

Réanalyse des données de l'étude :

*Surveillance de l'imprégnation à l'arsenic
chez la population du quartier Notre-Dame
(2005-2006)*

Auteur

Sébastien Rivest, agent de recherche en science des données
Observatoire national sur les incidences des émissions de contaminants sur la santé et l'environnement

Sous la direction de

Daniel Proulx, directeur
Observatoire national sur les incidences des émissions de contaminants sur la santé et l'environnement

Réviseur(e)s

Maryse Bouchard, professeure titulaire de santé environnementale
Centre Armand-Frappier Santé Biotechnologie, Institut national de la recherche scientifique

Yohann Moanahere Chiu, professeur à la Faculté de médecine et des sciences de la santé
Département de médecine de famille, Université de Sherbrooke

Leonara Marro, cheffe de section des biostatistiques
Santé Canada

Nolwenn Noisel, professeure agrégée à l'École de santé publique
Département de santé environnementale et santé au travail, Université de Montréal

Annie St-Amand, cheffe de section de la biosurveillance nationale
Santé Canada

Observatoire national sur les incidences des émissions de contaminants sur la santé et l'environnement
Université du Québec en Abitibi-Témiscamingue

445 boulevard de l'Université
Rouyn-Noranda (Québec) J9X 5E4
onicse@uqat.ca

Table des matières

Résumé	1
Introduction	2
Méthodologie	3
Résultats et discussion	4
Références	6
Annexe 1	8

Résumé

Une étude de biosurveillance menée auprès de la population de Rouyn-Noranda en 2005-2006 visait à comparer l'exposition à l'arsenic des résidents du quartier Notre-Dame à celle d'un groupe témoin situé dans le quartier Évain. Toutefois, l'étude présente plusieurs limites méthodologiques et statistiques qui en réduisent la portée interprétative. Les avancées méthodologiques en analyse statistique au cours des dernières décennies, de même que l'intégration des connaissances désormais disponibles, justifient une réanalyse des données issues de cette étude de biosurveillance.

L'étude initiale ne faisait état d'aucune différence d'imprégnation entre les groupes exposé et témoin en considérant la population dans son ensemble. Toutefois, dans une démarche de biosurveillance, la pratique est d'accorder une attention particulière aux différents groupes d'âge, notamment les jeunes, qui constituent une sous-population à risque concernant les contaminants environnementaux. La réanalyse des données de 2005-2006 met en évidence une interaction entre l'âge et l'exposition, indiquant que l'effet de l'exposition à l'arsenic sur l'imprégnation est significativement plus marqué chez les enfants de moins de dix ans que dans les autres groupes d'âge. De plus, l'examen ciblé de cette sous-population montre une imprégnation à l'arsenic significativement plus élevée chez les jeunes enfants exposés que chez les témoins.

Cette interprétation doit néanmoins être envisagée avec prudence. L'étude de biosurveillance de 2005-2006 comporte plusieurs limites méthodologiques qui restreignent la robustesse des conclusions pouvant en être tirées. Les résultats obtenus soulignent néanmoins la nécessité de mener des investigations complémentaires plus approfondies auprès des jeunes enfants, chez qui un risque accru de surexposition à l'arsenic apparaît plausible.

Introduction

En 2005-2006, une étude de biosurveillance a été réalisée par la Direction de santé publique (DSPu) de l'Abitibi-Témiscamingue auprès de la population de Rouyn-Noranda afin de comparer l'exposition à l'arsenic des résidents du quartier Notre-Dame (QND), le secteur le plus influencé par les émissions diffuses et fugitives de la Fonderie Horne, à celle d'un quartier moins exposé, Évain (D. Gagné 2007). L'objectif consistait à déterminer si les habitants du QND présentaient une imprégnation à l'arsenic plus élevée que ceux d'Évain.

Lors de cette étude, l'arsenic urinaire avait été retenu comme biomarqueur d'exposition. Au total, 206 participantes et participants ont été recrutés dans le QND, tandis que 113 participantes et participants l'ont été dans le quartier Évain (voir Gagné, 2007, pour plus de détails méthodologiques). Les participantes et participants étaient répartis en six classes d'âge : 1 à 4 ans, 5 à 9 ans, 10 à 14 ans, 15 à 24 ans, 25 à 60 ans, et 60 ans et plus. Plusieurs membres d'une même famille pouvaient prendre part à l'étude. Cinq campagnes d'échantillonnage ont été réalisées auprès des participantes et participants à différentes saisons. Pour chaque campagne, les prélèvements étaient effectués à une date fixe pour l'ensemble des participantes et participants, à partir d'un échantillon d'urine du matin. Cette étude ne faisait état d'aucune différence d'imprégnation entre le QND et le quartier Évain.

L'étude de biosurveillance de 2005-2006 comporte cependant plusieurs limites méthodologiques et erreurs statistiques réduisant substantiellement sa portée interprétative. La transformation des données, réalisée pour satisfaire l'hypothèse de normalité des erreurs, apparaît inadéquate (voir Annexe 1 pour plus de détails sur les enjeux statistiques). De plus, aucune vérification des hypothèses du modèle n'est rapportée, ce qui empêche d'évaluer la validité de l'ajustement statistique. Par ailleurs, la non-indépendance des mesures au sein d'une même famille et d'une même personne (la plupart ont participé à plusieurs campagnes d'échantillonnage), ainsi que la possibilité d'une autocorrélation spatiale n'ont pas été considérées. À ces limites s'ajoutent des erreurs, notamment l'inversion récurrente entre variables explicatives et variables réponses dans les modèles, compromettant la robustesse des conclusions avancées.

L'étude de biosurveillance de 2005-2006 a évalué l'ensemble des participantes et participants dans chacun des groupes étudiés, mais a accordé une attention limitée aux sous-populations constituant ces groupes pouvant présenter une susceptibilité accrue. Or, une démarche de biosurveillance devrait s'intéresser aux groupes plus vulnérables, tels que les enfants, les femmes enceintes ou les personnes présentant des conditions de santé particulières, dont les profils biologiques, les comportements d'exposition ou l'utilisation des milieux de vie peuvent entraîner des niveaux d'imprégnation distincts (M. Gagné et al. 2014). L'absence d'analyse stratifiée selon l'âge réduit ainsi la capacité d'identifier des disparités d'exposition et peut conduire à une sous-estimation des risques pour les segments les plus sensibles de la population.

Malgré ces limites statistiques et méthodologiques, les résultats de l'étude de biosurveillance de 2005-2006 continuent d'être mobilisés dans les discussions portant sur les émissions de contaminants de la Fonderie Horne à Rouyn-Noranda. Compte tenu des faiblesses statistiques de cette étude et des progrès substantiels réalisés en analyse statistique et interprétation des données de biosurveillance au cours des dernières décennies, une réanalyse des données originales apparaît justifiée. Le présent rapport propose une réévaluation de ces données en recourant à des méthodes statistiques plus avancées visant à réduire les biais statistiques, à mieux tenir compte des structures de dépendance et à renforcer la validité des inférences produites.

Méthodologie

Les données issues de l'étude de biosurveillance de 2005-2006 menée par la DSPu de l'Abitibi-Témiscamingue ont été transmises afin d'en faire la réanalyse (D. Gagné 2007). Elles comprennent les concentrations urinaires d'arsenic ainsi que les informations complètes des participantes et des participants.

L'étude de 2005-2006 répartissait ces personnes en six classes d'âge : 1-4 ans, 5-9 ans, 10-14 ans, 15-24 ans, 25-60 ans, et 60 et plus. Bien que cette structure d'âge soit biologiquement pertinente, le faible effectif de certaines catégories en réduisait la robustesse statistique. En effet, la classe des 1-4 ans ne comptait que 23 individus, dont seulement sept dans le groupe témoin (le sous-groupe avec le moins d'individus) et 16 dans le groupe exposé. À noter que cette faible taille d'effectif avait contraint l'étude originale à combiner les classes d'âge 1-4 ans et 5-9 ans dans certaines analyses. La présente étude propose une réanalyse basée sur quatre groupes d'âge : moins de 10 ans, 10-18 ans, 19-59 ans, 60 ans et plus. Cette organisation correspond approximativement à une différenciation entre jeunes enfants, adolescentes et adolescents, adultes et personnes âgées. Elle repose principalement sur la combinaison des groupes d'âge 1-4 ans et 5-9 ans de l'étude originale. Un examen visuel des distributions entre classes d'âge a montré que ces deux classes étaient similaires. De plus, la définition d'adulte de l'étude de biosurveillance de 2019 (Bilodeau 2019) a été appliquée pour établir une distinction claire entre les adolescentes et adolescents et les adultes. Cette classification est également proche de celle proposée dans l'étude du Bureau d'étude sur les substances toxiques (Savoie et Weber, 1975) fondée sur l'exposition aux contaminants tel que l'arsenic à Rouyn-Noranda et d'études de biosurveillance récentes sur l'arsenic (Campolina et al. 2024).

Pour évaluer l'effet de l'exposition à l'arsenic en fonction de la classe d'âge sur l'imprégnation à l'arsenic, un modèle bayésien multiniveau linéaire généralisé a été spécifié en utilisant la fonction « brm » de la librairie « brms » du logiciel R (Bürkner 2017; R core team 2025). L'utilisation d'un modèle bayésien a permis de spécifier conjointement la loi de probabilité et la structure de dépendance des données tout en modélisant explicitement les valeurs inférieures à la limite de détection sans leur attribuer une valeur artificielle, ce qui préserve mieux l'information et limite les biais associés aux méthodes de substitution. Un modèle initial a été construit en utilisant la concentration urinaire d'arsenic comme variable réponse et le groupe (témoin ou exposé), la classe d'âge et leur interaction comme variables explicatives. Cette approche permet de vérifier s'il y a un effet de l'exposition et de la classe d'âge et si l'effet de l'exposition diffère entre les différentes classes d'âge. Afin de refléter la structure hiérarchique du jeu de données et la non-indépendance des mesures répétées, la famille, l'individu et la campagne d'échantillonnage ont été inclus comme effets aléatoires. Compte tenu de l'asymétrie marquée vers la gauche des concentrations urinaires d'arsenic, une distribution log-normale a été utilisée. Cette distribution est appropriée lorsque la variable dépendante est continue, strictement positive et asymétrique (Bolker 2008). Compte tenu de la variabilité substantielle observée dans les concentrations urinaires d'arsenic entre les différentes campagnes d'échantillonnage, l'écart-type résiduel a été spécifié comme une fonction de cette variable ($\sigma \sim \text{campagne}$). Une portion des mesures d'arsenic urinaires était inférieure à la limite de détection de 0,7 µg/L (~ 29 % des mesures). Pour intégrer l'incertitude inhérente à ces valeurs, une censure à gauche a été appliquée aux observations situées sous le seuil de détection (*censored regression models*). Cette approche permet une plus grande robustesse du modèle, tout en évitant les biais introduits par les pratiques de substitution (remplacement par la moitié du seuil; la méthode utilisée dans l'étude de 2005-2006, par zéro ou par le seuil lui-même) (Helsel 2006).

Quatre chaînes de Markov ont été lancées, chacune comportant 4000 itérations dont 1000 ont été retirées pour la phase de « burn-in ». Des distributions a priori non informatives (uniformes) ont été choisies pour l'ensemble des paramètres afin de permettre une inférence objective et non biaisée. Un intervalle de

décimation de 10 a été fixé afin de réduire l'autocorrélation entre les échantillons issus des chaînes de Markov Monte-Carlo (MCMC). La convergence des chaînes a été contrôlée visuellement à l'aide des tracés des chaînes de Markov. La validation du modèle a également été effectuée au moyen de la statistique de Gelman–Rubin (Gelman et Rubin 1992), par méthode de validation croisée (*leave-one-out cross-validation*) et par vérification à posteriori de l'ajustement du modèle et des résidus du modèle.

La présence d'autocorrélation spatiale a été vérifiée par approche des k-plus-proches voisins ($k = 10$), définis à partir des coordonnées de chaque adresse. Un terme conditionnel autorégressif de type BYM2 a été intégré au modèle afin de représenter la dépendance spatiale entre adresses. La présence d'autocorrélation spatiale a été évaluée par l'indice de Moran, ainsi qu'en comparant la performance prédictive, par validation croisée *leave-one-out*, à celle du modèle initial dépourvu de composante spatiale. L'ajout du terme spatial n'ayant pas amélioré la performance, et l'indice de Moran n'étant pas significativement différent de zéro, le modèle initial a été retenu. L'influence de covariables susceptibles d'influencer la concentration urinaire d'arsenic comme le statut tabagique, la consommation de fruits de mer, le niveau d'éducation, la présence d'un puits contaminé et le travail dans le quartier Notre-Dame a été évaluée en les introduisant individuellement dans le modèle initial, par analyse de sensibilité ainsi qu'en comparant la performance prédictive par validation *leave-one-out*. L'effet du niveau d'éducation avait été vérifié dans l'étude de 2005-2006 bien que cette variable présente une forte corrélation intrinsèque avec l'âge (par exemple, il est impossible pour un jeune enfant d'avoir un diplôme universitaire). Dans cette réanalyse, le niveau d'éducation des participantes et des participants de 18 ans et moins a été substitué par le niveau d'éducation moyen des adultes (plus de 18 ans) demeurant à la même adresse. Aucune de ces covariables n'ayant affecté considérablement les effets des variables du modèle initial ou montré d'amélioration (du point de vue de la validation *leave-one-out*), le modèle initial a été conservé.

Résultats et discussion

Comme ce qui avait été rapporté dans l'étude de biosurveillance de 2005-2006, considérée dans son ensemble, la population de Rouyn-Noranda ne présente pas de différence significative entre les concentrations urinaires d'arsenic des groupes exposés et témoins. Pour cette comparaison globale (mesurée par effet marginal du groupe puisque le modèle contient une interaction entre le groupe et la classe d'âge), l'intervalle de crédibilité bayésien (ICB) chevauche zéro (ICB 95 % [-0.15, 0.40]). Dans une analyse bayésienne, une absence de chevauchement indique une relation statistiquement significative. Toutefois, la réanalyse des données de 2005-2006 met en évidence une interaction entre l'âge et l'exposition, indiquant que l'effet de l'exposition à l'arsenic sur l'imprégnation est significativement plus marqué chez les enfants de moins de 10 ans par rapport aux autres groupes d'âge (ICB 95 % [-1.40, -0.27] pour la comparaison aux 10 à 18 ans, [-1.14, -0.16] pour la comparaison aux 19 à 59 ans, [-1.77, -0.13] pour la comparaison aux 60 ans et plus; voir Figure 1). De plus, l'examen ciblé de cette sous-population montre une différence statistiquement significative entre l'imprégnation à l'arsenic des jeunes enfants exposés comparativement aux témoins (ICB 95 % [0.02, 0.99]). Comparativement, aucun effet mesurable de l'exposition n'a été détecté dans les autres classes d'âge. La différence en imprégnation entre le groupe exposé et témoin était relativement constante entre les campagnes d'échantillonnage, suggérant que les risques d'exposition à l'arsenic sont présents tout au long de l'année.

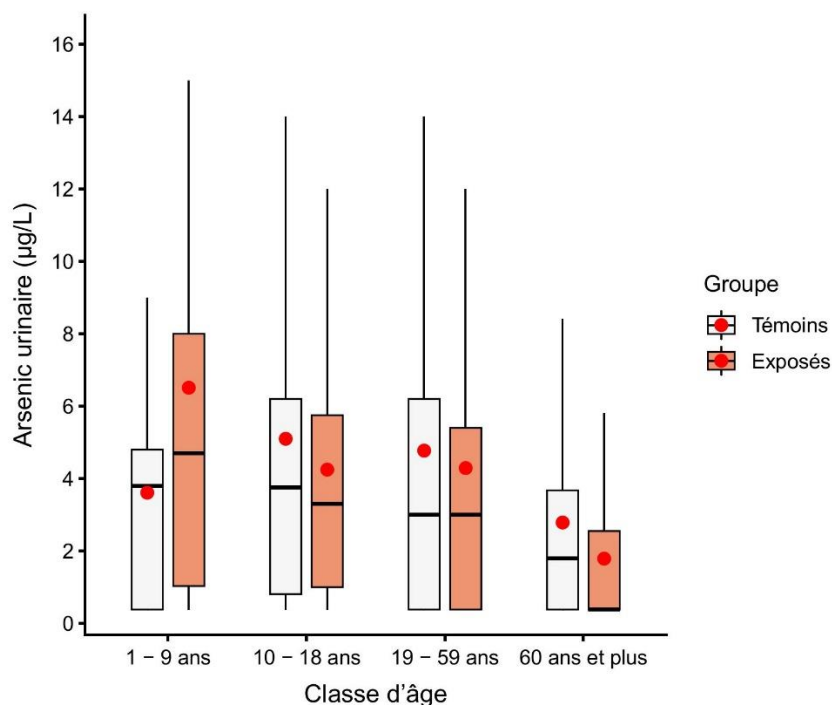


Figure 1. Concentrations urinaires d'arsenic chez le groupe exposé et le groupe témoin en fonction de la classe d'âge. Les médianes et les quartiles sont représentés par les graphiques en boîtes à moustaches, alors que les moyennes arithmétiques sont représentées par les points rouges. Les moustaches (lignes) s'étendent jusqu'aux plus petites et plus grandes observations situées dans les bornes définies par 1,5 fois l'intervalle interquartile. Les observations au-delà de ces bornes, considérées comme des valeurs extrêmes, ne sont pas affichées pour faciliter la visualisation des tendances.

Une association entre l'âge et l'exposition avait également été observée dans l'étude initiale (voir pages 35 et 36 de Gagné 2007), bien que la méthode utilisée pour en vérifier la robustesse présentait certaines limites statistiques (voir Annexe 1). En effet, globalement, la réanalyse confirme les principaux constats de l'étude originale : absence d'un effet global attribuable à l'exposition, présence d'une relation entre l'âge et l'effet de l'exposition et faible capacité explicative de la majorité des facteurs confondants potentiels. Cependant, le lien entre l'âge et l'effet de l'exposition n'avait pas été intégré aux conclusions de l'étude de biosurveillance menée en 2005-2006, et aucune analyse approfondie des variations de l'effet de l'exposition entre classes d'âge n'avait été réalisée. Dans ces conditions, affirmer une absence d'effet de l'exposition dû à un effet global non-significatif de l'exposition, comme cela avait été fait dans l'étude initiale, comporte un risque de dissimuler des effets propres à certaines sous-populations.

En se penchant sur ces sous-populations, la réanalyse de l'étude de biosurveillance de 2005-2006 permet une interprétation plus nuancée des données, mettant en lumière certaines différences dans les patrons d'imprégnation entre classes d'âge. Il est cependant crucial de tenir compte de certaines limitations méthodologiques inhérentes à l'étude dans l'interprétation de ces résultats. En particulier, la taille d'échantillon était limitée dans certains groupes d'âge, en particulier les jeunes enfants et les personnes âgées. L'interprétation des résultats devrait être faite avec prudence en particulier dans ces sous-groupes.

De plus, la réanalyse repose nécessairement sur les hypothèses méthodologiques de l'étude originale, notamment que l'arsenic urinaire dit « non-alimentaire » constitue un indicateur spécifique de l'exposition à l'arsenic inorganique d'origine environnementale. À la lumière des connaissances actuelles en biosurveillance (par exemple voir l'*Enquête canadienne sur les mesures de la santé*), cette hypothèse devrait être interprétée avec prudence. L'évaluation récente de la toxicité de l'acide diméthylarsinique (DMA), fraction importante de l'arsenic urinaire mesuré, pourrait, par ailleurs, avoir une incidence importante sur l'interprétation des résultats (OMS et FAO, 2026). Ainsi, bien que la réanalyse améliore la rigueur statistique de l'étude de 2005-2006, elle ne permet pas de surmonter les limitations associées au biomarqueur utilisé dans l'étude originale. Il est important de noter que des approches analytiques plus récentes, fondées sur la spéciation complète de l'arsenic et la quantification spécifique de ses formes inorganiques (As(III) et As(V)), combinées à des limites de détection plus faibles, pourraient améliorer la capacité à détecter ce type d'exposition.

Plutôt que de déterminer de façon décisive l'effet de l'exposition à l'arsenic sur l'imprégnation des résidentes et des résidents de Rouyn-Noranda, les résultats de cette réanalyse soulignent la nécessité de mener des investigations complémentaires plus approfondies auprès des jeunes enfants, chez qui un risque accru de surexposition à l'arsenic apparaît plausible. Des études de biosurveillance fondées sur un biomarqueur distinct ont eu lieu après celle de 2005-2006. Elles ont mis en évidence une surexposition des jeunes enfants du quartier Notre-Dame (Bilodeau 2019, 2020) et l'étude de 2005-2006 contribue ainsi à renforcer cette conclusion par la convergence de résultats issus de biomarqueurs différents.

Références

- Bilodeau, Frédéric. 2019. Rapport de l'étude de biosurveillance menée à l'automne 2018 sur l'imprégnation au plomb, au cadmium et à l'arsenic des jeunes enfants du quartier Notre-Dame à Rouyn-Noranda. Direction de santé publique de l'Abitibi-Témiscamingue, Unité de santé environnementale. https://www.cisss-at.gouv.qc.ca/wp-content/uploads/2019/11/Rapport_final_biosurveillance_2018.pdf.
- Bilodeau, Frédéric. 2020. Rapport de l'étude de biosurveillance menée à l'automne 2019 sur l'imprégnation à l'arsenic de la population du quartier Notre-Dame de Rouyn-Noranda. Avec Stéphane Bessette, Daniel Proulx, Phélix Bussière, Lyse Landry, et Omobola Sobanjo. Direction de santé publique de l'Abitibi-Témiscamingue, Unité de santé environnementale. https://www.cisss-at.gouv.qc.ca/partage/BIOSURVEILLANCE/Rapport_biosurveillance_2019_F.pdf.
- Bolker, Benjamin M. 2008. Ecological models and data in R. Princeton University Press. <https://www.degruyterbrill.com/document/doi/10.1515/9781400840908/html>.
- Bürkner, Paul-Christian. 2017. Brms: an R package for Bayesian multilevel models using Stan. *Journal of Statistical Software* 80 (août): 1-28. <https://doi.org/10.18637/jss.v080.i01>.
- Campolina, Délio, Virginia S. T. Ciminelli, Jack C. Ng, et al. 2024. Speciated and total urinary arsenic levels in Belo Horizonte, the largest Brazilian city within the mineral-rich region "iron quadrangle". *Exposure and Health* 16 (1): 101-18. <https://doi.org/10.1007/s12403-023-00540-5>.

- Gagné, Daniel. 2007. Surveillance de l'imprégnation à l'arsenic chez la population du quartier Notre-Dame (décembre 2005 à octobre 2006) : rapport final. Direction de santé publique de l'Abitibi-Témiscamingue. <https://numerique.banq.qc.ca/patrimoine/details/52327/46678>.
- Gagné, Michelle, Mathieu Valcke, Pierre Ayotte, et al. 2014. Réflexions sur l'élaboration d'une stratégie de biosurveillance en appui aux actions de santé publique au Québec: rapport synthèse. Direction de la santé environnementale et de la toxicologie, Institut national de santé publique Québec. https://www.inspq.qc.ca/sites/default/files/publications/1882_elaboration_strategie_biosurveillance.pdf.
- Gelman, Andrew, et Donald B. Rubin. 1992. Inference from iterative simulation using multiple sequences. *Statistical Science* 7 (4): 457-72. <https://doi.org/10.1214/ss/1177011136>.
- Helsel, Dennis R. 2006. Fabricating data: how substituting values for nondetects can ruin results, and what can be done about it. *Chemosphere, Environmental Chemistry*, vol. 65 (11): 2434-39. <https://doi.org/10.1016/j.chemosphere.2006.04.051>.
- Organisation mondiale de la santé et Organisation des Nations unies pour l'alimentation et l'agriculture. 2026. Evaluation of certain contaminants in food: one-hundred-and-first report of the Joint FAO/WHO Expert Committee on Food Additives. <https://iris.who.int/server/api/core/bitstreams/fa7f924f-f6b6-4994-9811-4a8d2a837b13/content>.
- R core team. 2025. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria Édition. R Foundation for Statistical Computing, Vienna, Austria Édition. <https://www.R-project.org/>.
- Statistique Canada. 2025. Enquête canadienne sur les mesures de la santé (ECMS). <https://www.statcan.gc.ca/fr/enquete/menages/5071>.
- Weber, Jean-Philippe. 1979. Étude de la distribution de certains toxiques dans la population de Rouyn-Noranda. Bureau d'étude sur les substances toxiques.

Annexe 1

Dans l'ensemble des analyses, aucune vérification explicite des hypothèses des modèles statistiques (homoscédasticité, normalité des résidus, absence de multicolinéarité élevée et indépendance des erreurs) n'est rapportée, de sorte qu'il n'est pas possible d'évaluer la pertinence du choix de ces modèles. Dans l'étude de biosurveillance de 2005-2006, les concentrations urinaires d'arsenic inférieures à la limite de détection de 0,7 µg/L ont été remplacées par la moitié de cette valeur (page 18, paragraphe 3). Une transformation logarithmique a ensuite été appliquée afin de satisfaire l'hypothèse de normalité des tests statistiques utilisés (page 19, paragraphe 3). Or, la proportion élevée de valeurs inférieures à la limite de détection (environ 29 %) engendre une inflation de la distribution log-transformée à proximité de zéro. Le rapport initial le souligne d'ailleurs : « On remarque à la figure 7 que la courbe aurait probablement eu une allure très semblable à une courbe normale si on avait pu quantifier de façon précise les valeurs sous le seuil de détection (qui sont ici représentées par l'histogramme le plus élevé pour la valeur -0,5). La proportion importante de ces valeurs a une influence sur la distribution qui s'en trouve artificiellement déportée vers la gauche. Le même phénomène se constate aussi chez les non-exposés. ». Dans ces conditions, la transformation logarithmique ne paraît pas, *a priori*, suffisante pour garantir la normalité requise par les modèles utilisés, normalité qui devrait être examinée sur les résidus des modèles, ce qui n'est pas documenté dans le rapport. Une telle distribution de valeurs censurées nécessite en principe une modélisation explicite afin de respecter les hypothèses du modèle. De plus, les pratiques de substitution (remplacement par la moitié du seuil, comme dans l'étude de 2005-2006, par zéro ou par le seuil lui-même) sont susceptibles d'introduire des biais dans les analyses (Helsel 2006).

En plus de la supposition de normalité, celle de non-indépendance des mesures ne paraît pas respectée. Le modèle ne comporte pas de variables aléatoires permettant de prendre en compte la structure de dépendance induite par des mesures répétées sur un même individu ou au sein d'une même famille. De plus, la possibilité d'une autocorrélation devrait être examinée puisque les données incluent une composante spatiale. Il convient toutefois de rappeler que la correction de cette non-indépendance tend généralement, toutes choses étant égales, à diminuer la confiance accordée à un effet statistique, alors que l'étude de 2005-2006 concluait déjà à l'absence d'effet de l'exposition à l'arsenic.

Les équations et graphiques représentant les relations linéaires entre variables montrent une inversion systématique entre variables explicatives et variables réponses (voir les figures 9, 19 et 20). Cette inversion soulève des enjeux méthodologiques, notamment parce que la transformation logarithmique appliquée aux concentrations urinaires d'arsenic n'est justifiable que lorsqu'elle concerne la variable réponse. Lorsqu'elle est appliquée à une variable explicative, cette transformation ne contribue pas au respect de la supposition de normalité des modèles. Le rapport ne permet pas de déterminer si cette erreur se retrouve dans l'ensemble des modèles statistiques utilisés.

Une relation entre l'âge et l'imprégnation à l'arsenic a été rapportée dans Gagné 2007 et formulée ainsi : « On remarque à la figure 9, une relation inverse entre l'âge et le niveau d'arsenic urinaire chez les exposés. C'est-à-dire que plus on avance en âge, moins on a des niveaux élevés d'arsenic urinaire. La grande dispersion des valeurs autour de la droite de régression montre cependant que l'équation de la droite de régression linéaire explique moins de 7 % de la variance. Bien que ce phénomène montre que la relation inverse de l'arsenic avec l'âge présente de fortes variations individuelles chez les exposés, elle n'en est pas moins plus importante que chez les non-exposés (Évain) comme on peut le constater au tableau 20. Cette relation est beaucoup plus faible chez les non-exposés. ». Cette analyse comporte plusieurs limites. Le modèle présuppose une relation linéaire entre l'âge et l'exposition sans fournir de justification biologique

soutenant cette hypothèse. Ensuite, comme indiqué précédemment, une inversion entre variable explicative et variable réponse est présente dans cette analyse.

En complément à cette analyse, une comparaison entre les classes d'âge traitées de manière catégorique a été réalisée séparément dans les groupes exposés et témoins (ANOVA, voir les tableaux 19 et 20). Cette stratégie apparaît méthodologiquement plus pertinente, puisqu'elle n'impose pas de relation linéaire entre âge et imprégnation. Il aurait cependant été nécessaire d'intégrer les deux groupes au sein d'un même modèle afin de permettre une comparaison statistique directe entre groupes. De plus, l'utilisation de la moyenne géométrique dans le Tableau 19 et de la moyenne arithmétique dans le Tableau 20 (illustrant les différences d'imprégnation entre classes d'âge pour le groupe exposé et témoin, respectivement) rend toute comparaison qualitative difficile entre exposés et témoins et soulève des interrogations quant à la cohérence des analyses effectuées.