Vous souhaitez étudier le rôle de l’exposition aux perturbateurs endocriniens *(pour cela vous mesurez la concentration de multiples dérivés de ces polluants chimiques dans les urines, au total 54 biomarqueurs, chacun distribué selon un loi Normale)* sur la fonction respiratoire *(une mesure du souffle que l’on nommera VEMS pour volume expiratoire maximale en 1 seconde, une variable continue qui est distribuée selon une loi Normale)* dans la population française.

Vous avez recruté 5050 individus dans différentes régions et avez réussi à mesurer l’ensemble des biomarqueurs et la fonction respiratoire sur un échantillon de 5000 individus. Vous disposez de l’ensemble des données sur les facteurs confondants pour 4005 individus. On soupçonne pour 3 biomarqueurs un possible effet plus prononcé chez les fumeurs actifs compte tenu de résultats d’études antérieures.

**1- Pourriez-vous préciser les différentes étapes du protocole d’analyse en précisant à chaque étape la méthode statistique envisagée, les paramètres statistiques estimés et l’intérêt de cette étape ?**

|  |  |  |  |
| --- | --- | --- | --- |
| N° | Nom de l’étape | Méthode statistique et paramètres statistiques reportés dans l’article | Intérêt de conduire cette étape |
| 1 |  |  |  |

Type de réponse attendu :

|  |  |  |  |
| --- | --- | --- | --- |
| N° | Nom de l’étape | Méthode statistique et paramètres statistiques reportés dans l’article | Intérêt de conduire cette étape |
| 0 | Nettoyage des données | Recherche d’outliers | Corriger ou supprimer les valeurs aberrantes |
| 1 | Analyse descriptive de la population sur les expos, le VEMS et les possibles facteurs confondants | Calcul de moyenne ± sd ou % | Caractériser la population d’étude (sera utile pour comparer notre population à celle étudiée dans des études précédentes et donc discuter les résultats) |
| 2 | Mesure de la corrélation entre les 54 biomarqueurs | Matrice de corrélation - coefficient de Pearson et p associés | Comprendre le lien entre les expositions |
| 3 | Imputation des données manquantes | Imputation multiple | Permet de ne pas réduire la taille de l’analyse entre les analyses univariées et mulitvariées. |
| 4 | Identification des facteurs de confusions | Régressions linaires entre chaque biomarqueurset chaque facteur de confusion potentiel, puis entre chaque facteur de confusion potentiel et le VEMS. | On conservera comme facteurs de confusion les facteurs associés avec le VEMS et ≥1 biomarqueur à p<0.20 |
| 5 | Test de l’interaction avec le tabac sur le 3 biomarqueurs (candidats pour interaction avec tabac) sur le VEMS | 3 régressions linaires multiples avec terme d’interaction tabac\*biomarqueur  VEMS = a + béta\*biomarqueur + c\*tabac + d\*tabac\*biomarqueur + fi\*Confi+ e  🡺d [IC95%] et p associé | Test le rôle du tabac comme effet modificateur dans la relation biomarqueur-VEMS. Si le test d’interaction est non significatif on ne gardera pas le terme d’interaction pour la suite de l’analyse pour ce biomarqueur, sinon, le biomarqueur devra être étudié séparément chez les fumeurs et non fumeurs. |
| 6 | Mesure de l’association univariée ente chaque biomarqueur et le VEMS | 51-54 (selon résultats de l’étape 5) régressions linéaires simples  (VEMS = a + béta\*biomarqueur + e)  🡺 Béta [IC95%] et p associé | Caractériser le sens et le niveau de l’association (Béta), son IC, et le degré de certitude que l’on a pour chaque association non-ajustée observée |
| 7 | Mesure de l’association ajustée sur les facteurs de confusion ente chaque biomarqueur et le VEMS | 51-54 régressions linéaires multiples  (VEMS = a + béta\*biomarqueur + c1\*Confound1+…+e)  🡺 Calcul du Béta [IC95%] et p associé | Caractériser le sens et le niveau de l’association (Béta), et le degré de certitude que l’on a pour chacune des associations ajustées. |
| 8 | Correction pour les tests multiples | Méthode de Bonferoni ou du FDR ou autres méthodes permettant de prendre en compte le degré de corrélation entre chaque expo selon les résultats de l’étape 2- A faire en considérant les 51-54 tests | C’est avec ces résultats que l’on va conclure à un effet significatif pour les biomarqueurs dont la valeur du p-corrigé est <0.05 |

**2- Les associations entre les analyses non ajustées et ajustées sur les facteurs de confusion sont différentes. Certaines associations qui étaient significatives deviennent non-significatives et une association non-significative devient significative. Comment interprétez-vous cela ?**

🡺 La réponse doit inclure une explication sur **la taille d’échantillons** (si l’étudiant ne propose pas de méthodes d’imputation dans le 1) – on perd 1/5° de la population ce qui a un impact direct sur les IC (plus grand) et p et donc peut expliquer une partie des associations qui passent non significative) et **sur les effets des facteurs de confusion sur la force de l’association**. Un facteur de confusion peut jouer dans les 2 sens !