



DataPOST : Analyse de données de polyexpositions et de santé des travailleurs

Note interdisciplinaire réalisée dans le cadre du Réseau doctoral de Santé Publique (RDSP)

Cassandra BARBEY

Thèse en Santé Publique, spécialité Santé, Environnement et Travail
ED Sciences de la vie et de la santé (SVS)

Département Métrologie des Polluants, INRS
Equipe ELIXIR, IRSET
EHESP

Sous la direction de Nathalie BONVALLOT et Frédéric CLERC



Contexte

L'exposome est un concept transversal à l'ensemble des disciplines de recherche, désignant l'ensemble des expositions auxquelles un individu est soumis tout au long de sa vie, depuis la période prénatale jusqu'à la mort (1). Il permet d'appréhender la santé comme le résultat d'interactions complexes entre facteurs biologiques, environnementaux et sociaux. Dans le contexte professionnel, il se décline sous le terme de worksome (2), ou exposome professionnel (3), regroupant l'ensemble des expositions spécifiques au travail et à son environnement ainsi que des facteurs modulant leurs effets sur la santé. Le travail occupe une place importante dans la vie des individus. C'est un déterminant social reconnu de la santé qui influence la qualité de vie au travers de ses différents aspects (économique, social ou psychologique) (4). Selon le contexte dans lequel il s'exerce, il peut avoir des effets ambivalents sur la santé des individus : il peut constituer un cadre favorable offrant un soutien social, financier et des opportunités personnelles, tout comme il peut exposer à des contraintes susceptibles de porter atteinte à l'équilibre physique et mental. Chaque exposition est susceptible d'avoir un effet individuel sur la santé des travailleurs dont l'intensité et la gravité dépendent de nombreux facteurs (durée d'exposition, fréquence, intensité, vulnérabilités individuelles...). Ces effets peuvent conduire à l'apparition de troubles variés, allant d'affections bénignes et réversibles à des maladies chroniques, voire mortelles.

Au quotidien, les situations de travail comportent généralement de multiples expositions pouvant avoir des effets d'interaction sur le travailleur. Elles peuvent non seulement amplifier les problèmes de santé connus, mais également engendrer de nouveaux risques. Ce phénomène, appelé « polyexposition », se définit comme l'exposition simultanée ou différée d'un individu à plusieurs agents, ou substances, pouvant interagir (5). La prise en compte de la polyexposition dans la santé au travail est relativement récente. Les mesures de préventions actuelles restent majoritairement conçues pour des expositions uniques, sans prendre en compte leurs possibles interactions. Malgré l'intérêt croissant de la communauté scientifique, les liens entre la polyexposition et la santé des travailleurs restent à ce jour insuffisamment documentés. De plus, l'évaluation de la polyexposition reste complexe et aucune démarche standardisée n'existe à ce jour. Néanmoins, les entreprises se retrouvent déjà confrontées à cette problématique notamment sur la question des mélanges de substances chimiques (article R. 4412-6 du Code du travail). Cependant, le manque de connaissance sur les effets combinés des expositions, l'absence de méthode d'évaluation harmonisée ainsi que le caractère individuel de l'élaboration du document unique d'évaluation des risques professionnels (DUERP) permettent de supposer que la majorité des entreprises reste principalement centrée sur la priorisation isolée des expositions, sans inclusion de possibles interactions.

Il apparaît donc nécessaire de développer les connaissances sur la polyexposition et ses effets sur la santé afin de faire évoluer les stratégies de prévention la concernant et de mettre au point de nouvelles méthodes d'évaluation permettant une prise en compte plus représentative des situations de travail. L'exploitation conjointe de multiples bases de données apparaît alors comme une solution prometteuse. Cette approche permet de combiner des informations hétérogènes présentant différents aspects d'une exposition, d'une maladie ou d'une situation de travail. Leur exploitation conjointe permet d'obtenir une vision plus complète de ces situations et favorise une meilleure compréhension des enjeux de santé au travail, tout en contribuant à l'amélioration des mesures de prévention existantes. C'est dans cette perspective que s'inscrit le projet « DataPOST : Analyse de données de polyexpositions et de santé des travailleurs », initié en 2021 par l'Institut national de recherche et de sécurité pour la prévention des accidents du travail et des maladies professionnelles (INRS). Ce projet vise à assembler plusieurs bases de données, provenant de l'INRS et de plusieurs organismes de santé au travail, afin de produire des portraits de la polyexposition et de les comparer avec les connaissances disponibles dans la littérature scientifique.

Résumé de la thèse

Le travail de thèse s'inscrit dans le cadre du projet DataPOST. Il a pour objectif de concevoir et d'appliquer une méthodologie d'analyse croisée de dix bases de données française en santé au travail afin d'orienter les perspectives de recherche sur la prévention des polyexpositions professionnelles. Cette analyse vise à : (1) identifier des profils de polyexposition professionnelle, (2) mettre en évidence les secteurs d'activité et les métiers les plus concernés, et (3) déterminer les disparités de profils et de maladies selon le sexe. La réponse à cet objectif a été organisée en trois grands axes faisant intervenir différents aspects du projet (Figure 1): l'axe « **Documentation** » permettant d'identifier et de rassembler les informations nécessaires à la compréhension du sujet, l'axe « **Conception** » permettant de concevoir une méthodologie permettant l'intégration d'informations hétérogènes provenant de plusieurs bases de données, et l'axe « **Application** » permettant de mettre en œuvre la méthodologie développée à travers plusieurs études de cas concrets liés à la santé au travail.

L'axe « **Documentation** » est constitué de deux parties :

- La première consiste en une recherche bibliographique visant à établir un état de l'art sur les connaissances disponibles en matière de polyexpositions professionnelles et des maladies qui leur sont associées (6). Une revue de 63 articles issus des bases de données Web of Science et PubMed a été menée selon un processus itératif en trois étapes : recherche, sélection par titre et abstract, et lecture. Les études retenues devaient intégrer des termes liés à la santé, aux expositions multiples et au travail. Trois types d'études ont été exclus, jugés hors du champ de cette revue : les recherches in vivo ou in vitro, celles portant spécifiquement sur des populations mineures, ainsi que les études sur les expositions prénatales. Vingt-huit pays et sept secteurs d'activités sont identifiés. Les problèmes de santé recensés se répartissent en six catégories : troubles du système nerveux (combinaisons chimiques), troubles auditifs (bruit et solvants), troubles mentaux (facteurs psychosociaux-organisationnels), troubles musculo-squelettiques (combinaisons physiques-biomécaniques et psychosociales-organisationnelles), ainsi que divers troubles affectant d'autres systèmes (cardiovasculaire, oculaire, rénal, cutané, biologique ou liés au déclin fonctionnel général). Dans l'ensemble, 91% des associations entre polyexposition et santé sont statistiquement significative. Cette revue montre néanmoins que peu d'études font état de la polyexposition en la reliant à des effets sur la santé et souligne que la prévention visant à réduire les risques multiples pour la santé des travailleurs est encore sous-développée.
- La deuxième réalise une description détaillée des bases de données utilisées. Chacune possède des caractéristiques propres, liées à leur objectif de création et aux missions des organismes détenteurs. Cela peut aller de la caractérisation des expositions professionnelles, à la surveillance des risques émergents, en passant par le développement du système assurantiel des maladies professionnelles. Ces bases peuvent être regroupées selon le type d'information qu'elles apportent : des informations uniquement sur les expositions professionnelles (six bases), des informations uniquement sur les maladies professionnelles (une base) ou des informations sur les deux (trois bases). La combinaison de ces bases de données fait actuellement l'objet d'un article en cours de rédaction avec les organismes partenaires du projet DataPOST.

L'axe « **Conception** » décrit la démarche méthodologique développée pour formaliser l'utilisation conjointe de plusieurs bases de données hétérogènes (7). Elle repose sur une ontologie, un référentiel structuré définissant les concepts et descriptions du projet, et sur une approche d'intégration de données, l'Ontology-Based Data Integration (OBDI), qui relie les données réelles, issues des bases, à des concepts théoriques. L'intégration se déroule en trois étapes :

1. **Définition du groupe professionnel** : Un clé statistique commune à l'ensemble des bases est conçue par combinaison des variables caractérisant les travailleurs à savoir le secteur d'activité, codé par la nomenclature d'activité française (NAF), le métier, défini par la nomenclature des professions et catégories socioprofessionnelles (PCS), et le sexe. Cette combinaison définit le groupe professionnel, correspondant à un ensemble d'individus de même sexe exerçant le même métier dans le même secteur d'activité.
2. **Identification des expositions professionnelles** : Les variables d'exposition sont standardisées afin de faciliter la mise en correspondance d'informations entre les bases. Les définitions des expositions présentes dans les protocoles de collecte sont comparées à celles du thésaurus des expositions professionnelles (TEP). Ces comparaisons sont évaluées et validées par un groupe d'expert constitué des représentants des organismes partenaires du projet.
3. **Standardisation des maladies professionnelles** : Les variables de maladies sont regroupées en bloc de maladies afin d'obtenir un volume de données suffisant par groupe professionnel pour les analyses. Cette étape repose sur la correspondance des variables avec leurs code équivalent dans la classification internationale des maladies, 10^{ème} révision (CIM-10).

L'axe « **Application** » illustre l'utilisation des données intégrées au travers de trois études de cas représentatives de situations concrètes liées à la santé au travail (7), permettant de répondre aux objectifs de la thèse :

- La première étude de cas repose sur une analyse quantitative visant à construire un **indicateur d'intérêt** pour chaque exposition et chaque maladie. Cet indicateur conçoit une valeur unique permettant de prioriser les situations les plus préoccupantes pour un groupe professionnel, un métier ou un secteur d'activité, et est représenté sous forme de heatmap. Il permet d'identifier des expositions préoccupantes confirmées par la littérature scientifique, permettant de valider sa méthode de construction, ainsi que de souligner les métiers et secteurs les plus à risque. L'indicateur d'intérêt permet une analyse transversale ne tenant pas compte des spécificités des bases permettant d'orienter les actions de prévention.
- La deuxième étude de cas repose sur une analyse qualitative, basée sur le calcul d'un **score de cohérence** pour chaque exposition et chaque maladie. Ce score mesure le degré de concordance entre les informations issues des différentes bases et permet d'identifier les expositions et maladies les plus et les moins documentées, toujours représentées sous forme de heatmap. Elle permet de mettre en évidence des disparités de scores entre les expositions et les maladies. Les expositions les plus récurrentes obtiennent généralement un score élevé, traduisant une bonne surveillance par les dispositifs de santé au travail. En revanche, certaines maladies récurrentes affichent des scores moyens, révélant des limites liées aux caractéristiques des bases de données et au cadre réglementaire de reconnaissance des maladies professionnelles. Le score de cohérence permet la mise en évidence de la complémentarité des informations.

- La dernière étude de cas porte sur le secteur de la construction. Deux méthodes d'analyses descriptives sont utilisées pour dresser un portrait de la polyexposition dans ce secteur. La première consiste en la production de **tableaux descriptifs** présentant les expositions, co-expositions et maladies pour chaque groupe professionnel, différenciés par sexe et complétés par des données de quantification des expositions. Cette approche met en lumière les disparités d'information entre hommes et femmes, les expositions et maladies les plus documentées, ainsi que celles associées aux plus fortes proportions de travailleurs exposés ou malades. Elle facilite la communication auprès des services de santé et de prévention. La seconde analyse repose sur une **matrice de co-expositions**, qui recense l'ensemble des associations d'expositions observées dans les bases. Elle permet d'identifier les co-expositions et familles d'expositions les plus fréquentes, d'évaluer la capacité de détection de l'outil en retrouvant des associations déjà connues, et de repérer des co-expositions encore peu étudiées par la littérature scientifique. Cet outil ouvre la voie à de futures recherches en santé et représente un premier pas vers une évaluation de la polyexposition professionnelle.

Le travail de thèse a montré l'intérêt d'exploiter conjointement plusieurs bases de données pour caractériser les situations de polyexposition et orienter les actions de prévention. Les analyses ont confirmé la validité de l'approche en retrouvant des co-expositions connues, tout en révélant des combinaisons encore peu étudiées ainsi que des disparités selon le sexe, les métiers et les secteurs, ouvrant de nouvelles perspectives de recherche. L'usage d'une ontologie a facilité la structuration des données et ouvert la voie à des outils d'analyse adaptés aux enjeux de santé au travail. Malgré certaines limites liées à l'hétérogénéité des bases et aux limites des données de maladies professionnelles, les résultats soulignent le potentiel de cette démarche pour faire évoluer les pratiques de prévention vers une approche plus globale, intégrant les polyexpositions et les différences entre populations de travailleurs.

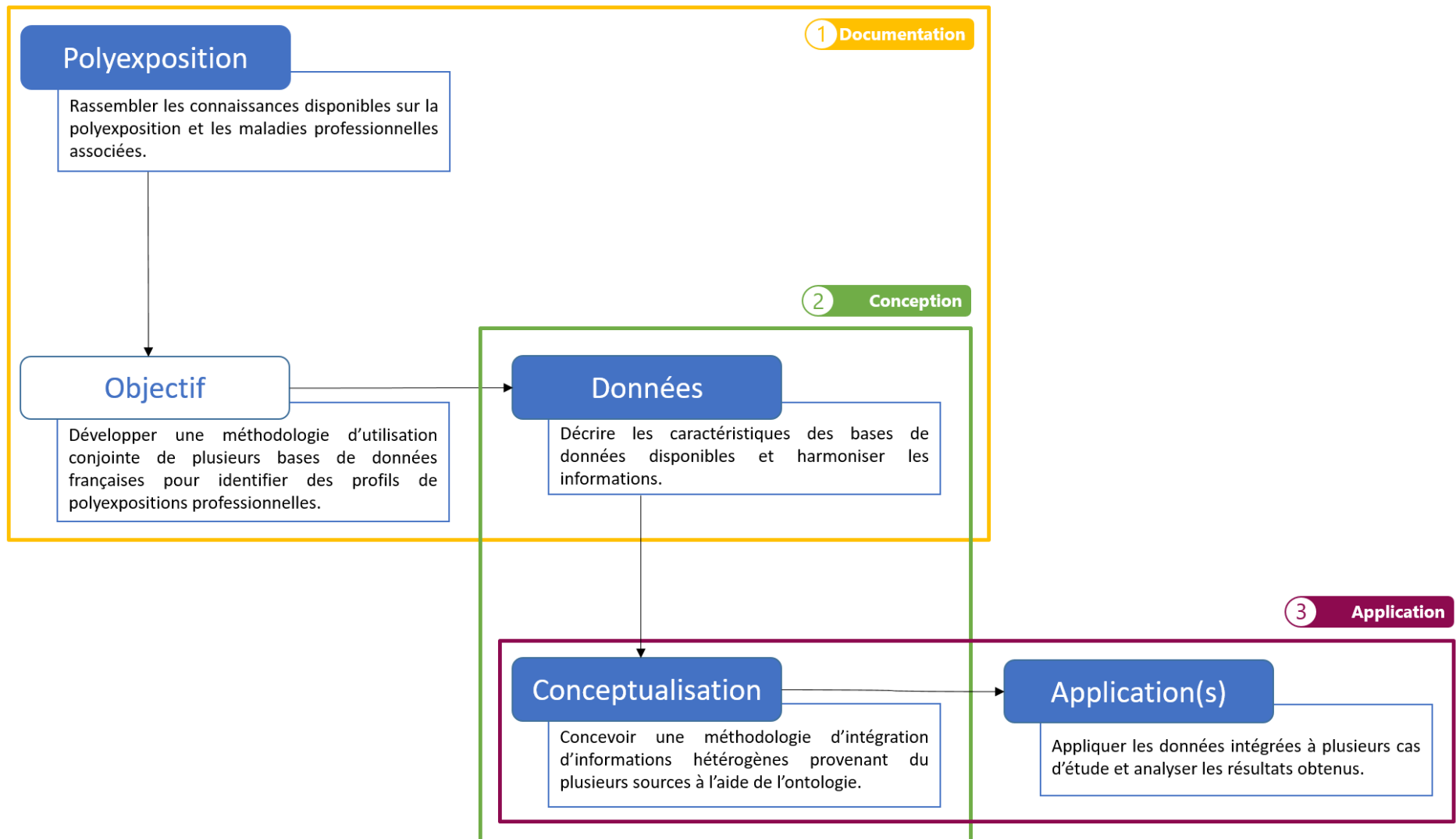


Figure 1. Stratégie générale de la thèse

Ouverture sur l'interdisciplinarité

1. Epidémiologie et biostatistiques

L'approche utilisée dans le cadre de cette thèse s'appuie sur le concept de *worksome*, elle intègre différents aspects d'une même exposition afin d'obtenir une caractérisation plus représentative des situations de travail : un aspect objectif avec l'inclusion de mesures réelles réalisées sur les lieux de travail, un aspect subjectif avec le ressenti des travailleurs vis-à-vis de leurs conditions de travail, un aspect normatif avec la reconnaissance d'exposition par utilisation de seuil réglementaire et un aspect historique avec l'utilisation de données, de documentation et d'études réalisées par le passé.

Le travail de conceptualisation permet de formaliser une démarche généralisable d'utilisation conjointe d'informations hétérogènes pour constituer une ontologie représentative de la problématique. Cette démarche, conçue pour être évolutive, peut être ajustée et appliquée à d'autres domaines que celui de la santé au travail. En intégrant des données complémentaires issues de disciplines diverses, elle peut être déclinée dans des contextes variés et contribuer à l'analyse de problématique de recherche dépassant le cadre initialement envisagé.

Cette approche permet l'identification de profils de travailleurs exposés, co-exposés ou malades, différenciés selon le sexe, le métier et le secteur d'activité. L'ajout d'informations sur l'imputabilité entre exposition et maladie permettrait de rendre ces profils encore plus précis et représentatifs. Elle s'inscrit dans une perspective « *outcome-wide* » (8) qui relie systématiquement un ensemble d'expositions et de co-expositions à plusieurs effets de santé. Une telle orientation offre la possibilité de mieux appréhender les interactions complexes entre conditions de travail et santé, tout en fournissant des bases solides pour des analyses épidémiologiques intégrées et différenciées, ainsi que des conseils de prévention plus adaptés.

2. Sciences des données et modélisation

La méthodologie repose sur la conception d'une unité statistique commune à l'ensemble des bases de données, ce qui permet de regrouper les données individuelles au sein de sous-populations de travailleurs et d'effectuer les analyses sur une échelle plus générale. L'utilisation de données intégrées, données standardisées à l'aide d'un référentiel commun, permet la conception d'une multitude de méthodes d'analyses, appliquées au sein d'études de cas concrètes, pouvant aller de la construction d'un indicateur de santé à l'étude d'un secteur spécifique, en passant par diverses représentations graphiques.

3. Santé, Environnement et Travail

L'identification des combinaisons d'expositions offre la possibilité d'intégrer progressivement la polyexposition dans l'évaluation des risques.

Sur le plan théorique, ces profils de co-expositions peuvent être confrontés aux connaissances disponibles dans la littérature scientifique. L'absence de résultats sur certains effets de santé suggère des risques insuffisamment étudiés et donc des mesures de prévention possiblement inadaptées.

Sur le plan pratique, la recherche ouvre la voie à la conception d'outils opérationnels pour les acteurs de terrain. Par exemple, des matrices hétérogènes d'expositions pourraient permettre d'identifier les polyexpositions selon les situations concrètes rencontrées en entreprise. En mobilisant les données existantes pour chaque exposition, il serait possible de regrouper ces dernières selon leurs effets sur un même organe, système biologique ou mécanisme d'action (9). Un code couleur, inspiré des heatmaps des associations proposées par Peters S et al. (10), pourrait mettre en évidence les situations de polyexposition pour lesquelles des effets sont documentés par la recherche. Les acteurs de terrain pourraient ainsi prioriser les situations présentant le ou les risques les plus élevés pour la santé, offrant ainsi des évaluations standardisées et généralisables, directement mobilisables par les entreprises. Ces outils pourraient être enrichis par des informations sur les méthodes de prévention identifiées dans la littérature scientifique. De plus, ils pourraient permettre également, en partenariat avec les entreprises, d'identifier les situations pour lesquelles aucun effet ou mesure de prévention n'est connu, orientant ainsi les futures recherches.

Le travail de recherche mené dans le cadre de cette thèse représente un levier concret pour améliorer l'évaluation des polyexpositions professionnelles sur le terrain, renforcer la prise de décision et soutenir la prévention systémique des situations de travail.

4. Sciences Humaines et Sociales, Sciences du Comportement

Ce travail de thèse met en évidence des enjeux sociologiques liés à la reconnaissance et à la déclaration des maladies professionnelles. Le biais de sous-déclaration résulte non seulement de limites réglementaires, mais aussi de facteurs sociaux et organisationnels tels que la méconnaissance des dispositifs de réparation et des droits sociaux par les travailleurs, la complexité et lourdeur des procédures administratives, la crainte de répercussions professionnelles ou encore les délais contraints pour l'instruction des dossiers (11). Ces freins sont documentés dans la littérature sociologique, qui souligne l'influence des inégalités sociales et de la précarité de l'emploi sur la capacité des travailleurs à déclarer une maladie professionnelle (12). Ces éléments contribuent à un manque de représentativité des données sur les maladies professionnelles et limitent la documentation de l'imputabilité entre une maladie et le travail.

5. Droits et Politiques de Santé

La prise en compte des polyexpositions en entreprise constitue un objectif pertinent pour la santé au travail, mais se heurte aux limites des compétences et aux priorités des entreprises, qui considèrent la prévention avant tout comme une contrainte réglementaire. Le travail présenté ici s'inscrit donc dans une perspective résolument de recherche, destinée à un public d'experts, dont sa diffusion à l'ensemble des entreprises françaises ne peut être envisagée à ce stade. L'objectif principal reste de développer et de renforcer les connaissances et méthodes dans ce domaine spécialisé, avant de pouvoir envisager de les appliquer plus largement lorsque la gestion des risques « simples » sera pleinement maîtrisée.

6. Conclusion

Cette thèse illustre l'intérêt d'une approche interdisciplinaire pour l'étude des polyexpositions professionnelles, combinant épidémiologie, modélisation, sciences humaines et politiques de santé. Elle permet de caractériser les situations de travail et les profils de travailleurs exposés, tout en ouvrant sur de nouvelles perspectives au-delà du champ de la santé au travail. S'appuyant sur l'approche d'exposome, elle intègre des facteurs sociologiques, opérationnels et environnementaux modulant l'effet du travail sur la santé, et contribue à une compréhension systémique des interactions entre conditions de travail et santé. Elle fournit ainsi un cadre méthodologique généralisable, capable de soutenir des analyses intégrées et des recommandations de prévention différenciées, scientifiquement fondées et adaptées aux besoins des acteurs de terrain.

Références

1. Wild CP. Complementing the Genome with an “Exposome”: The Outstanding Challenge of Environmental Exposure Measurement in Molecular Epidemiology. *Cancer Epidemiol Biomarkers Prev.* 2005 Aug 1;14(8):1847–50.
2. Eyles E, Manley D, Jones K. Occupied with classification: Which occupational classification scheme better predicts health outcomes? *Soc Sci Med.* 2019 Apr;227:56–62.
3. Roquelaure Y, Luce D, Descatha A, Bonvallot N, Porro B, Coutarel F. Un modèle organisationnel de l'exposome professionnel. *médecine/sciences.* 2022 Mar;38(3):288–93.
4. Armenti K, Sweeney M, Lingwall C, Yang L. Work: A Social Determinant of Health Worth Capturing. *Int J Environ Res Public Health.* 2023 Jan 10;20(2):1199.
5. Patrascu C, Clerc F, Pouyatos B. Risques multiples et polyexpositions : la prévention à l'épreuve de la complexité. :46.
6. Barbey C, Bonvallot N, Clerc F. Health Outcomes Related to Multiple Exposures in Occupational Settings: A Review. *Saf Health Work.* 2024 Dec;15(4):382–95.
7. Barbey C, Bonvallot N, Smaïl-Tabbone M, Clerc F. Integrating and Analysing Occupational Health Data Using a Multi-Ontology Approach. 2025;
8. Anguita-Ruiz A, Amine I, Stratakis N, Maitre L, Julvez J, Urquiza J, et al. Beyond the single-outcome approach: A comparison of outcome-wide analysis methods for exposome research. *Environ Int.* 2023 Dec;182:108344.
9. La Rocca B, Sarazin P. MiXie, an Online Tool for Better Health Assessment of Workers Exposed to Multiple Chemicals. *Int J Environ Res Public Health.* 2022 Jan 15;19(2):951.
10. Peters S, Undem K, Solovieva S, Selander J, Schlünssen V, Oude Hengel KM, et al. Narrative review of occupational exposures and noncommunicable diseases. *Ann Work Expo Health.* 2024 July 8;68(6):562–80.
11. Barlet B, Prete G. La sous-déclaration et la sous-reconnaissance des maladies professionnelles en France (1990–2020) : revue de littérature pluridisciplinaire. *Arch Mal Prof Environ.* 2022 Jan;83(1):27–44.
12. Kyung M, Lee SJ, Dancu C, Hong O. Underreporting of workers' injuries or illnesses and contributing factors: a systematic review. *BMC Public Health.* 2023 Mar 24;23(1):558.