins un des biomarqueurs présente au moins 60 % des valeurs détectées. Comme les biomarqueurs imputés ne sont obtenus que si au moins 70 % des valeurs sont détectées, les variables de somme seront calculées à l'aide d'une technique d'imputation simple pour chacun des biomarqueurs formant cette variable somme, les niveaux inférieurs à la LOQ seront imputés.

La somme molaire des métabolites est obtenue en divisant la concentration de chaque métabolite par son poids moléculaire, puis en les additionnant.

En plus de la somme des composés, une analyse en composantes principales au sein de chaque groupe de substances pourra être appliquée et le facteur latent sera utilisé pour analyser l'association avec les variables prédictives dans une analyse de régression multiple.

Les sommes de métabolites suivantes seront envisagées. Les métabolites qui constituent ces sommes seront également analysés individuellement sauf les hydroxyfluorènes, les hydroxyphénanthrènes et les isomères DCCA.

- 2-HO-fluorène + 3-HO-fluorène + 9-HO-fluorène
- 1-HO-phénanthrène + 2-HO-phénanthrène + 3-HO-phénanthrène + 4-HO-phénanthrène
- 1-naphtol + 2-naphtol
- PCB-138 + PCB-158 + PCB-180
- c-DCCA + t-DCCA

# 3 Statistiques descriptives

## 3.1 Caractéristiques des participants

Les caractéristiques sociodémographiques des individus seront résumées numériquement. Les variables qualitatives seront présentées au moyen de proportion avec IC respectif à 95 %, et les variables quantitatives sous forme de moyenne (SD) ou de médiane (intervalle interquartile - IQR) selon la normalité de leur distribution.

Concernant les variables catégorielles, les modalités de réponse devraient présenter un minimum de 5 individus et pourraient être combinés le cas échéant afin de faciliter l'analyse.

## 3.2 Biomarqueurs

Pour chaque biomarqueur (après préparation des données telle que décrite ci-dessus : imputation, ….), le nombre d'observations, le nombre et le pourcentage de valeurs manquantes, ainsi que le nombre et le pourcentage de valeurs inférieures à la LOQ et les centiles ($5^e$, $10^e$, $25^e$, $50^e$, $75^e$, $90^e$ et $95^e$) seront présentés. De plus, la moyenne arithmétique (M) et la moyenne géométrique (GM) avec leur intervalle de confiance (IC) à 95 % respectif, seront présentées. Les GM (IC à 95 %) seront obtenues en prenant l'antilog de la moyenne estimée et ses limites de confiance transformées en logarithme naturel.

Ces statistiques descriptives seront présentées pour les concentrations non corrigées et corrigées (pour la gravité spécifique CRT), dans les cas où cela s'applique.

Les biomarqueurs sont présentés de manière globale, selon leur origine (urine et sang), et selon les caractéristiques des individus (par exemple, groupes d'âge, sexe, niveau d'étude, revenus du ménage, nombre de personnes dans le ménage, situation professionnelle, et catégorie socio-professionnelle).

Table x. Distributions des concentrations des biomarqueurs de manière globale et selon les différentes caractéristiques des individus

| | **%>LOQ** | **n** | **P5** | **P10** | **P25** | **P50** | **P75** | **P90** | **P95** | **M (IC95%)** | **GM (IC95%)** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Biomarqueur x | | | | | | | | | | | |
| Total (n) | | | | | | | | | | | |
| **Groupes d'âge** | | | | | | | | | | | |
| Catégorie 1<br>Catégorie 2<br>… | | | | | | | | | | | |
| **Sexe** | | | | | | | | | | | |

| | %>LOQ | n | P5 | P10 | P25 | P50 | P75 | P90 | P95 | M (IC95%) | GM (IC95%) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Homme<br>Femme<br>.. | | | | | | | | | | | |
| **Niveau d'étude**<br>Primaire ou sans diplôme<br>Secondaire inférieur<br>Secondaire supérieur<br>Enseignement supérieur (court)<br>Enseignement supérieur (long) | | | | | | | | | | | |
| **Revenus du ménage (net)**<br>< 1000 €<br>1000 -1999 €<br>2000 – 2999 €<br>3000 – 3999 €<br>4000 – 4999 €<br>> 5000 € | | | | | | | | | | | |
| **Nombre de personnes dans le ménage** | | | | | | | | | | | |
| **Situation professionnelle**<br>Je travaille<br>Je suis sans emploi, je perçois « le chômage »<br>Je suis étudiant<br>Je suis en contrat d'apprentissage | | | | | | | | | | | |

| | %>LOQ | n | P5 | P10 | P25 | P50 | P75 | P90 | P95 | M (IC95%) | GM (IC95%) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Je suis au foyer, je ne perçois aucun revenu | | | | | | | | | | | |
| Je suis en congé de maladie/invalidité de longue durée | | | | | | | | | | | |
| Autre | | | | | | | | | | | |
| **Socio-professionnelle** | | | | | | | | | | | |
| Cadre-dirigeant | | | | | | | | | | | |
| Employé | | | | | | | | | | | |
| Ouvrier | | | | | | | | | | | |
| Profession libérale | | | | | | | | | | | |
| Indépendant | | | | | | | | | | | |
| Autre | | | | | | | | | | | |

Pour la création d'indicateurs (regroupement de variables) utile pour la description des biomarqueurs, le présent PAS sera modifié sur base des décisions du consortium.

# 4 Modélisation

## 4.1 Régression linéaire multiple pour les biomarqueurs continus

### 4.1.1 Spécification du modèle

Lors de l'analyse des biomarqueurs d'exposition (variables continues) en tant que variables dépendantes, les modèles suivants seront appliqués avec l'âge, IMC, genre comme facteurs confondants. D'autres facteurs confondants pourraient être ajoutés en fonction du biomarqueur modélisé (par exemple, le tabac).

Biomarqueur ~ CRT + autres variables indépendantes

### 4.1.2 Analyses simples et multiples

L'association entre chaque variable indépendante (données collectées selon les questionnaires, par exemple âge, sexe, caractéristiques sociodémographiques, alimentation, degré d'urbanisation, etc.) et les variables dépendantes (biomarqueurs) seront examinées au moyen d'un modèle de régression simple. Les variables indépendantes présentant une p-valeur <0,20 sont incluses dans un modèle de régression linéaire multiple généré pas à pas.

La sélection pas à pas des variables se fera de manière descendante (élimination) : le modèle prendra en compte toutes les variables indépendantes (modèle complet) et éliminera de manière itérative les variables les moins contributives (valeur de p fixée à <0,05) ; le processus de sélection s'arrêtera lorsque toutes les variables retenues sont significatives. Le R-carré – qui reflète le pourcentage de variabilité du niveau du biomarqueur qui pourrait être expliqué par son association avec les variables indépendantes retenue dans le modèle final – sera présenté pour les modèles univariés et le modèle final.

Les coefficients de régression linéaire obtenus seront présentés comme l'inverse des estimations transformées en log et seront exprimés comme le facteur de changement des concentrations pour l'augmentation unitaire par rapport à la variable indépendante. Ils seront de plus exprimés en pourcentage de variation pour chaque variable incluse dans le modèle final.

### 4.1.3 Analyses diagnostiques du modèle

Les conditions d'application sous-jacente à cette analyse de régression linéaire seront vérifiées pour garantir la validité de l'analyse comme suit.

- **Linéarité** : La relation entre les déterminants d'expositions et les biomarqueurs doit suivre des hypothèses de linéarité spécifiques pour la moyenne des résultats pour la régression linéaire. Cela se vérifiera en utilisant des tracés de diagnostic représentant les résidus par rapport aux valeurs prédites
- **Indépendance** : S'il existe des variables supposées corrélées aux résidus, la relation sera vérifiée graphiquement pour déterminer la présence de cluster ou si les résidus sont distribués de manière aléatoire.
- **Normalité des résidus** : l'histogramme, le graphique Q-Q et le test de Shapiro-Wilk seront utilisés pour examiner la normalité de la distribution des résidus. Si les résidus ne sont pas distribués normalement, les erreurs types, et donc les intervalles de confiance ainsi que les p-valeurs, pourraient être biaisés. En cas de violations de cette condition d'application, une transformation devra être appliquée aux données. Si la normalité des résidus n'était pas vérifiée, d'autres modèles pourraient être envisagés comme des modèles de régression quantile.
- **Variance résiduelle constante (homoscédasticité)** : les résidus seront comparés aux valeurs prédites (ajustées) par le modèle et testés par les tests Breusch-Pagan / Cook-Weisberg.
- **Multicolinéarité :** les variables indépendantes hautement corrélées seront examinées au moyen du facteur d'inflation de la variance (VIF). Une valeur VIF >5 sera considérée comme problématique et des méthodes pour réduire cette multicolinéarité pourront être appliquées en fonction du nombre de variables corrélées telles qu'en centrant la variable ou la réduction de dimension – par exemple au moyen d'une analyse en composantes principales.

### 4.1.4 Présentation de résultats

La concentration de chaque biomarqueur sera présentée selon les données individuelles (âge, sexe, caractéristiques sociodémographiques), alimentation, degré d'urbanisation, etc. à l'aide d'une régression linéaire (cf. section 4.1.2).

Le tableau suivant reprendra les résultats des régressions linéaires univariées pour les variables indépendantes significatives, et donc incluses dans le modèle multivarié, ainsi que les résultats du modèle multivarié. L'ensemble des résultats univariés seront par contre présentés en annexe.

Table x. Résultats de l'analyse de l'association entre la concentration du biomarqueur x et les différentes variables indépendantes

| | | | Biomarqueur x | | | |
|---|---|---|---|---|---|---|
| **Variables indépendantes** | | | **Simple** | | **Multiple (n = x)** | |
| | **n** | **Coef (IC95%)** | **p** | **Coef (IC95%)** | **P** |
| Variable qualitative | x | | | | |
| Catégorie 1 | Ref. | x | | | | |
| Catégorie 2 | | x | x.xx (IC95%) | x.xx | x.xx (IC95%) | x.xx |
| ⋮ | ⋮ | | ⋮ | ⋮ | ⋮ | ⋮ |
| Variable quantitative | | x | x.xx (IC95%) | x.xx | x.xx (IC95%) | x.xx |
| $R^2$ = 0.xxx | | | | | | |

## 4.2 Régression logistique pour les biomarqueurs binaires

### 4.2.1 Spécification du modèle

Lors de l'analyse des biomarqueurs d'exposition binaires (pour rappel, pour les biomarqueurs présentant plus de 70 % des observations inférieures à la LOQ, il est recommandé de diviser les observations en <LOQ versus ≥ LOQ) en tant que variables dépendantes, les modèles suivants seront appliqués, avec l'âge, l'IMC et le genre comme facteurs confondants (à nouveau, d'autres facteurs pourront être ajoutés en fonction du biomarqueur étudié).

$$\text{Ln}\,[Y1-Y] \sim \text{CRT} + \text{variables indépendantes}$$

Y : la probabilité de détection

### 4.2.2 Analyses simples et multiples

La sélection des prédicteurs significatifs pour les biomarqueurs binaires au niveau simple se fera tel que décrit à la section 4.1.2

Tous les prédicteurs significatifs en analyse simple seront inclus dans une régression logistique binaire multiple (modèle complet). En parallèle, un modèle le plus simple possible sera construit basé sur la sélection backward (descendante). Dans cette technique, le modèle prendra en compte toutes les variables indépendantes (modèle complet) et ajoutera ou supprimera de manière itérative les variables

basées sur le critère d'information d'Akaike (AIC). Ce critère permettra de comparer les modèles, le meilleur modèle étant celui avec la plus petite valeur de AIC.

Le modèle complet et le modèle obtenu après application de la procédure pas à pas seront comparés en utilisant la précision de la prédiction, l'AUC (aire sous la courbe, plus la valeur est élevée, meilleur est le modèle), la spécificité et la sensibilité. Sur cette base, la conclusion d'utiliser le modèle complet ou le modèle le plus simple possible sera justifiée.

Les résultats seront présentés au moyen des rapports de cotes et de leur intervalle de confiance à 95% correspondant (OR – IC à 95 %).

### 4.2.3 Analyses diagnostiques du modèle

Les conditions d'application sous-jacente à l'application de régressions logistiques binaires seront vérifiées en termes de linéarité, de valeurs aberrantes influentes et de multicolinéarité à l'aide des techniques telles que celles décrites aux sections 2.5.2 and 4.1.3.

### 4.2.4 Présentation de résultats

Le tableau suivant reprendra les résultats des régressions logistiques binaires pour les variables indépendantes significatives, et donc incluses dans le modèle multivarié, ainsi que les résultats du modèle multivarié. L'ensemble des résultats univariés seront par contre présentés en annexe.

Table x. Résultat de l'association entre le pourcentage de données détectées ou non pour le biomarqueur x et les prédicteurs

| **Variable indépendantes** | **Biomarqueur x** | | | **Simple** | | **Multiple (n = x)** | |
|---|---|---|---|---|---|---|---|
| | **n** | **Détectée (n)** | **Non détectée (n)** | **OR (IC95%)** | **p** | **OR (IC95%)** | **p** |
| Variable quantitative Moyenne ± SD | | | | | | | |
| | x | x.xx ±x.xx | x.xx ±x.xx | x.xx (x.xx-x.xx) | x.xx | x.xx (x.xx-x.xx) | x.xx |
| Variable quantitative Médiane (Q1-Q3) | | | | | | | |
| | x | x.xx (x.xx-x.xx) | x.xx (x.xx-x.xx) | x.xx (x.xx-x.xx) | x.xx | x.xx (x.xx-x.xx) | x.xx |
| Variable qualitative n (%) | | | | | | | |
| Catégorie 1 | x | x.xx (x.xx) | x.xx (x.xx) | Ref. | | Ref. | |
| Catégorie 2 | x | x.xx (x.xx) | x.xx (x.xx) | x.xx (x.xx-x.xx) | x.xx | x.xx (x.xx-x.xx) | x.xx |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |

## 4.3 Analyse de sensibilité

Les modèles de régression seront appliqués sur les données complètes ainsi que sur les données imputées afin d'évaluer la stabilité des estimations. De plus, dans le cadre de l'évaluation des modèles, ceux-ci seront examinés en tenant compte des données avec et sans valeurs aberrantes influentes. Sur cette base, des sources potentielles de biais et d'incertitude pourront être identifiées. Si les résultats d'une étude restent cohérents dans une gamme de scénarios plausibles, la confiance dans les conclusions de l'étude en sera renforcée. Au contraire, un changement substantiel entre les différents résultats indiquera la nécessité de faire preuve de prudence dans l'interprétation de ces résultats et permettra de mettre en évidence des domaines nécessitant des recherches ou des raffinements plus approfondis.

**Référence**

Analyse des déterminants d'exposition via le questionnaire des participants – Programme de biosurveillance BMH-Wal 1 & 2 (2019-2021) – No rapport : RP1-RAP-23-04308

Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of statistical software*, *45*, 1-67.