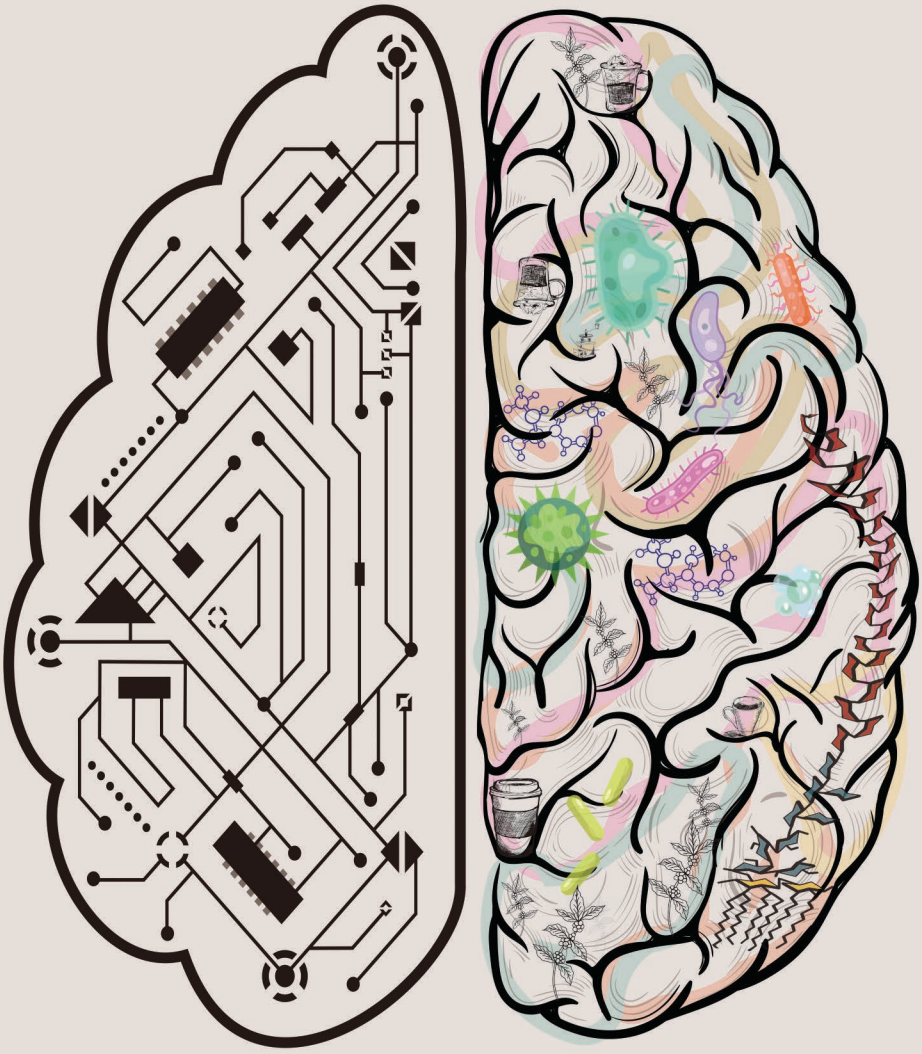
Exposomics to Unravel Parkinson's Disease Etiology



Yujia Zhao

Exposomics to Unravel Parkinson's Disease Etiology

Yujia Zhao

Yujia Zhao
Exposomics to Unravel Parkinson's Disease Etiology

Doctoral thesis, Utrecht University

Cover design: 大吉社
Layout by: Yujia Zhao
Printed by: Ipskamp Printing

ISBN: 978-90-393-7726-0
DOI: https://doi.org/10.33540/2422

Exposomics to unravel Parkinson's disease etiology

Exposomics om de etiologie van de ziekte van Parkinson te ontrafelen

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof. dr. H.R.B.M. Kummeling,
ingevolge het besluit van het College voor Promoties
in het openbaar te verdedigen op

dinsdag 8 oktober 2024 des middags te 2:15 uur

door

Yujia Zhao

geboren op 19 november 1993
te Shanxi, China

Promotor:
Prof. dr. R.C.H. Vermeulen

Copromotor:
Dr. S.M. Peters

Beoordelingscommissie:
Prof. dr. C.R. Berkers
Prof. dr. T.H. Hankemeier
Prof. dr. G. Pasterkamp
Prof. dr. W.M.M. Verschuren
Dr. R.H.S. Westerink

Dit proefschrift werd (mede) mogelijk gemaakt met financiële steun van de Stichting ParkinsonFonds en China Scholarship Council.

Table of Contents

Chapter 1

General Introduction

Parkinson's disease

Parkinson's disease (PD) is the second most common neurodegenerative disease after Alzheimer's disease. More than 10 million people worldwide are currently affected by PD. PD is an age-related disease, with incidence sharply rising at around age 65 (1). Over the past two decades, PD has emerged as the fastest-growing neurological disorder regarding disease burden (death and disability) around the world (2). Global estimates show that there may be nearly 13 million people living with PD by 2040, largely due to aging populations and increasing longevity (3).

PD is known as a movement disorder, characterized by four cardinal signs: tremor, rigidity, bradykinesia (slowness of movement) and postural imbalance (4). However, the clinical presentation of PD is multifaceted and includes many non-motor symptoms, such as olfactory loss, sleep dysfunction, autonomic dysfunction (e.g., constipation, urinary frequency), psychological or cognitive problems (e.g., cognitive decline, depression, anxiety) (4). Recognizing the onset of PD can be challenging, as there is typically an average delay of 10 years (prodromal phase) between the first noticeable symptom and the formal diagnosis (5). Individuals diagnosed with PD mostly experience a gradual progression of non-motor symptoms before the emergence of movement symptoms (4).

Genetic mutations play a role in only 5–10% of PD cases, indicating that the majority are related to a variety of lifestyle and environmental factors (6). While therapies and medicines are available to alleviate symptoms, there is no cure for the disease. PD is a slowly progressive disorder with decades of disease duration from prodromal to late stages, with patients increasingly requiring high medical care; thus, a substantial disease and economic burden are caused globally (2, 7).

Environmental exposures and Parkinson's disease

The hypothesis suggesting that environmental exposures could modifying PD risk emerged initially from the discovery of the neurotoxic effects of 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP), a byproduct of synthetic heroin. MPTP is converted into the toxic metabolite 1-methyl-4-phenylpyridinium (MPP^+), which induces an acute form of parkinsonism in humans by destroying dopaminergic neurons in the substantia nigra (8, 9). This kindled an interest in the potential role of exogenous toxicants in PD development, in particular compounds that are structurally similar to MPP^+, such as pesticides rotenone and paraquat. Epidemiological investigations have shown a consistent association between overall pesticide exposure and an increased risk of developing PD (10). Specific pesticide classes with suggested neurotoxic effects include insecticides rotenone and permethrin, organochlorines like beta-

hexachlorocyclohexane, and herbicides paraquat and 2,4-dichlorophenoxyacetic acid (2,4-D) (11, 12).

Besides pesticides, smoking and coffee consumption are two commonly reported inverse risk factors for PD. Smoking, contrary to its widely recognized adverse effects on cancers and other chronic diseases, exhibits neuroprotective effects in PD. Current smokers have approximately a 40% lower risk of PD compared to never smokers (13, 14). Similarly, coffee consumption has consistently shown an association with reduced PD risk in many epidemiological studies (15, 16). Both cigarette smoke and coffee are complex mixtures, and active components like nicotine and caffeine are assumed to contribute to their neuroprotective actions. The proposed mechanisms of action involve interactions with nicotinic receptors for nicotine and adenosine 2A receptor for caffeine.

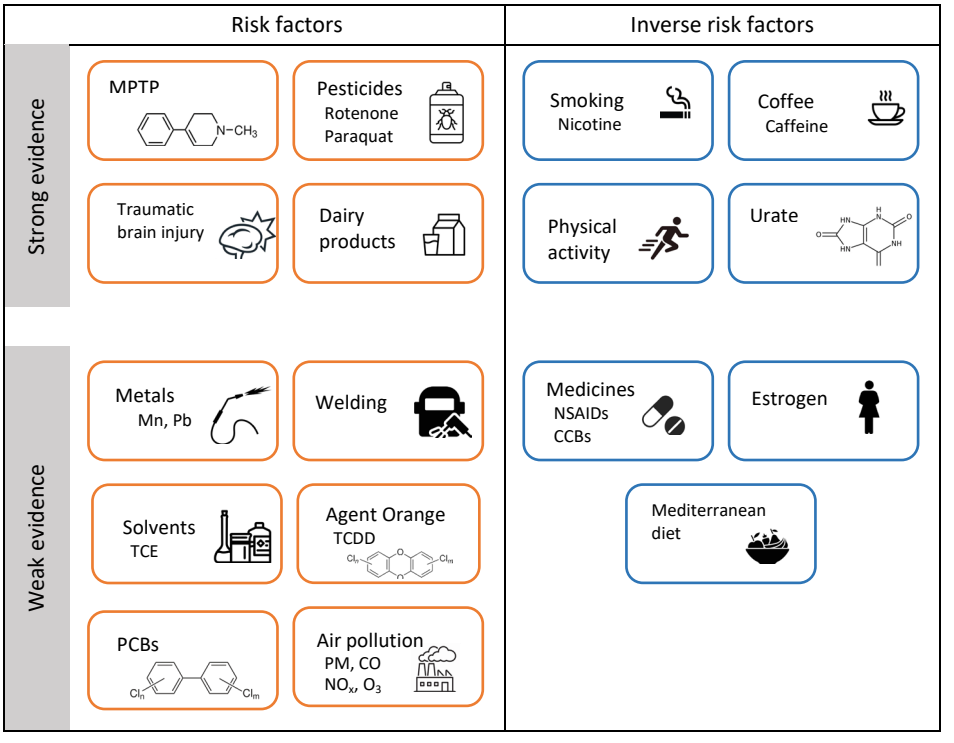
While certain environmental exposures demonstrate strong evidence, others yield inconsistent results and suspected associations were indicated. Exposure to various metals shows indications of potential links with PD. High-dose manganese exposure, which can be experienced in certain occupations such as welding and through contact with certain pesticides such as maneb and mancozeb, is known to cause a form of parkinsonism called ‘manganism’ (17). Exposure to lead may also be associated with a higher risk of PD (18, 19). Furthermore, there are various chemicals that have been associated with an increased PD risk, for example long-term exposure to chemical solvents, particularly trichloroethylene (TCE) (20). Occupational exposure to polychlorinated biphenyls (PCBs) has also been suggested to elevate PD mortality, especially in women (21). Agent Orange during the Vietnam War, containing herbicide 2,4-D and the highly toxic dioxin 2,3,7,8-tetrachlorodibenzodioxin (TCDD), resulted in a 30% higher PD incidence among exposed veterans compared to non-exposed veterans (22).

Although these suggested associations, confirming causality is challenging due to difficulties in accurate exposure measures. For instance, exposure to pesticides and welding is typically assessed through agricultural or occupational job history (11, 23), while smoking and coffee consumption rely on self-reported questionnaires (14, 16). These measures represent generalized estimates and may lead to exposure misclassification. Exposure-response analyses are limited, as exposure levels are often based on qualitative or semiquantitative data. Furthermore, these exposures are complex mixtures, and it is challenging to pinpoint the specific component(s) contributing to PD.

Beyond environmental chemicals, the gut microbiota is gaining attention as a potential mechanistic pathway for PD pathogenesis. This originated from the Braak’s hypothesis

of gut-to-brain spread of α-synuclein deposits, a pathological hallmark of PD (24). With the advent of high-throughput sequencing, the role of the gut microbiota has been specifically highlighted. Studies have shown alterations in the diversity and composition of the microbiota in PD patients, including enrichment of genera *Lactobacillus*, *Akkermansia*, and *Bifidobacterium*, and depletion of bacteria from the *Lachnospiraceae* family and the *Faecalibacterium* genus (25). This microbial dysbiosis may lead to a pro-inflammatory status, hyperpermeability of the gastrointestinal barrier, and α-synuclein misfolding in the enteric nervous system. However, these shifts in microbial profiles have not yet been linked to clinical presentations or other functional consequences of PD.

Box 1. An overview of environmental exposures related to Parkinson's disease risk



Evidence is indicated as stronger or weaker for each exposure. A stronger grade indicates an established association, based on consistent results in systematic reviews or meta-analysis. A weaker grade indicates a suspected association, where only incidental studies have reported the relationship, or where a systematic review or meta-analysis has not yielded consistent evidence.

Abbreviations: Mn, manganese; Pb, lead; TCE, trichloroethylene; TCDD, 2,3,7,8-tetrachlorodibenzodioxin; PCBs, polychlorinated biphenyls; PM, particular matter; NSAIDs, nonsteroidal anti-inflammatory drugs; CCBs, calcium channel blockers.

Untargeted high-resolution mass spectrometry for measuring the exposome

As a complement to the genome, the exposome is defined as environmental influences and associated biological responses throughout the lifespan, including exposures from the environment, diet, behavior, and endogenous processes (26). While no universal approach exists to measure the entire exposome, the human metabolome, which comprises all low-molecular weight (<2000 Da) metabolites in the human body, has emerged as one crucial exposome measure (27). The human metabolome can be measured from a variety of biological matrices ranging from biofluids such as blood, urine and feces, to organs, tissues or even cells. The metabolome includes not only metabolites directly linked to endogenous activities, but also those derived from food, medications, microbiota that inhabit the body, and environmental chemicals, serving as a crucial readout linking environmental effects to molecular mechanisms (28). High-resolution mass spectrometry (HRMS) is a widely used metabolomics platform due to its better sensitivity, high throughput and broad coverage of detectable metabolites (29). It is important to note that proteins, polymers, other large molecules and metals are not considered part of the metabolome because they require different measurement approaches.

One of the critical challenges for implementing the exposome in practice is developing methods that allow for measurement of exposures consistent with the scale experienced by an individual over a lifetime. To facilitate research in human metabolomics, the Human Metabolome Database (HMDB), a comprehensive and freely accessible online repository of human metabolites, has been established and consistently updated since 2007 (28). The 2021 version of HMDB contains records of 220,945 metabolites that have been detected and/or quantified in human tissues and biofluids. Unlike targeted analysis, which focuses on known, pre-defined analytes (up to a few hundreds), untargeted metabolomics has the capability to simultaneously detect a wide range of metabolites spanning various chemical classes. Especially with advancements in ultra-high-resolution mass spectrometry (UHRMS), the current untargeted metabolomic platform can measure 10,000-100,000 signals using sample volumes of less than 100µl (27, 30). Metabolomic profiling of blood samples in humans has indicated presence of metabolites from more than 80% of metabolic pathways documented in the Kyoto Encyclopedia of Genes and Genomes database, as well as a broad spectrum of environmental chemicals (31). Due to the ability to characterize an extensive series of endogenous and exogenous metabolites in biological samples, metabolomics has the potential to provide insights into the possible etiology and molecular mechanisms underlying disease of interest in epidemiological research.

Longitudinal studies for Parkinson's disease and the EPIC4PD cohort

Etiological studies of PD traditionally relied on a case-control study design, because PD remains relatively uncommon within the general population. However, many longitudinal studies have recently managed to accumulate sufficient cases through several decades of follow-up, which have the potential to largely enrich the understanding of the role of environmental factors in PD (1). Longitudinal studies have well-known advantages over case-control studies, as they reduce the reliance on participants' recall of past events. This is particularly crucial in the context of PD that often exhibits a long prodromal phase, and some pre-clinical symptoms may affect dietary and lifestyle choices.

Most of the large prospective cohorts with PD outcome ascertainment are based in the United States, including the Health Professionals Follow-up Study (HPFS) (number of participants at recruitment=51,529), Nurses' Health Study (NHS) (n=121,700) (32), Cancer Prevention Study-II Nutrition (CPS-IIN) (n=184,190) (33), and National Institutes of Health-American Association of Retired Persons Diet and Health Study (NIH-AARP) (n=318,260) (34). These cohorts have assessed the impact of diet and nutrients, smoking and alcohol consumption, and air pollution on PD risk, predominantly adopting external exposure assessment methods.

In Europe, there is a large cohort available to study PD: the European Prospective Investigation into Cancer and Nutrition (EPIC), which was initiated between 1992 and 2000 across ten countries (35). EPIC was originally designed to investigate the relationship between nutrition and cancer, and a subset of 220,494 subjects, known as the EPIC4PD cohort, underwent screening for PD diagnosis (36). At baseline, information on dietary intake and lifestyle data (smoking and alcohol consumption, physical activity, hormone therapy, occupation) was collected through a questionnaire. Anthropometric measurements were performed using a standardized approach. Blood samples were drawn at recruitment, from which plasma and erythrocytes were separated and aliquoted for future analysis (35). Up until 2012, the follow-up rate for the EPIC4PD cohort was 98.5%, with a median follow-up period of 12.8 years (maximum of 20.8 years). PD cases were initially identified through medical record linkage and then validated by movement disorder experts (36). In total, 734 incident PD cases were identified, each assigned a reliability label of diagnosis based on the quality of clinical data and expert confidence.

Thesis aims and outline

This thesis utilized the EPIC4PD cohort to explore etiological roles of multiple environmental exposures in PD occurrence, with a particular focus on applying metabolomics for exposure assessment. The specific aims and outline of the thesis are described below.

The goal of **Section One** (Chapters 2 and 3) was to summarize current evidence and to conduct an original study within the EPIC4PD cohort to explore the impact of metal exposures on PD. In **Chapter 2**, we performed a systematic review and meta-analysis to assess the evidence of existing epidemiologic studies on the relationship between metal exposures and PD risk. In **Chapter 3**, within the EPIC4PD cohort, we carried out a nested case-control study, measuring eleven metal species in erythrocytes collected before PD diagnosis. We then explored the association between prediagnostic metal levels and PD risk.

In **Section Two** (Chapters 4 and 5), our focus was on developing an untargeted metabolomics framework for internal exposure assessment in human studies. In **Chapter 4**, we introduced a comprehensive approach for evaluating a specific group of chemicals, exemplified by dioxin(-like) exposures in an occupational population. Apart from well-known parental dioxin(-like) compounds, we identified related halogenated compounds and metabolites of parent compounds using untargeted metabolomics. These identified chemicals composed a dioxin(-like) exposure portfolio (chemical-wide), which was then used to study their impact on the biological system using a combination of metabolome-wide analyses and associated immune responses. **Chapter 5** extended this chemical-wide approach to assess compounds related to caffeine metabolism. Relationship between caffeine intake and PD risk was first evaluated within the full EPIC4PD cohort using self-reported coffee consumption as a proxy for exposure to caffeine. Subsequently, we measured caffeine and its metabolites in blood within the aforementioned nested case-control study within EPIC4PD. This enabled us to investigate the frequently suggested inverse associations between caffeine and PD risk, using objective biomarkers.

In **Section Three** (Chapters 6 and 7), we aimed to provide insights into the microbiota-gut-brain axis in PD using prediagnostic biomarkers in the nested EPIC4PD case-control study. In **Chapter 6**, we assessed lipopolysaccharide-binding protein (LBP), a blood marker for bacterial invasion and intestinal permeability, and explored its association with PD risk. In **Chapter 7**, using untargeted metabolomics and most updated database of gut microbial metabolites, we characterized 167 microbial metabolites and examined their prospective associations with PD risk.

In **Chapter 8**, I summarized the main findings of this thesis and discuss the applications of metabolomics in internal exposome measurement, the role of environmental factors in PD, methodological considerations of this thesis, and future directions.

References

1. Ascherio A, Schwarzschild MA. The epidemiology of Parkinson's disease: risk factors and prevention. The Lancet Neurology. 2016;15(12):1257-72.

2. Global, regional, and national burden of neurological disorders during 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. The Lancet Neurology. 2017;16(11):877-97.

3. Dorsey ER, Bloem BR. The Parkinson Pandemic-A Call to Action. JAMA Neurol. 2018;75(1):9-10.

4. Armstrong MJ, Okun MS. Diagnosis and Treatment of Parkinson Disease: A Review. JAMA. 2020;323(6):548-60.

5. Gaenslen A, Swid I, Liepelt-Scarfone I, Godau J, Berg D. The patients' perception of prodromal symptoms before the initial diagnosis of Parkinson's disease. Mov Disord. 2011;26(4):653-8.

6. Tysnes OB, Storstein A. Epidemiology of Parkinson's disease. J Neural Transm (Vienna). 2017;124(8):901-5.

7. Yang W, Hamilton JL, Kopil C, et al. Current and projected future economic burden of Parkinson's disease in the U.S. NPJ Parkinson's disease. 2020;6:15.

8. Ballard PA, Tetrud JW, Langston JW. Permanent human parkinsonism due to 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP): seven cases. Neurology. 1985;35(7):949-56.

9. Langston JW, Ballard P, Tetrud JW, Irwin I. Chronic Parkinsonism in humans due to a product of meperidine-analog synthesis. Science. 1983;219(4587):979-80.

10. van der Mark M, Brouwer M, Kromhout H, Nijssen P, Huss A, Vermeulen R. Is pesticide use related to Parkinson disease? Some clues to heterogeneity in study results. Environ Health Perspect. 2012;120(3):340-7.

11. Brown TP, Rumsby PC, Capleton AC, Rushton L, Levy LS. Pesticides and Parkinson's disease--is there a link? Environ Health Perspect. 2006;114(2):156-64.

12. Dick FD. Parkinson's disease and pesticide exposures. Br Med Bull. 2006;79-80:219-31.

13. Gallo V, Vineis P, Cancellieri M, et al. Exploring causality of the association between smoking and Parkinson's disease. Int J Epidemiol. 2019;48(3):912-25.

14. Mappin-Kasirer B, Pan H, Lewington S, et al. Tobacco smoking and the risk of Parkinson disease: A 65-year follow-up of 30,000 male British doctors. Neurology. 2020;94(20):e2132-e8.

15. Palacios N, Gao X, McCullough ML, et al. Caffeine and risk of Parkinson's disease in a large cohort of men and women. Mov Disord. 2012;27(10):1276-82.

16. Kim IY, O'Reilly É J, Hughes KC, Gao X, Schwarzschild MA, Ascherio A. Differences in Parkinson's Disease Risk with Caffeine Intake and Postmenopausal Hormone Use. Journal of Parkinson's disease. 2017;7(4):677-84.

17. Lucchini R, Tieu K. Manganese-Induced Parkinsonism: Evidence from Epidemiological and Experimental Studies. Biomolecules. 2023;13(8).

18. Coon S, Stark A, Peterson E, et al. Whole-body lifetime occupational lead exposure and risk of Parkinson's disease. Environ Health Perspect. 2006;114(12):1872-6.

19. Weisskopf MG, Weuve J, Nie H, et al. Association of cumulative lead exposure with Parkinson's disease. Environ Health Perspect. 2010;118(11):1609-13.

20. Goldman SM, Weaver FM, Stroupe KT, et al. Risk of Parkinson Disease Among Service Members at Marine Corps Base Camp Lejeune. JAMA Neurology. 2023;80(7):673-81.

21. Steenland K, Hein MJ, Cassinelli RT, 2nd, et al. Polychlorinated biphenyls and neurodegenerative disease mortality in an occupational cohort. Epidemiology. 2006;17(1):8-13.

22. Song S, Kim JY, Lee Y, Jeong H, Kim S, Lee EE. Effects of defoliant exposure and medication use on the development of Parkinson's disease in veterans. Age Ageing. 2023;52(10).

23. Mortimer JA, Borenstein AR, Nelson LM. Associations of welding and manganese exposure with Parkinson disease: review and meta-analysis. Neurology. 2012;79(11):1174-80.

24. Braak H, Del Tredici K, Rüb U, de Vos RA, Jansen Steur EN, Braak E. Staging of brain pathology related to sporadic Parkinson's disease. Neurobiol Aging. 2003;24(2):197-211.

25. Romano S, Savva GM, Bedarf JR, Charles IG, Hildebrand F, Narbad A. Meta-analysis of the Parkinson's disease gut microbiome suggests alterations linked to intestinal inflammation. NPJ Parkinson's disease. 2021;7(1):27.

26. Miller GW, Jones DP. The nature of nurture: refining the definition of the exposome. Toxicol Sci. 2014;137(1):1-2.

27. Walker DI, Valvi D, Rothman N, Lan Q, Miller GW, Jones DP. The metabolome: A key measure for exposome research in epidemiology. Curr Epidemiol Rep. 2019;6:93-103.

28. Wishart DS, Guo A, Oler E, et al. HMDB 5.0: the Human Metabolome Database for 2022. Nucleic Acids Res. 2022;50(D1):D622-D31.

29. Zhang X-w, Li Q-h, Xu Z-d, Dou J-j. Mass spectrometry-based metabolomics in health and medical science: a systematic review. RSC Advances. 2020;10(6):3092-104.

30. Uppal K, Walker DI, Liu K, Li S, Go YM, Jones DP. Computational Metabolomics: A Framework for the Million Metabolome. Chem Res Toxicol. 2016;29(12):1956-75.

31. Jones DP, Park Y, Ziegler TR. Nutritional metabolomics: progress in addressing complexity in diet and health. Annu Rev Nutr. 2012;32:183-202.

32. Gao X, Cassidy A, Schwarzschild MA, Rimm EB, Ascherio A. Habitual intake of dietary flavonoids and risk of Parkinson disease. Neurology. 2012;78(15):1138-45.

33. Palacios N, Gao X, O'Reilly E, et al. Alcohol and risk of Parkinson's disease in a large, prospective cohort of men and women. Mov Disord. 2012;27(8):980-7.

34. Liu R, Guo X, Park Y, et al. Caffeine intake, smoking, and risk of Parkinson disease in men and women. Am J Epidemiol. 2012;175(11):1200-7.

35. Riboli E, Hunt KJ, Slimani N, et al. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. Public health nutrition. 2002;5(6b):1113-24.

36. Gallo V, Brayne C, Forsgren L, et al. Parkinson's Disease Case Ascertainment in the EPIC Cohort: The NeuroEPIC4PD Study. Neurodegener Dis. 2015;15(6):331-8.

Chapter 2

Metal Exposure and Risk of Parkinson's Disease: A Systematic Review and Meta-analysis

Yujia Zhao, Anushree Ray, Lützen Portengen, Roel Vermeulen, Susan Peters

Corresponding author: Susan Peters

Published in *American Journal of Epidemiology*, July 2023

https://pubmed.ncbi.nlm.nih.gov/37022311/

Abstract

Metal exposure has been suggested as a possible environmental risk factor for Parkinson's disease (PD). We searched the PubMed, EMBASE and Cochrane databases to systematically review the literature on metal exposure and PD risk and to examine the quality of the overall study and exposure assessment method. A total of 83 case-control studies and five cohort studies were included, of which 73 were graded as low or moderate overall quality. 69 studies adopted self-reported exposure and biomonitoring after disease diagnosis for exposure assessment approaches. The meta-analyses showed that concentrations of copper and iron in serum, and zinc in either serum or plasma were lower, while concentrations of magnesium in CSF and zinc in hair were higher in PD cases compared to controls. Cumulative lead levels in bone were found to be associated with increased risk of PD. We did not find associations between other metals and PD. The current level of evidence for associations between metals and PD risk is limited as biases from methodological limitations cannot be ruled out. High-quality studies assessing metal levels *before* the disease onset are needed to improve our understanding of the role of metals in the etiology of PD.

Key words: Parkinson's disease, metals, systematic review, meta-analysis

Introduction

Parkinson's disease (PD) is the second most frequent neurodegenerative disease, characterized by movement dysfunctions including bradykinesia, muscular rigidity, rest tremor and postural instability. The pathological features of PD are represented by the selective degeneration of dopaminergic neurons in the substantia nigra pars compacta and the Lewy body inclusions, leading to dopamine deficiency and motor defects (1). The estimated incidence is 14 per 100,000 overall and increases sharply to 160 per 100,000 people after the age of 65 (2). The global burden of PD has more than doubled over the past decades, showing a faster growth than any other neurological disorder (3), which cannot be fully explained by the increasing aging population.

Although the precise pathological mechanisms remain undetermined, current thinking is that PD arises from an interaction between genetic and environmental factors. Causative genetic mutations only explain a small proportion of PD patients, and about 90% of cases are sporadic, suggesting a significant role for environmental risk factors (2). Among these factors, heavy metal exposure is one of the concerns in PD pathogenesis. Possible mechanisms of metals in the onset and progression of PD include mitochondrial dysfunction and oxidative stress, promotion of α-synuclein aggregation and fibril formation, and activation of microglial cells and inflammation (4, 5). Human studies have shown that manganese inhalation from mining and welding fume could induce parkinsonism (6), and dental amalgam filling restoration was associated with an elevated risk of PD (7). Moreover, numerous studies on specific metals and PD risk have been published, but results are inconsistent. Methodological limitations may hinder drawing conclusions on the associations between metal exposures and PD risk.

Here, we conducted a systematic review and meta-analysis aiming to evaluate the current epidemiological evidence of associations between metal exposure and the risk of PD, with specific consideration of the quality of studies and validity of the exposure assessment methods.

Materials and methods

Study search strategy

We searched the PubMed, EMBASE and Cochrane databases up to July 2021. The search string consisted of a combination of medical subject headings [MeSH] and text words (search queries are provided in Table S1). We included 'Parkinson's disease', 'Parkinson*', 'PD' and 'neurodegenerative*' terms for PD, in combination with 'metal' and terms for specific metals (aluminum, calcium, cadmium, chromium, copper, iron,

lead, magnesium, manganese, mercury, nickel, selenium, zinc), as well as 'exposure' or 'exposed'. We further scrutinized the reference lists of relevant reviews and meta-analyses for additional publications.

Inclusion-exclusion criteria and study selection

Eligible publications in our systematic review were selected based on the following criteria: 1) original, peer-reviewed research papers; 2) human observational studies: case-control or cohort design; 3) exposure included the metal species as described above or general metal exposure; 4) outcome was sporadic PD; and 5) written in English. Exclusion criteria were: 1) animal study; 2) review, case report and case series, editorial, letter or conference abstract without original data; 3) repeated or overlapping publication; 4) exposure was welding or welding fume, not estimating specific or general metal exposure; and 5) outcome was parkinsonism, manganism, motor dysfunction, or neuropsychological dysfunction.

After duplicate removal, all articles were screened by title and abstract to exclude records on irrelevant topics and based on the exclusion criteria. Full texts for the remaining articles were retrieved and assessed by one reviewer (YZ). Any uncertainty was discussed with the second reviewer (SP). In case of multiple publications from the same study, the most complete and/or most recent paper was included. Re-analyses of previously reported studies without updates on the association between metal exposure and PD were excluded.

Data extraction

The following information was extracted from the candidate articles: first author's last name, year of publication, country or region, study design, sample size, age and gender distribution of participants, case ascertainment and control selection, matching variables or adjusted confounders, exposure assessment method, and analysis technique for measuring metal levels. Additional information for cohort studies included the follow-up period and the number of cases that developed the outcome (PD diagnosis/mortality).

For studies with quantitative exposure assessment, the mean and standard deviation (SD) of metal concentrations for the case and control groups were collected. When the mean and/or SD were not available, alternative statistic parameters for location (median, geometric mean), variability (geometric standard deviation, standard error (SE), interquartile range (IQR), range) and alternative statistical tests (t statistic, p-value, confidence interval (CI)) were considered. For studies only presenting numerical data in figures, WebPlotDigitizer was used for digitizing the data points from the figure. For studies with dichotomous/ordinal exposure categories, the numbers in each category

from each group and the crude/adjusted odds ratio (OR) or relative risk (RR) with 95% CI were extracted.

Study quality assessment

Study quality was assessed in terms of both study design and exposure assessment (EA) method. The Newcastle-Ottawa Scale (NOS) (8) was adapted separately for case-control studies and cohort studies (Table S2-S3). There were four parameters to evaluate the methodological quality: subject selection, comparability of the groups, ascertainment of either exposure or outcome for case-control or cohort studies respectively, and statistical analysis. We then appraised the EA methods using an adapted version of a previously published EA rating system (9) (Table S4). EA methods were considered as uninformative (EA score of 1) when based on self-reported exposure, which could have led to non-differential misclassification, or registry job history/self-reported job history in industrial cohorts, which are often inaccurate and incomplete (10). Biomonitoring, environmental monitoring and food frequency questionnaires (FFQ) after disease onset were considered not completely valid (EA score of 2) due to possible reverse causation, while bone level measurements of lead, cadmium and chromium after disease onset were regarded as accurate (EA score of 4) due to their slow elimination from the human body. An EA score of 3 was given to job histories from company records, a valid but not agent-specific approach. Approaches considered as valid and agent-specific (EA score of 4), included a job-exposure matrix (JEM), case-by-case expert assessment, and environmental monitoring or biomonitoring before disease onset. Two reviewers (YZ and AR) independently performed the quality assessment of all selected studies. Any disagreements were discussed between the two reviewers, and if no consensus was reached, they were resolved by a third reviewer (SP).

Statistical analysis

For case-control studies assessing metals in biological matrices (except for bone), the between-group standardized mean difference (SMD) (Hedges' *g*) was used as the effect measure for each study. SMD was calculated using the mean and SD on the log-transformed scale (11), due to skewed distributions and small sample sizes in many of the included studies. For case-control studies assessing dietary and occupational/environmental metal exposures, the OR for 'ever/higher metal exposure' versus 'never/background metal exposure' was used as the effect measure for each study. Covariate-adjusted ORs were preferred over crude ORs to reduce possible confounding. When studies reported ORs for stratified exposure groups (e.g., quartiles, as was done in seven studies), the pooled OR for a single study was calculated by within-study random-effects meta-analysis of the nonreference groups (12). When the mean/SD or OR/SE were not available, they were estimated from alternative statistics

according to recommendations of the Cochrane Handbook (13). When metal levels in the same matrix or source were presented as continuous data in some studies and as categorical data in other studies, effect measures, SMDs and ORs, were mutually converted by the formula, $SMD = \sqrt{3}/\pi \ln OR$ (14). All formulae are provided in the Formula appendix.

Meta-analyses were conducted for each of the different metals (aluminum, calcium, cadmium, chromium, copper, iron, lead, magnesium, manganese, mercury, nickel, selenium, zinc , and general metal exposure) from various biological matrices (bone, cerebrospinal fluid (CSF), hair, (whole)blood, erythrocyte, plasma, serum, urine) and sources (diet, occupation/environment) separately, provided there were at least two studies remaining when low-quality papers were excluded. Studies assessing plasma and serum were additionally combined because they both assessed metals in the blood. As considerable between-study heterogeneity was anticipated, a random-effects model was used to pool effect sizes. The restricted maximum likelihood estimator (15) was used to calculate the heterogeneity variance tau-squared, τ^2. Knapp-Hartung adjustment (16) was applied to calculate the confidence interval around the pooled effect.

Cochran's *Q*-test and I^2 (17) were used to assess and quantify between-study heterogeneity. A *P* value less than 0.05 was considered significant statistical evidence of heterogeneity. I^2 values below 25% were deemed low, 25-75% as medium and above 75% as high degree of heterogeneity (17). In an attempt to explain heterogeneity, subgroup analyses for geological locations and detection methods were performed if the original meta-analysis contained at least ten studies. Separate estimates of τ^2 were assumed in each subgroup. To explore the robustness of meta-analyses, we calculated different influence diagnostics (DFFITS value, Cook's distance, hat value, DFBETAS value) of individual studies based on the leave-one-out method, each time omitting one study. A study was considered as influential if any of the above influential measures reached the chosen cut-offs (18). The presence of publication bias was checked using a funnel plot and Egger's test (19) if the number of studies was more than ten, and then applying the trim-and-fill method (20). Analyses were performed with the *meta*, *metafor* and *dmetar* packages in R 3.6 (21).

Results

Study selection

After duplicate removal, a total of 4,045 papers from multiple electronic databases as well as relevant reviews were screened. From these, 83 case-control studies and five cohort studies were selected based on the inclusion and exclusion criteria (Figure 1).

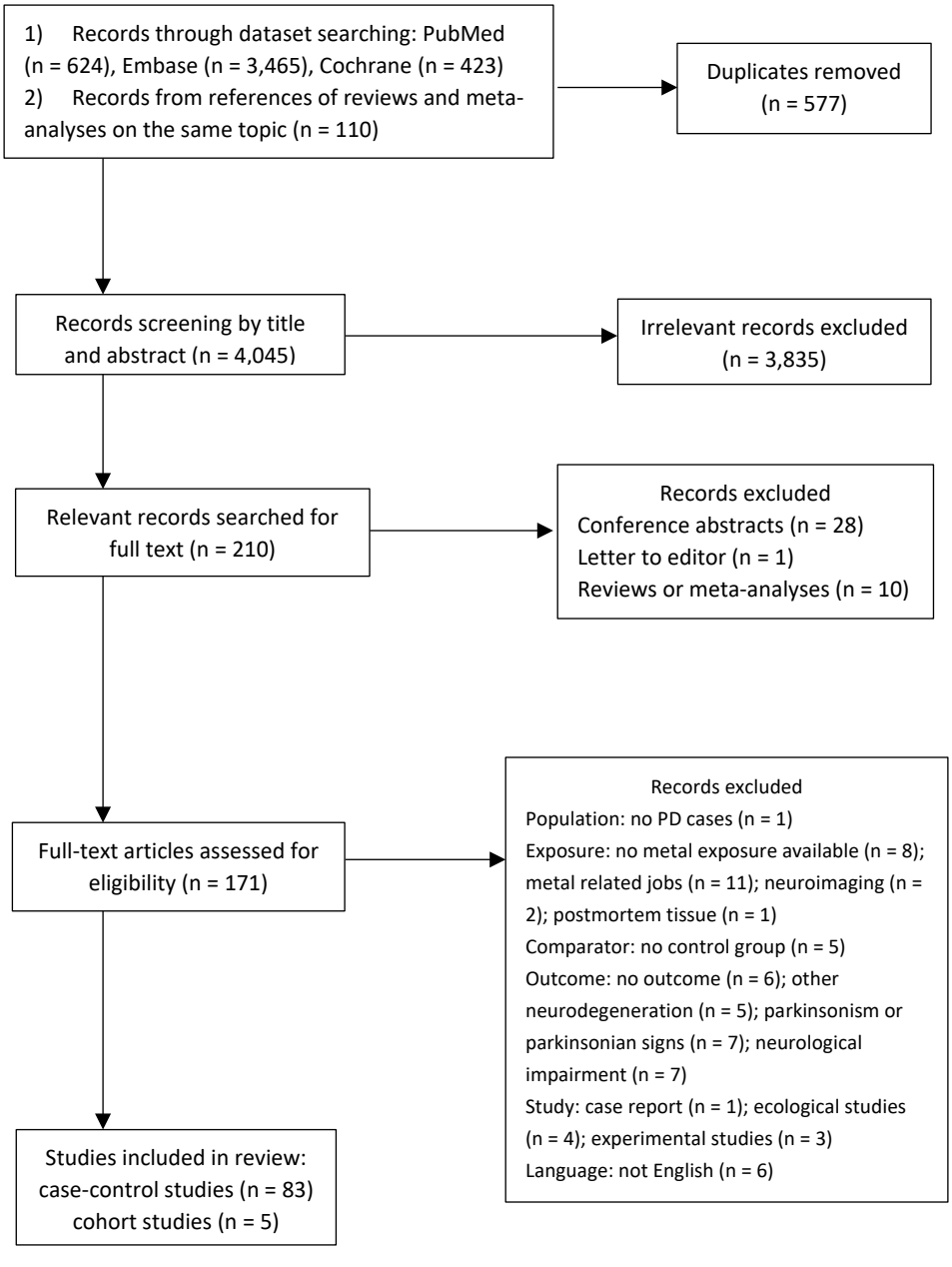


Figure 1. Flowchart of study selection

Basic information of candidate studies is summarized in Table S5 for case-control studies (22-104) and Table 1 for cohort studies (105-109). Overall, 35 (40%) of selected studies were carried out in Europe, 21 (24%) in Asia, 22 (25%) in North America and 10 (11%) in other parts of the world.

The number of case-control studies concerning each metal in different biospecimens/sources are presented in Table 2. Many studies (n=48, 58%) assessed more than one type of metal and 24 (29%) included more than one biological matrix or exposure source. The metals and exposure sources varied among the five cohort studies (Table 1).

Quality assessment

Study quality assessment results for all included papers are shown in Tables S6-S7. For the general study quality, most case-control studies (n=66, 80%) were scaled as moderate-quality, four as low-quality and thirteen as high-quality (Table 3). Three cohort studies were deemed as moderate-quality (107-109) and two as high-quality (105, 106). Concerning exposure assessment methods, most case-control studies (n=75, 90%) adopted uninformative or invalid approaches (EA scores of 1 or 2). Eight studies used more reliable methods (EA score of 4). All cohort studies assessed metal exposure before disease onset (EA score of 4).

Meta-analyses of metal levels in biological matrices

The majority of meta-analyses were based on less than five studies, and most of them included less than 250 PD cases (Table 4). Pooled SMDs for aluminum, calcium, chromium, manganese, mercury, nickel, and selenium did not show any statistically significant difference between PD cases and controls in any biospecimen. Statistically significant differences in effect size were observed for cadmium in blood (n=2, SMD -0.61 (95%CI -1.08, -0.13)), copper in serum (n=18, SMD -0.43 (95%CI -0.84, -0.02)), iron in serum (n=27, SMD -0.28 (95%CI -0.56, 0.00)), zinc in plasma or serum (n=18, SMD -0.53 (95%CI -0.92, -0.14)), which were lower in PD cases than in controls, and for magnesium in CSF (n=5, SMD 0.66 (95%CI 0.41, 0.91)) and zinc in hair (n=4, SMD 0.52 (95%CI 0.14, 0.90)), which were higher in PD cases. Forest plots of the meta-analyses of copper, iron, and zinc in plasma/serum, from >15 studies, are shown in Figure 2-4.

The two included studies on lead levels in bone reported an increased risk of PD for individuals with higher overall lead bone levels relative to the lowest quartile (OR 1.34 (95%CI 1.02, 1.76) (48); OR 1.32 (95%CI 1.04, 1.66) (65)). Further, a positive exposure-response relationship was observed for tibia bone lead (p_{trend}=0.012 (48); p_{trend}=0.06 (65)).

Table 1. Overview of cohort studies on metal exposure and Parkinson's disease

Study	Country	Cohort	Subjects	Exposure source and assessment method	Follow-up period	Outcome	Metal	Relative risk (95% CI)*
Logroscino 2008	USA	Health Professionals Follow-up Study (HPFS); Nurses' Health Study (NHS)	47,406 men from HPFS; 76,947 women from NHS	diet; food frequency questionnaire	1986-2000 for HPFS 1984-2000 for NHS	PD incidence	Iron	1.10 (0.92, 1.33) [a, #]
Feldman 2011	Sweden	Swedish Twin Registry	20,225 men	occupation; job exposure matrix	1967/1973-2009	PD incidence	Metals (non-specified)	0.9 (0.4-1.8) [b]
Palacios 2014	USA	Nurses' Health Study (NHS)	97,430 women	air; environment monitoring	1990-2008	PD incidence	Cadmium Chromium Lead Manganese Mercury Nickle	1.01 (0.86, 1.19) [c, #] 0.90 (0.76, 1.07) [c, #] 0.94 (0.80, 1.11) [c, #] 1.12 (0.95, 1.33) [c, #] 1.23 (1.05, 1.46) [c, #] 0.99 (0.85, 1.15) [c, #]
Brouwer 2015	The Netherlands	Netherlands Cohort Study on diet and cancer (NLCS)	58,279 men	occupation; job exposure matrix	1986-2003	PD mortality	Metals (non-specified)	1.02 (0.77, 1.35) [d, #]
Vinceti 2016	Italy	residents who consumed the high-Se tap water and lesser exposed comparison group	exposed cohort: 2,065; unexposed cohort: 95,715	drinking water; environment monitoring	1986-2012	PD mortality	Selenium	2.47 (1.15, 5.28) [e]

* Relative risk (RR) for ever/higher exposure versus never/background exposure
[#] Calculated from RRs in nonreference exposure groups using within-study random-effects meta-analysis
[a] Adjusted for age, smoking, total energy, caffeine, BMI, vitamin C, vitamin E, lactose, physical activity and alternate healthy eating index
[b] Adjusted for age, education and smoking
[c] Adjusted for age, smoking and population density
[d] Adjusted for smoking, non-occupational physical activity and BMI
[e] Adjusted for age and calendar year

Table 2. The number of case-control studies for each metal, categorized into different biospecimens or sources

	Blood	Bone	CSF	Erythrocyte	Hair	Plasma	Serum	Urine	Diet	Occupation/ environment	Total of studies
Aluminium	1	NA	4	NA	3	NA	4	1	NA	1	9
Calcium	1	NA	4	NA	3	NA	5	1	4	NA	13
Cadmium	2	NA	2	NA	1	NA	2	2	NA	2	7
Chromium	1	NA	5	NA	1	NA	6	3	NA	1	11
Copper	3	NA	11	1	3	8	19	5	3	6	45
Iron	1	NA	11	NA	4	5	28	4	6	2	46
Lead	3	2	4	NA	1	1	4	2	NA	5	17
Magnesium	1	NA	5	NA	2	NA	7	1	3	NA	14
Manganese	4	NA	8	NA	4	2	9	5	3	6	30
Mercy	4	NA	2	NA	3	NA	4	3	NA	3	12
Nickel	1	NA	3	NA	1	1	3	1	NA	2	9
Selenium	NA	NA	3	NA	1	3	7	2	2	NA	14
Zinc	3	NA	7	1	4	6	13	6	4	3	32
Metals*	NA	NA	NA	NA	NA	NA	NA	NA	NA	8	8
Total of studies	8	2	15	1	7	11	41	11	8	13	83

NA, not available

* Only reported general metal exposure, not specified certain kind of metal

Table 3. Quality assessment for included papers: exposure assessment method and general study design quality

EA score	Exposure assessment method	No. of studies	General study design quality		
			low	moderate	high
Case-control studies*					
1	Self-reported occupational/environmental exposure	6	1	3	2
2	Biomonitoring after disease onset	64	3	58	3
2	Food frequency questionnaire after disease onset	8	0	3	5
2	Environmental monitoring after disease onset	1	0	1	0
4	Biomonitoring, Pb in bone	2	0	1	1
4	Job exposure matrix	4	0	2	2
4	Expert assessment	2	0	1	1
	Total No. of studies		**4**	**66**	**13**
Cohort studies					
4	Food frequency questionnaire before disease onset	1	0	0	1
4	Job exposure matrix	2	0	2	0
4	Environmental monitoring before disease onset	2	0	1	1
	Total No. of studies		**0**	**3**	**2**

* Four studies adopted two methods of exposure assessment.

For many meta-analyses, between-study heterogeneity was considerable (Table 4). Studies assessing copper, iron, and zinc in plasma/serum had an I^2 of >90%. Subgroup analyses revealed a subtle change in effect sizes between geographic locations (Table S8). Significant differences were observed among the detection techniques for copper in CSF ($p_{subgroup}$=0.014), copper in plasma/serum ($p_{subgroup}$<0.001), iron in CSF ($p_{subgroup}$<0.001), iron in serum ($p_{subgroup}$=0.005), manganese in plasma/serum ($p_{subgroup}$=0.025), and zinc in serum ($p_{subgroup}$=0.034). Influential studies were detected in some meta-analyses, including those of iron in plasma/serum, selenium in plasma/serum, and zinc in serum (Table S9). Removal of these influential studies caused small deviations from both the original pooled effects and between-study heterogeneity.

Funnel plot and Egger's test did not reveal any significant evidence of publication bias, except for studies on copper in CSF (Egger's test, p=0.03) (Table S10, Figure S1). After trim-and-fill method adjustment, the pooled effect of -0.23 (-0.49, 0.02) in the meta-analysis of iron in plasma/serum changed to 0.02 (-0.27, 0.32).

Table 4. Pooled effect estimates for metal levels in biospecimens and Parkinson's disease

Metal	Biological matrix	No. of studies	No. of PD cases	No. of controls	Pooled SMD (95%CI)	I^2 (%)
Aluminum	CSF	4	219	140	-0.50 (-1.05, 0.04)	53
	hair	3	186	243	0.92 (-1.15, 3.00)	94
	serum	4	464	447	-0.44 (-2.53, 1.64)	97
Calcium	CSF	4	219	140	0.30 (-0.10, 0.71)	18
	hair	3	163	75	-0.58 (-1.27, 0.11)	10
	serum	5	497	546	0.80 (-0.69, 2.30)	99
Cadmium	blood	2	49	37	-0.61 (-1.08, -0.13)	0
	CSF	2	68	33	-1.20 (-12.21, 9.82)	92
	serum	2	97	137	-0.88 (-7.43, 5.68)	84
	urine	2	49	37	-0.04 (-4.21, 4.13)	53
Chromium	CSF	5	182	178	-0.40 (-1.58, 0.78)	92
	serum	6	440	586	0.10 (-0.14, 0.34)	33
	urine	3	79	64	-0.14 (-0.45, 0.17)	0
Copper	blood	2	114	42	0.42 (-3.76, 4.59)	64
	CSF	11	418	336	0.16 (-0.38, 0.70)	86
	hair	3	150	56	-0.03 (-0.80, 0.73)	14
	plasma	7	603	746	0.27 (-0.58, 1.12)	97
	serum	18	1,147	1,164	-0.43 (-0.84, -0.02)	94
	plasma+serum	25	1,750	1,910	-0.23 (-0.60, 0.14)	96
	urine	4	198	127	-0.11 (-1.21, 0.98)	84

(**Table 4** continued)

Iron	CSF	11	483	312	-0.29 (-0.71, 0.13)	81
	hair	4	176	89	-0.13 (-1.03, 0.77)	78
	plasma	5	525	601	0.02 (-0.86, 0.90)	95
	serum	27	2,060	2,380	-0.28 (-0.56, 0.00)	89
	plasma+serum	32	2,585	2,981	-0.23 (-0.49, 0.02)	92
	urine	4	223	152	0.27 (-1.34, 1.87)	88
Lead	blood	2	49	37	0.37 (-6.35, 7.09)	81
	CSF	4	154	133	-0.60 (-2.59, 1.40)	95
	serum	4	380	516	-0.13 (-1.48, 1.22)	91
	plasma+serum	5	530	691	0.09 (-1.01, 1.19)	94
Magnesium	CSF	5	239	155	0.66 (0.41, 0.91)	0
	hair	2	137	42	-0.35 (-1.32, 0.62)	0
	serum	6	572	580	0.45 (-0.19, 1.09)	82
Manganese	blood	3	209	139	0.02 (-0.83, 0.87)	66
	CSF	8	296	243	-0.15 (-0.64, 0.34)	76
	hair	4	199	257	2.70 (-3.84, 9.23)	99
	plasma	2	375	300	0.43 (-7.02, 7.88)	98
	serum	8	589	664	0.11 (-0.43, 0.66)	89
	plasma+serum	10	964	964	0.18 (-0.29, 0.65)	93
	urine	4	205	130	-0.61 (-1.33, 0.11)	64
Mercury	blood	4	182	286	-0.20 (-1.69, 1.30)	93
	CSF	2	68	33	-1.05 (-4.14, 2.04)	12
	hair	3	179	273	-0.20 (-1.85, 1.45)	90
	serum	4	195	301	-0.66 (-1.91, 0.59)	90
	urine	3	103	133	-0.62 (-4.55, 3.01)	91
Nickel	CSF	3	208	150	-0.81 (-1.80, 0.17)	46
	serum	3	130	236	0.25 (-0.92, 1.42)	75
	plasma+serum	4	355	361	0.75 (-1.09, 2.59)	98
Selenium	CSF	3	100	106	0.71 (-0.04, 1.46)	28
	plasma	3	285	356	0.16 (-1.15, 1.47)	76
	serum	7	254	309	0.16 (-0.88, 1.20)	94
	plasma+serum	10	539	665	0.16 (-0.52, 0.84)	92
	urine	2	52	54	0.04 (-0.43, 0.50)	0
Zinc	blood	2	114	42	0.40 (-7.49, 8.29)	90
	CSF	7	312	213	-0.06 (-0.85, 0.73)	83
	hair	4	176	89	0.52 (0.14, 0.90)	0
	plasma	5	551	522	-1.04 (-2.07, -0.01)	92
	serum	13	815	837	-0.33 (-0.75, 0.09)	85
	plasma+serum	18	1,366	1,359	-0.53 (-0.92, -0.14)	94
	urine	4	198	127	-0.01 (-0.33, 0.30)	0

SMD, standardized mean difference

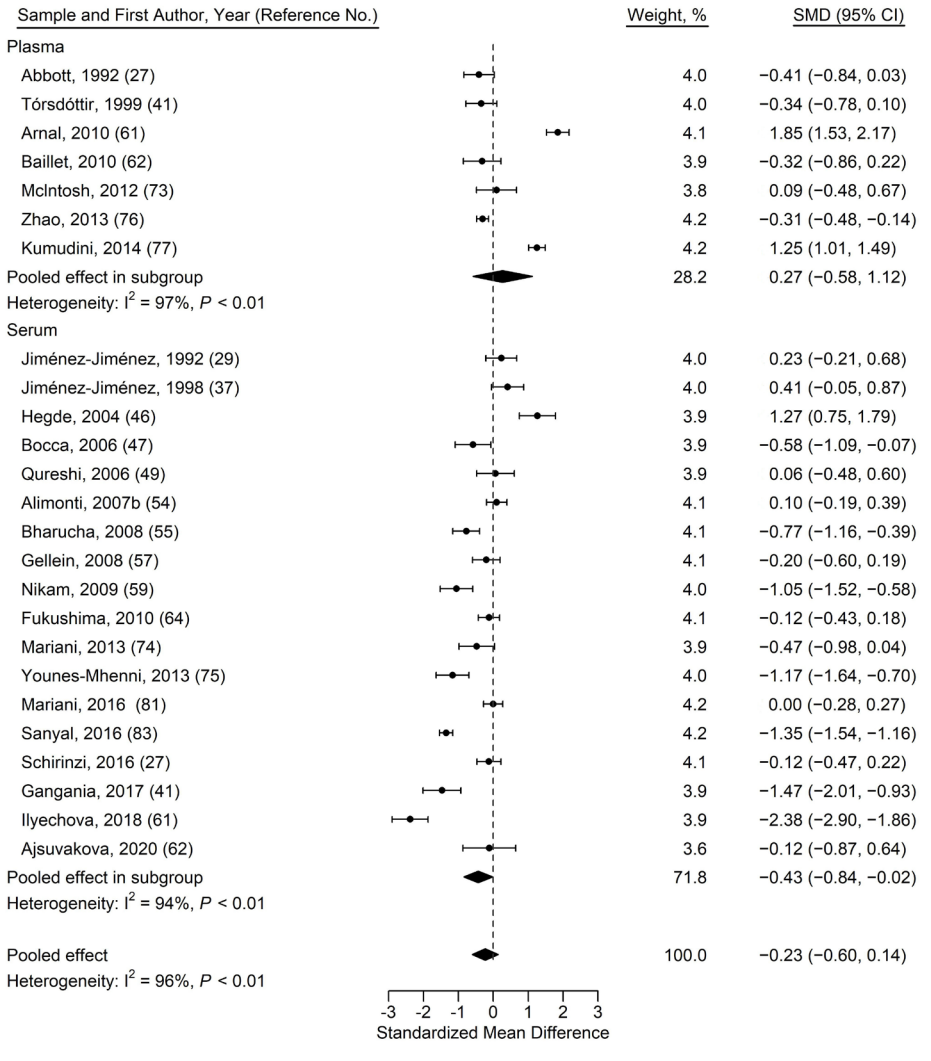


Figure 2. Forest plot of copper level in plasma and serum and Parkinson's disease

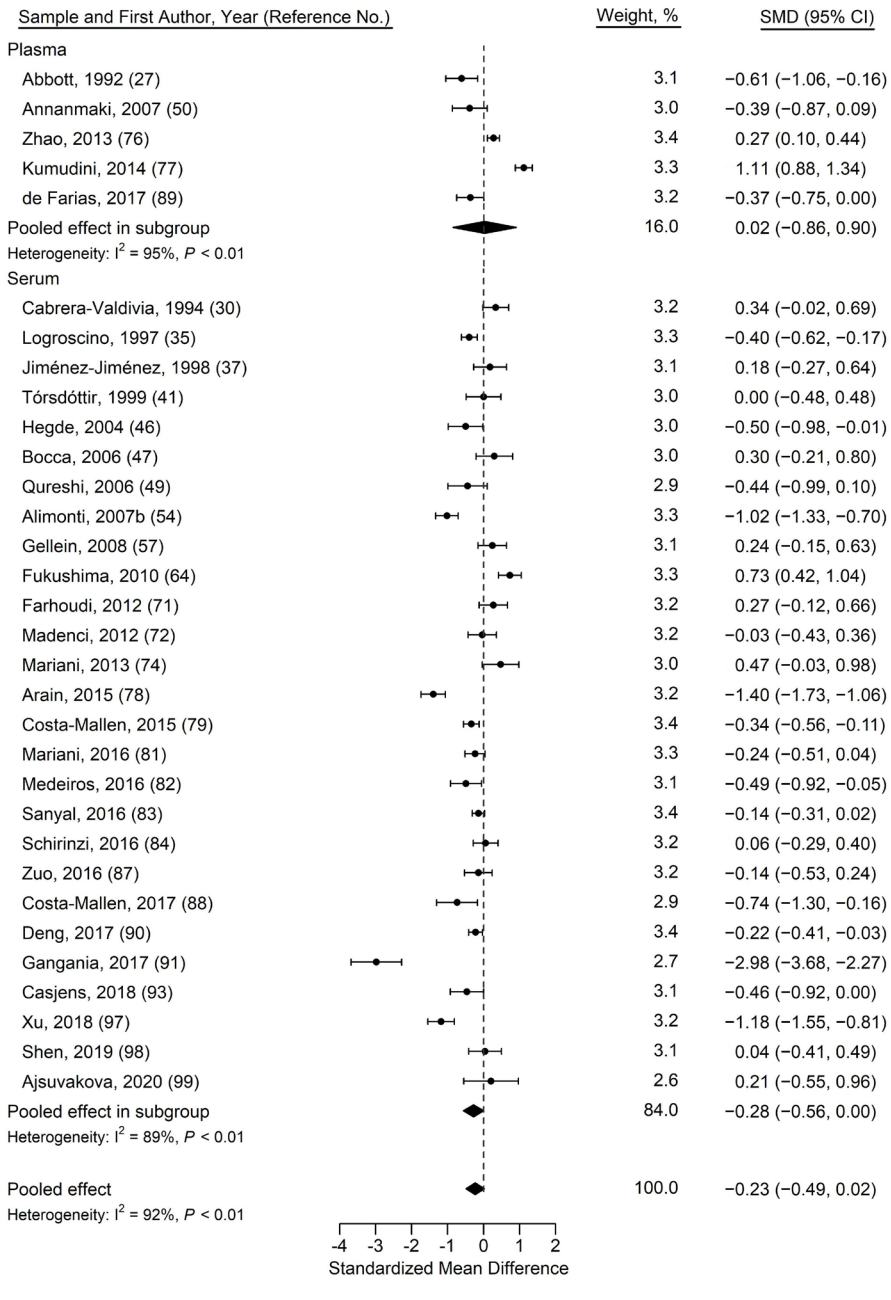


Figure 3. Forest plot of iron in plasma and serum and Parkinson's disease

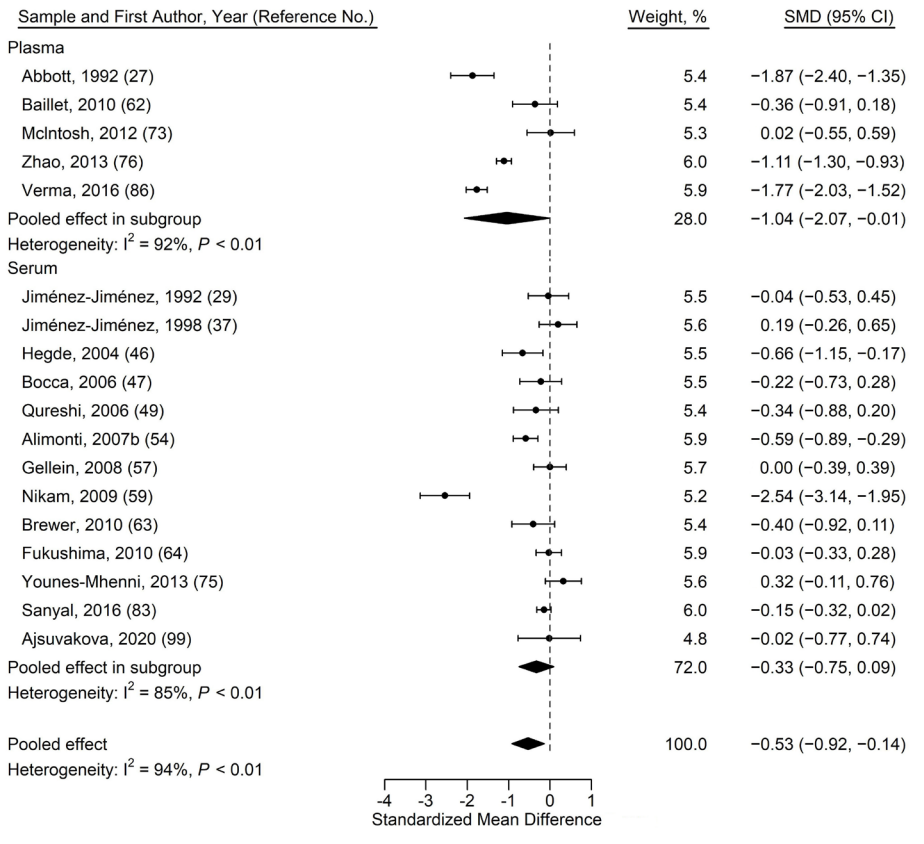


Figure 4. Forest plot of zinc in plasma and serum and Parkinson's disease

Meta-analyses of metal exposure from diet and occupation/environment

Case-control studies mainly focused on essential nutritional metals (calcium, copper, iron, magnesium, zinc) and did not show consistent results in meta-analyses (Table 5). An overall OR of 1.11 (95%CI 0.70, 1.76) was estimated for manganese, indicating no significant difference in dietary manganese intake between cases and controls. In a cohort study from Logroscino et al. (106), a modest increase in PD risk was associated with dietary iron intake (the highest vs. the lowest quintile, RR 1.30 (95%CI 0.94, 1.80)).

As for occupational/environmental metal exposure, a borderline significant OR from combing four studies (OR 1.04 (95%CI 1.01, 1.06)) was found for manganese exposure and PD risk (Table 6). Lead exposure was associated with an elevated risk (OR=1.14), but the effect was not statistically significant (95%CI 0.64, 2.01). The same for non-specified metal exposure (OR 1.22 (95%CI 0.70, 2.14)). The impacts of exposure to

copper, iron, mercury and zinc were inconclusive, and the confidence intervals for mercury and zinc were wide.

Feldman et al. (108) and Brouwer et al. (107) explored the association between occupational metal exposures and PD among men in large population-based prospective cohorts in Sweden and the Netherlands, respectively, but neither of them observed any significant association (Table 1). Palacios et al. (105) found a positive monotonic association with airborne mercury exposure and risk of PD (hazard ratio (95%CI) through quartiles, Q2 1.15 (0.87, 1.52), Q3 1.24 (0.93, 1.65), Q4 1.33 (0.99, 1.79)) in a cohort of female nurses, while relations with other hazardous metals (cadmium, chromium, lead, manganese, nickel) showed little evidence of differences. Vinceti et al. (109) found that high selenium levels in drinking water were associated with excess PD mortality, with an RR of 2.47 (95%CI 1.15, 5.29) compared with the control region.

Table 5. Pooled effect estimates for dietary metal intake and Parkinson's disease

Metal	No. of studies	No. of PD cases	No. of controls	Pooled OR (95%CI)	I^2 (%)
Calcium	4	826	1,151	1.03 (0.77, 1.39)	64
Copper	3	700	719	0.83 (0.30, 2.27)	85
Iron	6	1,140	1,704	0.99 (0.60, 1.61)	69
Magnesium	3	700	719	0.89 (0.22, 3.63)	89
Manganese	3	700	719	1.11 (0.70, 1.76)	0
Selenium	2	122	111	1.24 (0.44, 3.51)	0
Zinc	4	740	748	0.85 (0.42, 1.72)	83

Table 6. Pooled effect estimates for occupational/environmental metal exposures and Parkinson's disease

Metal	No. of studies	No. of PD cases	No. of controls	Pooled OR (95%CI)	I^2 (%)
Copper	3	1,163	2,779	1.11 (0.68, 1.80)	0
Iron	2	911	2,453	1.08 (0.91, 1.29)	0
Lead	4	1,351	1,571	1.14 (0.64, 2.01)	41
Manganese	4	1,547	25,893	1.04 (1.01, 1.06)	0
Mercury	2	524	840	1.02 (0.009, 111.70)	17
Zinc	3	947	1,045	1.56 (0.06, 44.07)	66
Metal	7	2,526	2,971	1.22 (0.70, 2.14)	65

Discussion

In this systematic review and meta-analysis, we assessed the current literature to summarize the evidence on the association between metal exposure and PD risk. Most case-control studies were biomonitoring studies and were of moderate-quality. Overall, there are no consistent associations regarding most metals in biospecimens or from dietary, occupational or environmental sources. Only for lead exposure was there an indication of a possible increased risk of PD, given the higher bone lead level among PD cases reported by two studies. Prospective studies assessing metal exposure prior to the outcome occurrence were limited and most did not find changes in risk of PD after metal exposure, except for the increased risk after exposure to airborne mercury and elevated PD mortality among residents drinking water with high selenium concentrations in one single study.

Trace metals are responsible for a wide variety of neuronal functions, and disturbances of metal homeostasis have been implicated in the progression of PD. In mechanistic studies, excessive levels of some metals (e.g., manganese, iron, lead, mercury, aluminum, cadmium) have been shown to induce injury in dopaminergic neurons (5, 110-113), which are the cells primarily affected in PD, while magnesium is expected to act as a neuroprotective agent by inhibiting N-methyl-D-aspartate (NMDA) receptors activity and oxidative stress (114). However, the role of other metals (e.g., zinc, copper, selenium) remains unclear and complicated as both beneficial and deleterious actions have been postulated in PD (115, 116).

To date, human studies on the relationship between metal exposures and the risk of PD face several limitations. The number of studies available for most metal-biospecimen combinations is less than five and based on small-scale research, often including less than 50 PD patients. Further, few studies on metal exposure from diet, occupation or the environment are available to date, although they included larger numbers of PD cases. Such data sparsity makes the pooled effects in this review less accurate because the standard random-effects meta-analysis method can lead to serious distortions in the presence of few studies and/or limited sample sizes (117). Additionally, consistent lower levels of iron and copper in serum were drawn from respectively 18 and 27 studies, but the result became ambiguous when adding a few studies assessing metals in plasma, making the inverse association of iron and copper in the combined matrices with PD undecisive. Another concern when utilizing biomonitoring studies is that circulating metal is not necessarily representative for long-term exposure due to rapid elimination in biological fluids. Pathogenesis and progression of PD is slow, thus chronic exposures to environmental stimuli will play a major role in the disease etiology.

Bone lead, an exception among biomonitoring, is a proxy of distant past exposure because of the decades half-life of lead in bone. Two large-scale case-control studies assessing bone lead (451 PD patients and 722 controls in total) consistently reported increased risk of PD related to cumulative exposure to lead. Further considering the relatively good quality of study design, these studies have indicated lead as a possible environmental risk factor for PD.

In our meta-analysis, PD patients had somewhat increased manganese blood levels in comparison with controls, but with wide confidence intervals and considerable heterogeneity across studies (I^2>90%). Studies assessing occupational/environmental exposure, however, indicated a possible association (OR 1.04 (95%CI 1.01, 1.06), I^2=0%). This limited evidence regarding manganese as risk factor for PD seemed contradictory to the well-established finding of manganese-induced parkinsonism. The reason behind the inconsistence might be the different pathogenesis. Unlike PD, manganese-induced parkinsonism does not involve degeneration of midbrain dopamine neurons and that levodopa is not an effective therapy (118). The contribution of manganese to these two movement disorders may therefore be different.

The overall lack of consistency among studies limits drawing firm conclusions on associations. The high level of between-study heterogeneity was confirmed among many evaluated studies, which indicates that effects might differ in certain contexts. From our subgroup analysis, metal detection methods in biomonitoring studies might have contributed to the high heterogeneity. Other relevant factors such as age distribution, gender ratio, disease severity and disease duration may also have resulted in heterogeneity, but no sufficient data were available to address their impact in our current analyses. What is more, the near null effect of iron in either serum or plasma after trim-and-fill correction indicates that the pooled effect in the meta-analysis might be overestimated due to small-study effects.

More importantly, methodological limitations in the available studies could result in serious bias and distort the association between metal exposure and the risk of PD. First, there is possible case selection bias, as some studies identified PD outcome through death certificates or health care registers (40, 107-109). Register-based case ascertainment is likely to omit patients with early or mild disease, leading to results only based on more severe cases, which may not be translatable to all PD cases. Overlapping clinical features with other neurodegenerations and secondary parkinsonism, as well as symptom-based diagnosis, might also obscure the association for PD, since disease aetiologies may be different. Second is the selection of controls, which is often based on patients from the same hospital. Hospital controls, however, may not be representative of the source population, whereas the use of relatives as controls (30,

65, 84) may be affected by overmatching due to shared living conditions, activities and life habits resulting in a similar exposure status. Third, self-reported information on exposure in case-control studies (43, 100) can be affected by the awareness of disease status, resulting in differential recall between cases and controls. Furthermore, PD manifestations may have changed the toxicokinetics of metals, and altered metal levels after diagnosis may erroneously be thought to have an aetiological role, so-called reverse causality. Fourth, almost half of the case-control studies did not adopt matching between case and control groups. Confounding introduced by age, gender, smoking status, alcohol consumption and comorbidities could bias effect estimates and adjustment should be considered.

Conclusion and future directions

To our knowledge, this is the first meta-analysis and systematic review to investigate the association between metal exposures from various routes and the risk of PD. Besides consistency of results, we also considered the impact of exposure assessment and study design, which recently have been recommended when applying pooled estimates to causal inference in observational studies (119). In conclusion, because of inadequate study quality, high heterogeneity of reported results and methodological limitations, the work in PD epidemiology is yet insufficient to establish an association between specific metal exposures and the disease risk. Future research on the association between metals and PD risk should aim to address the above challenges effectively to provide more reliable evidence. This further evidence will heavily rely on large prospective cohort studies, with comprehensive lifelong exposure history, sufficient follow-up period, well-established biobanks and careful case ascertainment.

Chapter 2 Supplemental materials

Table S1. Search queries

PubMed
(“Parkinson Disease”[MeSH Terms] OR “parkinson*”[All Fields] OR “pd”[All Fields] OR “neurodegenerative*”[All Fields]) AND (“Metals”[MeSH Terms] OR “Manganese”[MeSH Terms] OR “Iron”[MeSH Terms] OR “Copper”[MeSH Terms] OR “Lead”[MeSH Terms] OR “Mercury”[MeSH Terms] OR “Aluminum”[MeSH Terms] OR “Calcium”[MeSH Terms] OR “Selenium”[MeSH Terms] OR "Zinc"[MeSH Terms] OR "Magnesium"[MeSH Terms] OR "Cadmium"[MeSH Terms] OR "Chromium"[MeSH Terms] OR "Nickel"[MeSH Terms]) AND ("exposure*"[All Fields] OR "exposed*"[All Fields]) AND ((humans[Filter]) AND (english[Filter]))) NOT ((((“Meta-Analysis" [Publication Type])) OR ("Review" [Publication Type])) OR ("Systematic Review" [Publication Type]))
EMBASE
('parkinson disease'/exp OR 'parkinson disease' OR 'pd'/exp OR 'pd' OR 'parkinson*' OR 'neurodegenerative*') AND ('metal*' OR 'manganese*' OR 'iron*' OR 'copper*' OR 'lead*' OR 'mercury*' OR 'aluminium*' OR ‘aluminum*’ OR 'calcium*' OR 'selenium*' OR 'zinc*' OR 'magnesium*' OR ‘cadmium*’ OR ‘chromium*’ OR ‘nickel*’) AND ('exposure*' OR 'exposed*') NOT ([cochrane review]/lim OR [systematic review]/lim OR [meta analysis]/lim) NOT [review]/lim AND [english]/lim AND [humans]/lim AND [embase]/lim
Cochrane
(parkinson disease OR pd OR parkinson* OR neurodegenerative*) AND (metal* OR manganese* OR iron* OR copper* OR lead* OR mercury* OR aluminium* OR aluminum* OR calcium* OR selenium* OR zinc* OR magnesium* OR cadmium* OR chromium* OR nickel*) AND (exposure* OR exposed*)

Table S2. Newcastle-Ottawa scale for case-control studies

Assessment items	Points
Subject selection	
1. Adequate case definition	
a. Independent validation by neurologists or confirmation of PD diagnosis by reference to secure records (e.g., hospital record)	2
NB: registers of confirmed PD included	
b. Record linkage (e.g., insurance registry or other health registry)	1
c. Self-report	0
d. No description	0
2. Representativeness of cases	
a. Cases from multiple communities/hospitals/clinics	2
b. Cases from single community/hospital/clinic	1
c. Potential for selection bias or not stated	0
3. Selection of cases	
a. Incident cases only	1
b. Including prevalent cases	0
c. No description	0
4. Selection of controls	
a. Community controls	2
b. Hospital controls	1
c. Family/relative/partner controls or no description	0
5. Definition of controls	
a. No history of PD or confounding condition	1
b. No description	0
Comparability	
6. Comparability of cases and controls on the basis of the design or analysis	
a. Besides age and sex, other potential confounders are considered (e.g., ethnicity, tobacco and alcohol consumption)	2
b. Controls only matched with age and sex	1
c. Not controlled or not stated	0
Exposure assessment	
7. Ascertainment of exposure	
a. Blinded to case/control status	1
b. Unblinded	0
c. No description	0
8. Participation rate	
a. All eligible subjects participated or the same rate for both groups	2
b. Rate different and non-participants described	1
c. No description	0

Statistical analysis	
9. Statistical analysis a. Detailed and appropriate analysis: If continuous (in biomonitoring and some dietary exposure studies), parameters for center and variability (e.g., mean and standard deviation, median and interquartile range) of metal levels and the significance of differences between case and control groups (calculated by unpaired t-test, one-way analysis of variance or other reasonable statistics) were provided. If categorical (in occupational or environmental exposure or some dietary exposure studies), counts/percentage of case/control subjects in each group and crude/adjusted odds ratio and 95% confidential interval (calculated by logistic regression or other reasonable statistics) were provided.	2
b. Statistical test and results stated but limited	1
c. Inappropriate analysis or no description	0
Total maximum score	15

Low quality: 0-5; moderate quality: 6-10; high quality: 11-15.

Table S3. Newcastle-Ottawa scale for cohort studies

Assessment items	Points
Selection	
1. Representativeness of the exposed cohort	
a. Truly representative of the exposed population in the community/industry	2
b. Somewhat representative of the exposed population in the community/industry	1
c. Selected group of the exposed population *NB. state which group*	0
d. No description	0
2. Representativeness of the nonexposed cohort	
a. From the same community/industry as the exposed cohort	1
b. From a different source	0
c. No description	0
3. Demonstration that PD was not present at the start of the study	
a. Yes	1
b. No	0
Comparability	
4. Comparability of cohorts on the basis of the design or analysis	
a. Besides age and sex, other potential confounders are considered, (e.g., ethnicity, tobacco and alcohol consumption)	2
b. Controls only matched with age and sex	1
c. Not controlled or not stated	0
Outcome	
5. Assessment of outcome	
a. Independent validation by neurologists or confirmation of PD diagnosis by reference to secure records (e.g., hospital records) *NB: registers of confirmed PD included*	2
b. Record linkage (e.g., insurance registry or other health registry)	1
c. Self-report	0
d. No description	0
6. Follow up long enough for outcomes to occur	
a. More than 10 years	1
b. Less than 10 years	0
7. Adequacy of follow up cohorts	
a. Complete follow-up, all subjects accounted for	2
b. Subjects lost to follow-up are unlikely to introduce bias, small number lost < 20%	1
c. Follow up rate < 80%	0
d. No description	0

Analysis	
8. Statistical analysis	
a. Detailed and appropriate analysis	
Reporting of: the number of participants and exposure status at baseline, follow-up period, the number of participants with or without PD at the end of the study, relative risk/hazard ratio with 95% confidential interval are reported.	2
b. Statistical test and results stated but limited	1
c. Inappropriate analysis or no description	0
Total maximum score	13

Low quality: 0-4; moderate quality: 5-9; high quality: 10-13.

Table S4. Exposure assessment (EA) rating score for exposure assessment method*

EA score	Exposure assessment method	Design	Interpretation
1	Self-reported exposure Registry job history	(Hospital-based) case-control Industrial cohort	Uninformative
	Self-reported job history	Industrial cohort	
2	Self-reported job history Self-reported job history and task	(Hospital-based) case-control Community-based cohort	Findings not completely valid
	Environmental monitoring after disease onset Biomonitoring after disease onset# Food frequency questionnaire (FFQ) after disease onset	(Hospital-based) case-control Community-based cohort Industrial cohort Nested case-control	
3	Company job history	Industrial cohort Nested case-control	Findings valid, but not agent-specific
4	Job-exposure matrix (JEM) Case-by-case assessment by expert(s) Environmental monitoring before disease onset Biomonitoring before disease onset FFQ before disease onset	(Hospital-based) case-control Community-based cohort Industrial cohort Nested case-control	Findings are valid and agent-specific

* Adapted from Sutedja et al. (9)

Except for lead, cadmium and chromium in bone (e.g., metals that accumulate in the body, with slow elimination), then biomonitoring after disease onset is given a score of 4.

Formula appendix: Formulas for effect measure calculation

(1) Between-group standardized mean difference (SMD):

$$SMD = \frac{\bar{x}_{case} - \bar{x}_{ctr}}{s_{pooled}}$$

$$s_{pooled} = \sqrt{\frac{(n_{case}-1)SD_{case}^2 + (n_{ctr}-1)SD_{ctr}^2}{(n_{case}-1)+(n_{ctr}-1)}}$$

Standard error of SMD:

$$SE_{SMD} = \sqrt{\frac{n_{case}+n_{ctr}}{n_{case}n_{ctr}} + \frac{SMD^2}{2(n_{case}+n_{ctr})}}$$

Small-sample correction for SMD, Hedges's g:

$$g = SMD \times (1 - \frac{3}{4n-9})$$

$\bar{x}_{case}$, SD_{case}, n_{case}: mean, standard deviation and sample size of case group
$\bar{x}_{ctr}$, SD_{ctr}, n_{ctr}: mean, standard deviation and sample size of control group
n, total sample size of the study

(2) If the median and interquartile range (IQR) were reported, mean and standard deviation (SD) were estimated according to Wan et al. (120) as follows:

$$\bar{x} \approx \frac{q_1 + m + q_3}{3}$$

$$SD \approx \frac{q_3 - q_1}{2\Phi^{-1}(\frac{0.75n - 0.125}{n + 0.25})}$$

$\bar{x}$, mean; SD, standard deviation
q_1, the first quartile; m, median; q_3, the third quartile; n, sample size

(3) If the range was reported, SD was estimated using a tabulated conversion factor (*f*) (121):

$$SD \approx f \times range$$

(4) If 95% confidence interval (CI) or standard error (SE) were reported, SD was estimated as follows:

$$SD = \sqrt{n} \times (upperlimit - lowerlimit)/(t_{0.05,n-1} \times 2)$$

$$SD = SE \times \sqrt{n}$$

(5) If geometric mean and 95% CI were reported, mean and SD were estimated according to Higgins et al (11) as follows:

$$\bar{z} = lng$$

$$s_z = \frac{\bar{z} - lng_l}{t_{0.05,n-1}} \times \sqrt{n}$$

$$\bar{x} \approx \exp\,(\bar{z} + \frac{{s_z}^2}{2})$$

$$SD \approx \sqrt{(\exp {s_z}^2 - 1) \times \exp\,(2 \times \bar{z} + {s_z}^2)}$$

g, geometric mean; g_l, lower limit of 95% CI for g
$\bar{z}$, s_z: mean and standard deviation of log-transformed measurements

(6) If continuous exposure was reported in two subgroups (merely sub-samples of a whole group), the below formula can be used to combine into the single sample size, mean and SD for the whole group.

$$n = n_1 + n_2$$

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$$

$$s = \sqrt{\frac{(n_1 - 1){s_1}^2 + (n_2 - 1){s_2}^2 + \frac{n_1 n_2}{n_1 + n_2}({\bar{x}_1}^2 + {\bar{x}_2}^2 - 2\bar{x}_1\bar{x}_2)}{n_1 + n_2}}$$

$\bar{x}$, s, n: mean, SD and sample size for the whole group

$\bar{x}_1$, s_1, n_1: mean, SD and sample size for the subgroup 1

$\bar{x}_2$, s_2, n_2: mean, SD and sample size for the subgroup 2

When more than two groups are to be combined, the first two groups are combined first, the results are then combined with the third group, then sequentially with each subsequent group

(7) If studies assessing categorical exposure only reported observed numbers of cases/controls in each exposure group and no odds ratio (OR), OR and associated SE of logOR were calculated as below.

$$OR = \frac{a/b}{c/d}$$

$$logOR = lnOR$$

$$SE_{logOR} = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$$

a, b: the number of exposed and non-exposed subjects in case group

c, d: the number of exposed and non-exposed subjects in control group

(8) Conversion between standardized mean difference (SMD) and OR:

$$SMD = \frac{\sqrt{3}}{\pi}\ln OR$$

$$SE_{SMD} = \frac{\sqrt{3}}{\pi}SE_{lnOR}$$

(9) Conversion to mean and SD on log-transformed scale from these of raw scale (11):

$$\bar{z} = ln\bar{x} - \frac{1}{2}\ln\left(\frac{s^2}{\bar{x}^2} + 1\right)$$

$$s_Z = \sqrt{\ln\left(\frac{s^2}{\bar{x}^2} + 1\right)}$$

$\bar{x}$, s: mean and SD of raw-scaled measurements

$\bar{z}$, s_z: mean and standard deviation of log-transformed measurements

Table S5. Overview of case-control studies on metal exposures and Parkinson's disease

Study	Country	Subjects source[1]		Sample size		Age (years)		Gender (%male)		Biospecimen or exposure source[2]	Metals[3]
		PD	HC	PD	HC	PD	HC	PD	HC		
Barbeau 1963(22)	Canada	NR	H	50	76	49.2[a]	46[a]	NR	NR	serum	Mg
Schwab 1964(23)	USA	H	C	75	34	NR	NR	NR	NR	serum	Mg
Pall 1987(24)	UK	NR	H	31	38	54.6±11.2	49.5±11.9	64.5	44.7	CSF	Cu, Fe, Mn
Ngim 1989(25)	Singapore	H	H	54	95	64.4±9.6	64.7±9	66.7	63.2	blood, hair, urine	Hg
Wechsler 1991(26)	USA	H, P	H	34	22	68.4[a]	58.9[a]	55.9	40.9	occupation	Al, Cd, Cr, Cu, Pb, Mn, Hg, Ni
Abbott 1992(27)	UK	H	H	45	41	62±9	58±9	51.1	51.2	plasma	Fe, Cu, Zn
Gazzaniga 1992(28)	Italy	NR	NR	11	22	64.9 (49-78)[b]	NR	90.9	90.9	CSF	Cu, Fe, Mn
Jiménez-Jiménez 1992(29)	Spain	H	S	39	39	67.3±1.34[c]	66.2±1.52[c]	51.3	48.7	serum	Zn, Cu
Cabrera-Valdivia 1994(30)	Spain	H	S	68	68	65.8±0.96[c]	65.8±1.00[c]	52.9	47.1	serum, urine	Fe
Jiménez-Jiménez 1995a(31)	Spain	NR	NR	32	27	66.7±1.5[c]	67.1±1.9[c]	46.9	44.4	serum, urine	Mn, Cr
Jiménez-Jiménez 1995b(32)	Spain	NR	NR	29	30	66.3±1.62[c]	67±1.75[c]	51.7	50.0	serum, urine	Se
Seidler 1996(33)	Germany	H	C	380	376	56.2±6.6	56.2±7.0	66.1	NR	occupation	Pb, Hg, Zn, Cu, Cd, Ni
Gorell 1997(34)	USA	M	C	144	464	NR	NR	62.5	63.1	occupation	Cu, Fe, Zn, Hg, Mn, Pb
Logroscino 1997(35)	Italy	C	M	104	352	NR	NR	NR	NR	serum, diet	Fe
Aguilar 1998(36)	Spain	H	H	28	43	65.5±9.1	65.2±13	39.3	44.2	CSF, serum	Se, Cr
Jiménez-Jiménez 1998(37)	Spain	H	H	37	37	65.7±8.8	62.4±17.8	37.8	43.2	CSF, serum	Fe, Cu, Zn, Mn
Smargiassi 1998(38)	Italy	H	H	86	86	66.4±9.7	63.1±8.7	58.1	55.8	occupation	metal
Boll 1999(39)	Mexico	H	H	49	26	NR	58.7±2.37[c]	65.3	46.2	CSF	Cu
Johnson 1999(40)	USA	M	C	126	432	NR	NR	61.9	62.7	diet	Fe, Ca
Tórsdóttir 1999(41)	Iceland	NR	NR	40	40	67 (47-82)[b]	NR	55.0	NR	plasma, serum	Cu, Fe
Kocatürk 2000(42)	Turkey	NR	NR	30	24	64[a]	61[a]	NR	NR	plasma, erythrocyte	Cu, Zn
Pals 2003(43)	Belgium	P, C	S	423	205	NR	NR	60.5	34.1	environment	metal, Zn, Pb, Cu
Powers 2003(44)	USA	H	H	250	388	70 (37-88)[d]	71 (38-65)[d]	62.4	62.1	diet	Zn, Mn, Mg, Cu, Ca
Bocca 2004(45)	Italy	NR	NR	26	13	64.9±10.8	63.8±13.7	92.3	46.2	urine, serum, blood, CSF	Cd, Cr, Hg, Ni, Pb
Hegde 2004(46)	India	H	NR	52	25	NR	55.4±6.4	53.8	52.0	serum	Ca, Mg, Cu, Zn, Fe, Al, Mn
Frigerio 2005(103)	USA	M	C	196	196	71 (41-97)[d]	NR	61.7	NR	occupation	metal
Bocca 2006(47)	Italy	NR	NR	91	18	65.5±9.7	63.3±13.8	70.3	55.6	CSF, blood, serum, urine, hair	Al, Ca, Cu, Fe, Mg, Mn, Zn

Coon 2006(48)	USA	M	C	121	414	72.3[a]	69.7[a]	62.8	54.8	bone (calcaneus and tibia)	Pb
Qureshi 2006(49)	Sweden	H	NR	36	21	NR	62±11	63.9	61.9	serum, CSF	Fe, Cu, Zn, Se
Alimonti 2007a(53)	Italy	NR	C	42	20	64.5±10.7	66.2±14.7	85.7	85.0	CSF	Al, Ca, Cd, Cr, Cu, Fe, Hg, Mg, Mn, Ni, Pb, Zn
Alimonti 2007b(54)	Italy	H	C	71	124	65.5±9.4	44.8±12.7	74.6	65.3	serum	Al, Ca, Cd, Cr, Cu, Fe, Hg, Mg, Mn, Ni, Pb, Zn
Annanmaki 2007(50)	Finland	H	S	40	29	60.8±6.5	60.2±5.1	57.5	44.8	plasma, diet	Fe
Dick 2007(51)	Scotland, Sweden, Italy, Romania, Malta	H	H, C	767	1989	69.8±9.2	69.8±10	55.5	53.1	occupation	Fe, Mn, Cu
Finkelstein 2007(52)	Canada	H	H	232	22914	NR	NR	46.1	NR	environment	Mn
Bharucha 2008(55)	USA	H	NR	91	40	NR	49±9.9	75.8	65.0	serum	Cu
Boll 2008(56)	Mexico	H	H	22	41	NR	NR	NR	NR	CSF	Cu
Gellein 2008(57)	Norway	H	NR	33	99	NR	NR	48.5	48.5	serum	Ca, Cr, Cu, Fe, Hg, Mg, Mn, Ni, Pb, Se, Zn
Petersen 2008(58)	Denmark	C, H	C	79	154	74.4±9.5	75.2±9.6	54.4	55.2	blood, hair, occupation	Hg
Nikam 2009(59)	India	H	NR	40	40	NR	NR	NR	NR	serum	Cu, Zn, Se
Powers 2009(60)	USA	H	H	420	560	NR	71 (38-87)[d]	63.3	62.7	diet	Fe
Arnal 2010(61)	Argentina	C	N	87	79	70±4.6	77.8±3.7	57.5	48.1	plasma	Cu
Baillet 2010(62)	France	H	NR	24	30	57.8±8.5	39.4±11.3	70.8	36.7	plasma	Se, Cu, Zn
Brewer 2010(63)	USA	H	H	30	29	67.4 (52-85)[b]	68.6 (56-80)[b]	60.0	31.0	serum, urine	Zn
Firestone 2010(104)	USA	M	C	404	526	69 (29-88)[d]	71 (38-86)[d]	62.4	62.0	occupation	metal, Cu, Mn, Pb
Fukushima 2010(64)	China	H	H	82	82	63.95±9.4	63.65±9.35	57.3	57.3	serum, blood, diet	Cu, Fe, Zn, Mn
Weisskopf 2010(65)	USA	H, C	S, F, C	330	308	66.5±9.5	69.4±8.8	65.5	55.8	bone (tibia and patella)	Pb
Fukushima 2011(66)	China	H	H	71	71	63.7±9.7	63.4±9.7	57.7	57.7	urine	Fe, Cu, Zn, Mn
Hozumi 2011(67)	Japan	NR	NR	20	15	68.7±5.8	48.4±22.2	45.0	40.0	CSF	Cu, Fe, Mg, Mn, Zn
Ling 2011(68)	Thailand	H	C	41	26	NR	50.9±3.8	NR	53.8	serum	Fe, Cu
Miyake 2011a(69)	Japan	H	H	249	368	68.5±8.6	66.6±8.5	37.3	38.3	diet	Ca
Miyake 2011b(70)	Japan	H	H	249	368	68.5±8.6	66.6±8.5	37.3	38.3	diet	Fe, Mg, Zn, Cu, Mn
Farhoudi 2012(71)	Iran	H	NR	50	50	64.53±10.18	63.53±9.78	56.0	50.0	serum	Fe
Madenci 2012(72)	Turkey	H	H	60	42	68.5±9.2	66.9±8.3	55.0	52.4	serum	Fe
McIntosh 2012(73)	USA	H	H	23	24	70±9	73±6	47.8	41.7	plasma, blood, urine	Cd, Cr, Cu, Hg, Pb, Se, Zn
Mariani 2013(74)	Italy	H	H	22	49	73±9.7	69.2±10.1	63.6	38.8	serum	Fe, Cu
Younes-Mhenni 2013(75)	Tunisia	H	H	48	36	65.8±10.2	59.7±12.1	54.2	38.9	serum	Se, Cu, Zn
Zhao 2013(76)	China	NR	NR	238	302	66.6±11.3	65.6±12.2	50.8	50.7	plasma	Se, Cu, Fe, Zn
Kumudini 2014(77)	India	H	H	150	175	55.7±10.6	53.73±10.9	71.3	68.6	plasma	Cu, Fe, Mn, Pb
Arain 2015(78)	Pakistan	H	R	49	201	NR	NR	57.1	NR	hair, serum	Mn, Al, Fe

Costa-Mallen 2015(79)	USA	H, C	H	128	226	69.01[a]	62.57[a]	68.8	46.0	serum	Fe
van der Mark 2015(80)	Netherlands	H	H	444	876	68 (34-91)[d]	68 (34-90)[d]	63.3	63.6	occupation	metal
Mariani 2016(81)	Italy	H	NR	92	112	70 (38-93)[d]	62 (31-87)[d]	67.4	35.7	serum	Cu, Fe
Medeiros 2016(82)	Brazil	H	C	40	46	65.95±12.3	62.3±10.17	45.0	41.3	serum	Fe
Sanyal 2016(83)	India	H	H	250	280	57.88±12.059	56.42±9.68	64.8	65.4	serum, CSF	Al, Ca, Cu, Cr, Fe, Pb, Mg, Mn, Zn
Schirinzi 2016(84)	Italy	H	R, F	300	300	70.6±10.4	69.4±9.4	51.0	50.0	serum	Fe, Cu, Hg, Mn
Stefano 2016(85)	Italy	NR	R	46	24	72.33±1.25[c]	68.25±1.83[c]	65.2	33.3	hair	Mg, Al, Ca, Cr, Mn, Fe, Ni, Cu, Zn, Se, Cd, Hg, Pb
Verma 2016(86)	India	NR	NR	225	125	56.84±8.82	57.24±7.97	72.9	72.0	plasma	Mn, Zn, Ni
Zuo 2016(87)	China	H	H	518	29	NR	NR	52.1	NR	CSF, serum	Fe
Costa-Mallen 2017(88)	USA	H, C	C	24	27	63.6±9	64±9.2	45.8	51.9	serum	Fe
de Farias 2017(89)	Brazil	H	NR	56	54	70.34±8.86	69.7±8.76	48.2	44.4	plasma	Fe
Deng 2017(90)	China	H	H	213	219	63.4±10.52	63.12±6.01	53.5	54.3	serum	Fe
Gangania 2017(91)	India	NR	C	35	33	57±12	50[a]	68.6	NR	serum	Fe, Cu
Hemmati-Dinarvand 2017(92)	Iran	H	NR	40	40	65.7±6.32	64.35±3.75	67.5	55.0	serum	Se
Casjens 2018(93)	USA	H	C	36	39	60 (55-67)[d]	54 (50-64)[d]	NR	NR	serum, blood	Fe, Mn
Dos Santos 2018(94)	Brazil	H	C	26	33	69.8±14.7	NR	69.2	NR	hair	Ca, Fe, Zn
Ilyechova 2018(95)	Russia	H	NR	50	50	64 (57-72)[d]	60 (55-74)[d]	42.0	48.0	serum	Cu
Maass 2018(96)	Germany	H	NR	36	42	67±11	65.5±13.1	66.7	57.1	CSF	Al, Ca, Cr, Cu, Fe, Mg, Mn, Ni, Pb, Se, Zn
Xu 2018(97)	China	H	NR	70	64	60.91±10.93	60.52±6.5	41.4	46.9	serum	Fe
Shen 2019(98)	China	H	H	43	33	69.5±5.9	67.6±4.3	51.2	42.4	serum, CSF	Fe
Ajsuvakova 2020(99)	Russia	H	NR	13	14	73.6[a]	NR	30.8	NR	hair, urine, serum	Cu, Fe, Mn, Zn
Belvisi 2020(100)	Italy	H	R	694	640	67.9±9	64.7±10.3	58.6	54.8	occupation	metal
Kissani 2020(101)	Morocco	H	C	30	30	59[a]	34[a]	73.3	63.3	blood, urine	Mn, Cu, Zn, Pd
Lucchini 2020(102)	Italy	H	H	334	444	71.5±9.9	70±9.8	62.3	60.8	occupation	metal

NR, not reported

[1] H, hospital; C, community; P, patient support group; S, spouses; M, medical/health system; F, friends; R, relatives; N, nurses

[2] CSF, cerebrospinal fluid

[3] Al, aluminum; Ca, calcium; Cd, cadmium; Cr, chromium; Cu, copper; Fe, iron; Pb, lead; Mg, magnesium; Mn, manganese; Hg, mercury; Ni, nickel; Se, selenium; Zn, zinc

[a] mean

[b] mean (range)

[c] mean ± standard error of mean (SEM)

[d] median (range)

Table S6. Quality assessment of case-control studies

Study	Item 1	Item 2	Item3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Total	Grade	EA score	EA method[1] *
Barbeau 1963(22)	0	0	0	1	0	0	1	0	0	2	low	2	biomonitoring
Schwab 1964(23)	2	2	0	0	2	0	1	0	0	7	moderate	2	biomonitoring
Pall 1987(24)	2	0	1	1	0	1	1	0	1	7	moderate	2	biomonitoring
Ngim 1989(25)	2	2	1	1	1	2	1	0	2	12	high	2	biomonitoring
Wechsler 1991(26)	0	2	0	1	1	0	0	0	1	5	low	1	self-report
Abbott 1992(27)	2	1	0	1	0	1	1	0	2	8	moderate	2	biomonitoring
Gazzaniga 1992(28)	2	0	0	0	0	1	1	0	2	6	moderate	2	biomonitoring
Jiménez-Jiménez 1992(29)	2	2	0	0	0	0	1	0	2	7	moderate	2	biomonitoring
Cabrera-Valdivia 1994(30)	2	2	0	0	0	0	1	0	2	7	moderate	2	biomonitoring
Jiménez-Jiménez 1995a(31)	2	0	0	2	0	1	1	0	2	8	moderate	2	biomonitoring
Jiménez-Jiménez 1995b(32)	2	0	0	2	0	1	1	0	2	8	moderate	2	biomonitoring
Seidler 1996(33)	2	2	0	2	0	2	1	0	1	10	moderate	4	JEM
Gorell 1997(34)	2	2	1	2	1	2	1	2	2	15	high	4	expert assessment
Logroscino 1997(35)	2	0	0	0	1	2	1	0	2	8	moderate	2	biomonitoring; FFQ
Aguilar 1998(36)	2	2	0	1	0	1	1	0	2	9	moderate	2	biomonitoring
Jiménez-Jiménez 1998(37)	2	2	0	1	0	1	1	0	2	9	moderate	2	biomonitoring
Smargiassi 1998(38)	2	1	0	1	0	1	1	2	2	10	moderate	4	expert assessment
Boll 1999(39)	2	1	0	1	1	1	1	0	1	8	moderate	2	biomonitoring
Johnson 1999(40)	2	2	1	2	1	2	1	2	2	15	high	2	FFQ
Tórsdóttir 1999(41)	2	0	0	0	1	1	1	0	2	7	moderate	2	biomonitoring
Kocatürk 2000(42)	0	0	0	0	1	0	1	0	2	4	low	2	biomonitoring
Pals 2003(43)	2	2	0	0	1	1	0	0	2	8	moderate	1	self-report
Powers 2003(44)	2	2	1	2	1	2	1	2	2	15	high	2	FFQ
Bocca 2004(45)	2	0	0	0	1	0	1	0	2	6	moderate	2	biomonitoring
Hegde 2004(46)	2	2	0	0	1	1	1	0	1	8	moderate	2	biomonitoring

Frigerio 2005(103)	2	2	0	2	1	1	1	2	2	13	high	4	JEM
Bocca 2006(47)	2	0	0	0	1	0	1	0	2	6	moderate	2	biomonitoring
Coon 2006(48)	2	2	0	2	1	2	1	0	1	11	high	4	bone lead
Qureshi 2006(49)	2	1	0	0	1	0	1	0	1	6	moderate	2	biomonitoring
Alimonti 2007a(53)	2	0	0	2	1	1	1	0	2	9	moderate	2	biomonitoring
Alimonti 2007b(54)	2	1	0	2	0	0	1	0	2	8	moderate	2	biomonitoring
Annanmaki 2007(50)	2	1	0	0	1	0	1	0	2	7	moderate	2	biomonitoring; FFQ
Dick 2007(51)	2	2	0	1	0	2	1	0	1	9	moderate	4	JEM
Finkelstein 2007(52)	2	2	0	1	1	1	1	0	2	10	moderate	2	environmental monitoring
Bharucha 2008(55)	2	1	0	0	1	0	1	0	1	6	moderate	2	biomonitoring
Boll 2008(56)	2	1	0	1	1	0	1	0	1	7	moderate	2	biomonitoring
Gellein 2008(57)	2	2	1	0	1	1	1	0	2	10	moderate	2	biomonitoring
Petersen 2008(58)	2	2	0	2	0	1	1	1	2	11	high	2 or 1	biomonitoring; self-report
Nikam 2009(59)	2	1	0	0	0	1	1	0	2	7	moderate	2	biomonitoring
Powers 2009(60)	2	2	1	2	1	2	1	0	2	13	high	2	FFQ
Arnal 2010(61)	2	1	0	0	1	0	1	0	1	6	moderate	2	biomonitoring
Baillet 2010(62)	2	1	0	0	1	0	1	0	2	7	moderate	2	biomonitoring
Brewer 2010(63)	2	1	0	1	0	0	1	0	1	6	moderate	2	biomonitoring
Firestone 2010(104)	2	2	1	2	1	2	1	0	2	13	high	1	self-report
Fukushima 2010(64)	2	1	0	1	1	2	1	0	2	10	moderate	2	biomonitoring; FFQ
Weisskopf 2010(65)	2	2	0	0	0	2	1	0	2	9	moderate	4	bone lead
Fukushima 2011(66)	2	1	0	1	1	2	1	0	2	10	moderate	2	biomonitoring
Hozumi 2011(67)	2	0	0	0	1	0	1	0	2	6	moderate	2	biomonitoring
Ling 2011(68)	2	1	0	2	0	1	1	0	1	8	moderate	2	biomonitoring
Miyake 2011a(69)	2	2	0	1	1	2	1	0	2	11	high	2	FFQ
Miyake 2011b(70)	2	2	0	1	1	2	1	0	2	11	high	2	FFQ
Farhoudi 2012(71)	2	1	0	0	0	2	1	0	2	8	moderate	2	biomonitoring
Madenci 2012(72)	2	1	0	1	0	1	1	0	2	8	moderate	2	biomonitoring
McIntosh 2012(73)	2	1	0	1	1	1	1	0	2	9	moderate	2	biomonitoring

Mariani 2013(74)	2	1	0	1	0	1	1	0	2	8	moderate	2	biomonitoring
Younes-Mhenni 2013(75)	2	1	0	1	0	0	1	0	2	7	moderate	2	biomonitoring
Zhao 2013(76)	2	0	0	0	1	1	1	0	2	7	moderate	2	biomonitoring
Kumudini 2014(77)	2	1	0	1	1	2	1	0	2	10	moderate	2	biomonitoring
Arain 2015(78)	2	1	0	0	1	0	1	0	1	6	moderate	2	biomonitoring
Costa-Mallen 2015(79)	2	2	0	1	1	0	1	0	2	9	moderate	2	biomonitoring
van der Mark 2015(80)	2	2	1	1	1	2	1	1	2	13	high	4	JEM
Mariani 2016(81)	2	2	0	0	1	0	1	0	2	8	moderate	2	biomonitoring
Medeiros 2016(82)	2	1	0	2	0	2	1	0	2	10	moderate	2	biomonitoring
Sanyal 2016(83)	2	2	0	1	1	1	1	0	2	10	moderate	2	biomonitoring
Schirinzi 2016(84)	2	1	0	0	1	1	1	0	2	8	moderate	2	biomonitoring
Stefano 2016(85)	2	1	0	0	1	0	1	0	2	7	moderate	2	biomonitoring
Verma 2016(86)	2	0	0	0	0	1	1	0	2	6	moderate	2	biomonitoring
Zuo 2016(87)	2	1	0	1	1	0	1	0	1	7	moderate	2	biomonitoring
Costa-Mallen 2017(88)	2	2	0	2	1	1	1	0	2	11	high	2	biomonitoring
de Farias 2017(89)	2	1	0	0	0	2	1	0	2	8	moderate	2	biomonitoring
Deng 2017(90)	2	2	0	1	0	1	1	0	2	9	moderate	2	biomonitoring
Gangania 2017(91)	2	0	0	2	1	0	1	0	2	8	moderate	2	biomonitoring
Hemmati-Dinarvand 2017(92)	2	1	0	0	0	2	1	0	2	8	moderate	2	biomonitoring
Casjens 2018(93)	2	1	0	2	1	0	1	0	0	7	moderate	2	biomonitoring
Dos Santos 2018(94)	2	1	0	2	1	1	1	0	2	10	moderate	2	biomonitoring
Ilyechova 2018(95)	2	1	0	0	1	1	1	0	1	7	moderate	2	biomonitoring
Maass 2018(96)	2	1	0	0	1	1	1	0	2	8	moderate	2	biomonitoring
Xu 2018(97)	2	1	0	0	1	1	1	0	2	8	moderate	2	biomonitoring
Shen 2019(98)	2	1	0	1	1	1	1	0	2	9	moderate	2	biomonitoring
Ajsuvakova 2020(99)	2	1	0	0	0	0	1	0	2	6	moderate	2	biomonitoring
Belvisi 2020(100)	2	2	0	0	1	2	0	0	2	9	moderate	1	self-report
Kissani 2020(101)	2	1	0	0	0	0	1	0	1	5	low	2	biomonitoring
Lucchini 2020(102)	2	2	0	1	1	2	0	0	2	10	moderate	1	self-report

EA, exposure assessment

[1] JEM, job exposure matrix; FFQ, food frequency questionnaire

* All studies assessed exposure after disease onset

Table S7. Quality assessment of cohort studies

Study	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Total	Grade	EA score	EA method[1] *
Logroscino 2008(106)	2	1	1	2	2	1	1	2	12	high	4	FFQ
Feldman 2011(108)	1	1	1	2	1	1	0	2	9	moderate	4	JEM
Palacios 2014(105)	1	1	1	2	2	1	1	2	11	high	4	environmental monitoring
Brouwer 2015(107)	1	1	0	2	1	1	0	2	8	moderate	4	JEM
Vinceti 2016(109)	2	1	0	2	1	1	0	2	9	moderate	4	environmental monitoring

EA, exposure assessment

[1] FFQ, food frequency questionnaire; JEM, job exposure matrix

* All studies assessed exposure before disease onset

Table S8. Subgroup analyses of metals in biological matrices

Subgroup	No. of studies	Pooled SMD	95% CI	*p*-value	I^2(%)	$p_{subgroup}$
Copper in CSF	11	0.16	-0.38, 0.70	0.526	86	
Geographic location						0.631
Western	9	0.07	-0.44, 0.58	0.750	82	
Non-western	2	0.65	-14.28, 15.57	0.680	96	
Detection method						0.014
AAS	7	0.21	-0.44, 0.85	0.462	86	
ICP-AES	2	-0.50	-0.52, -0.48	0.002	0	
ICP-MS	2	0.74	-12.93, 14.42	0.615	95	
Copper in serum	18	-0.43	-0.84, -0.02	0.043	94	
Geographic location						0.098
Western	10	-0.13	-0.39, 0.14	0.301	67	
Non-western	8	-0.80	-1.73, 0.12	0.079	95	
Detection method						<0.001
AAS	9	-0.63	-1.34, 0.07	0.071	95	
ICP-AES	4	0.16	-1.07, 1.39	0.709	89	
ICP-MS	3	-0.15	-0.27, -0.04	0.030	0	
Colorimetry	1		NA			
NR	1		NA			
Copper in plasma and serum	25	-0.23	-0.60, 0.14	0.210	96	
Geographic location						0.200
Western	15	-0.02	-0.37, 0.32	0.593	91	
Non-western	10	-0.54	-1.38, 0.30	0.180	98	
Detection method						<0.001
AAS	14	-0.37	-0.93, 0.18	0.180	97	
ICP-AES	4	0.16	-1.07, 1.39	0.709	89	
ICP-MS	5	0.21	-0.58, 1.00	0.502	94	
Colorimetry	1		NA			
NR	1		NA			
Iron in CSF	11	-0.29	-0.71, 0.13	0.155	81	
Geographic location						0.996
Western	7	-0.29	-0.91, 0.32	0.291	82	
Non-western	4	-0.29	-1.33, 0.75	0.436	84	
Detection method						<0.001
AAS	5	-0.13	-1.04, 0.78	0.709	87	
ICP-AES	3	-0.84	-1.11, -0.57	0.006	0	
ICP-MS	2	-0.25	-4.42, 3.91	0.583	61	
ELISA	1		NA			
Iron in serum	27	-0.28	-0.56, 0	0.049	89	
Geographic location						0.398
Western	15	-0.18	-0.42, 0.06	0.137	79	
Non-western	12	-0.43	-1.05, 0.19	0.152	94	
Detection method						0.005
AAS	4	-0.21	-0.60, 0.18	0.184	56	
ICP-AES	5	-0.09	-0.95, 0.77	0.784	94	
ICP-MS	3	0.15	-0.13, 0.42	0.149	0	
ELISA	1	N				
Colorimetry	9	-0.43	-1.23, 0.38	0.257	93	
NR	5	-0.59	-1.24, 0.06	0.065	89	
Iron in plasma and serum	32	-0.23	-0.49, 0.02	0.072	92	
Geographic location						0.878
Western	18	-0.22	-0.43, -0.02	0.034	76	
Non-western	14	-0.27	-0.84, 0.31	0.337	96	

Detection method						0.065
AAS	5	-0.09	-0.49, 0.31	0.562	85	
ICP-AES	5	-0.09	-0.95, 0.77	0.784	94	
ICP-MS	4	0.44	-0.36, 1.24	0.180	91	
ELISA	1		NA			
Colorimetry	12	-0.43	-0.99, 0.13	0.122	90	
NR	5	-0.59	-1.24, 0.06	0.065	89	
Manganese in plasma and serum	10	0.18	-0.29, 0.65	0.407	93	
Geographic location						0.533
Western	6	0.07	-0.71, 0.85	0.822	92	
Non-western	4	0.33	-0.56, 1.22	0.325	95	
Detection method						0.025
AAS	3	-0.31	-1.18, 0.57	0.272	74	
ICP-MS	7	0.40	-0.19, 0.99	0.146	93	
Selenium in plasma and serum	10	0.16	-0.52, 0.84		92	
Geographic location						0.735
Western	6	0.25	-0.47, 0.98	0.409	84	
Non-western	4	0.00	-2.16, 2.17	0.995	96	
Detection method						0.819
AAS	7	0.10	-0.97, 1.16	0.829	95	
ICP-MS	2	0.33	-2.83, 3.49	0.412	49	
NR	1		NA			
Zinc in serum	13	-0.33	-0.75, 0.09	0.113	85	
Geographic location						0.512
Western	7	-0.21	-0.48, 0.05	0.094	47	
Non-western	6	-0.50	-1.59, 0.58	0.289	93	
Detection method						0.034
AAS	7	-0.41	-1.29, 0.48	0.303	91	
ICP-AES	4	-0.37	-0.85, 0.12	0.098	65	
ICP-MS	2	0.00	-0.10, 0.09	0.739	0	
Zinc in plasma and serum	18	-0.53	-0.92, -0.14	0.011	94	
Geographic location						0.330
Western	10	-0.36	-0.77, 0.05	0.080	81	
Non-western	8	-0.74	-1.57, 0.08	0.071	97	
Detection method						0.735
AAS	10	-0.62	-1.28, 0.04	0.062	94	
ICP-AES	4	-0.37	-0.85, 0.12	0.098	65	
ICP-MS	4	-0.47	-1.91, 0.97	0.374	96	

NR, not reported; NA, not applicable

AAS, atomic absorption spectrometry; ICP-AES, inductively coupled plasma-atomic emission spectrometry; ICP-MS, inductively coupled plasma-mass spectrometry.

Table S9. Influential diagnosis of meta-analyses of metal levels in biological matrices

Metal	Sample	Pooled effect of original meta-analysis			Influential study	Pooled effect after influential study removal		|ΔSMD|
		No. of studies	SMD (95%CI)	I^2 (%)		SMD (95%CI)	I^2 (%)	
Chromium	serum	6	0.10 (-0.14, 0.34)	33	Gellein 2008 Sanyal 2016	0.10 (-0.37, 0.57)	35	0.00
Copper	CSF	11	0.16 (-0.38, 0.70)	86	none	NA		
	plasma	7	0.27 (-0.58, 1.12)	97	Arnal 2010	0.01 (-0.68, 0.70)	96	0.26
	serum	18	-0.43 (-0.84, -0.02)	94	none	NA		
	plasma+serum	25	-0.23 (-0.60, 0.14)	96	none	NA		
Iron	CSF	11	-0.29 (-0.71, 0.13)	81	none	NA		
	serum	27	-0.28 (-0.56, 0.00)	89	Gangania 2017	-0.20 (-0.40, 0.01)	87	0.08
	plasma+serum	32	-0.23 (-0.49, 0.02)	92	Gangania 2017	-0.16 (-0.36, 0.04)	91	0.07
Magnesium	serum	6	0.45 (-0.19, 1.09)	82	Hegde 2004	0.30 (0.07, 0.54)	25	0.15
Manganese	CSF	8	-0.15 (-0.64, 0.34)	76	none	NA		
	serum	8	0.11 (-0.43, 0.66)	89	Schirinzi 2016	-0.08 (-0.44, 0.28)	65	0.19
	plasma+serum	10	0.18 (-0.29, 0.65)	93	none	NA		
Selenium	serum	7	0.16 (-0.88, 1.20)	94	Nikam 2009	0.49 (-0.30, 1.28)	88	0.33
	plasma+serum	10	0.16 (-0.52, 0.84)	92	Nikam 2009	0.38 (-0.14, 0.89)	85	0.22
Zinc	CSF	7	-0.06 (-0.85, 0.73)	83	Hozumi 2011	-0.34 (-0.65, -0.03)	35	0.28
	serum	13	-0.33 (-0.75, 0.09)	85	Nikam 2009	-0.16 (-0.35, 0.02)	50	0.17
	plasma+serum	18	-0.53 (-0.92, -0.14)	94	none	NA		

NA, not applicable.

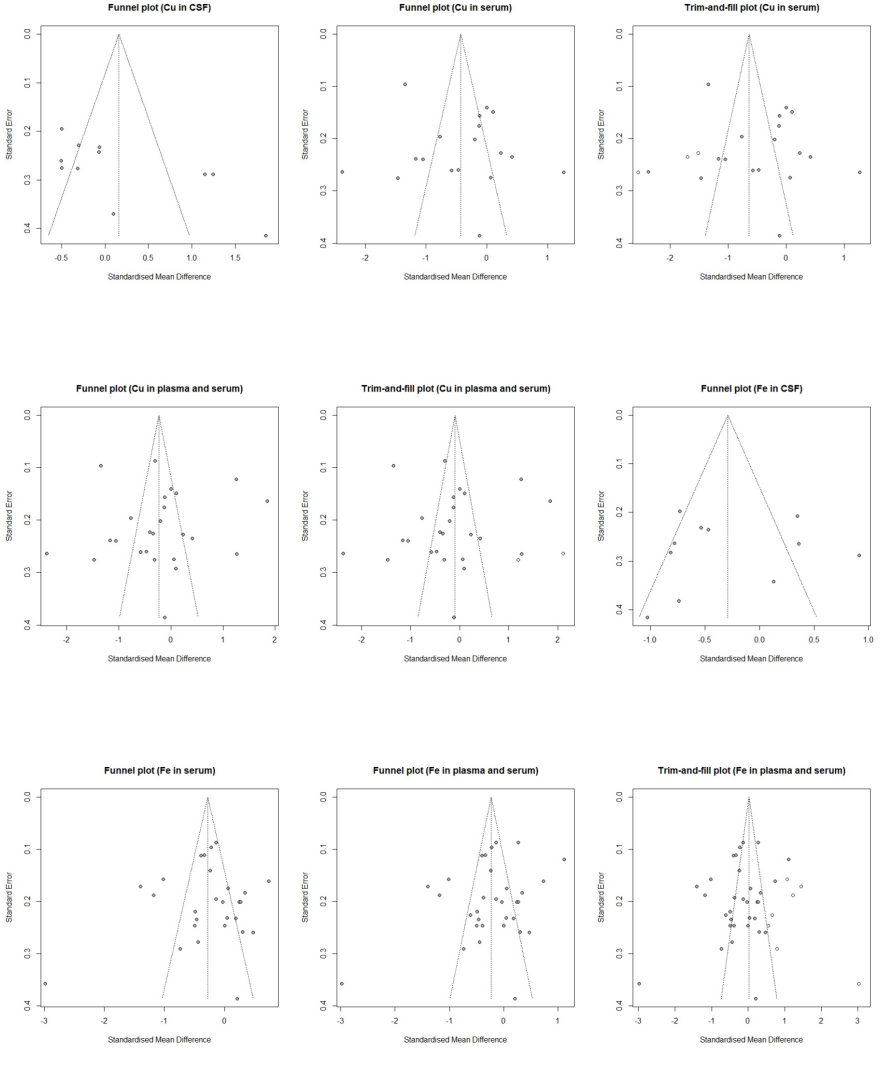
Table S10. Publication bias detection and adjustment of meta-analyses of metal levels in biological matrices

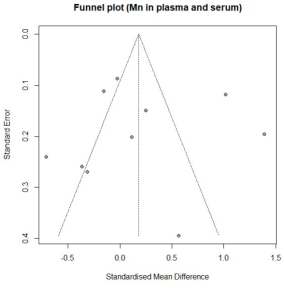
Metal	Sample	Pooled effect of original meta-analysis			Egger's test		Trim-and-fill method	
		No. of studies	Pooled SMD (95%CI)	I^2 (%)	Intercept estimate	p-value	No. of missing studies	Adjusted pooled SMD (95%CI)
Copper	CSF	11	0.16 (-0.38, 0.70)	86	8.661	0.030	none	NA
	serum	18	-0.43 (-0.84, -0.02)	94	2.713	0.347	3	-0.64 (-1.07, -0.21)
	plasma+serum	25	-0.23 (-0.60, 0.14)	96	-0.047	0.986	2	-0.10 (-0.49, 0.30)
Iron	CSF	11	-0.29 (-0.71, 0.13)	81	-0.642	0.854	none	NA
	serum	27	-0.28 (-0.56, 0.00)	89	-0.484	0.770	none	NA
	plasma+serum	32	-0.23 (-0.49, 0.02)	92	-2.180	0.194	7	0.02 (-0.27, 0.32)
Manganese	plasma+serum	10	0.18 (-0.29, 0.65)	93	-0.102	0.974	1	0.27 (-0.19, 0.73)
Selenium	Plasma+serum	10	0.16 (-0.52, 0.84)	92	-0.819	0.767	none	NA
Zinc	serum	13	-0.33 (-0.75, 0.09)	85	-1.478	0.426	none	NA
	plasma+serum	18	-0.53 (-0.92, -0.14)	94	1.086	0.633	3	-0.68 (-1.04, -0.31)

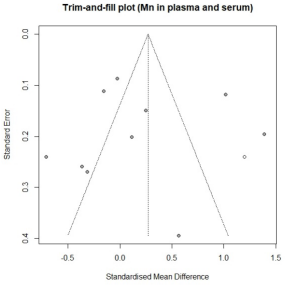
NA, not applicable

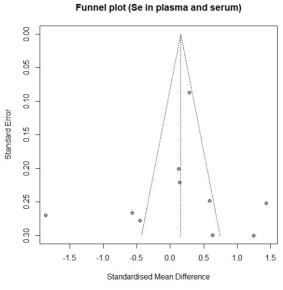
Figure S1. Funnel plots and trim-and-fill plots of meta-analyses of metal levels in biological matrices

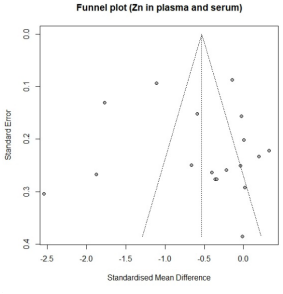
The closed dots indicate the observed studies, and the open dots indicate the missing studies imputed by the trim-and-fill method. The triangular area represents the 95% confidence limits, and the vertical line represents the overall effect size.

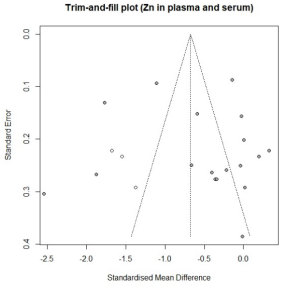



Funnel plot (Mn in plasma and serum)
Standard Error
Standardised Mean Difference


Trim-and-fill plot (Mn in plasma and serum)
Standard Error
Standardised Mean Difference


Funnel plot (Se in plasma and serum)
Standard Error
Standardised Mean Difference


Funnel plot (Zn in plasma and serum)
Standard Error
Standardised Mean Difference


Trim-and-fill plot (Zn in plasma and serum)
Standard Error
Standardised Mean Difference

References

1. Halliday G, Murphy K. Chapter 9 - Pathology of Parkinson's Disease. In: Schapira AHV, Lang AET, Fahn S, editors. Blue Books of Neurology. 34: Butterworth-Heinemann; 2010. p. 132-54.

2. Ascherio A, Schwarzschild MA. The epidemiology of Parkinson's disease: risk factors and prevention. The Lancet Neurology. 2016;15(12):1257-72.

3. Feigin VL, Abajobir AA, Abate KH, Abd-Allah F, Abdulle AM, Abera SF, et al. Global, regional, and national burden of neurological disorders during 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. The Lancet Neurology. 2017;16(11):877-97.

4. Bjorklund G, Stejskal V, Urbina MA, Dadar M, Chirumbolo S, Mutter J. Metals and Parkinson's Disease: Mechanisms and Biochemical Processes. Curr Med Chem. 2018;25(19):2198-214.

5. Raj K, Kaur P, Gupta GD, Singh S. Metals associated neurodegeneration in Parkinson's disease: Insight to physiological, pathological mechanisms and management. Neuroscience letters. 2021;753:135873.

6. Bowler RM, Roels HA, Nakagawa S, Drezgic M, Diamond E, Park R, et al. Dose-effect relationships between manganese exposure and neurological, neuropsychological and pulmonary function in confined space bridge welders. Occupational and environmental medicine. 2007;64(3):167-77.

7. Hsu YC, Chang CW, Lee HL, Chuang CC, Chiu HC, Li WY, et al. Association between History of Dental Amalgam Fillings and Risk of Parkinson's Disease: A Population-Based Retrospective Cohort Study in Taiwan. PloS one. 2016;11(12):e0166552.

8. Wells G, Shea B, O'Connell D, Peterson J, Welch V, Losos M, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality if nonrandomised studies in meta-analyses. Available from: URL http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp. The Ottawa Hospital Research Institute.

9. Sutedja NA, Veldink JH, Fischer K, Kromhout H, Heederik D, Huisman MH, et al. Exposure to chemicals and metals and risk of amyotrophic lateral sclerosis: a systematic review. Amyotroph Lateral Scler. 2009;10(5-6):302-9.

10. Teschke K, Olshan AF, Daniels JL, De Roos AJ, Parks CG, Schulz M, et al. Occupational exposure assessment in case-control studies: opportunities for improvement. Occupational and environmental medicine. 2002;59(9):575-93; discussion 94.

11. Higgins JP, White IR, Anzures-Cabrera J. Meta-analysis of skewed data: combining results reported on log-transformed or raw scales. Statistics in medicine. 2008;27(29):6072-92.

12. Vlaanderen J, Lan Q, Kromhout H, Rothman N, Vermeulen R. Occupational benzene exposure and the risk of lymphoma subtypes: a meta-analysis of cohort studies incorporating three study quality dimensions. Environmental health perspectives. 2011;119(2):159-67.

13. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al. Cochrane Handbook for Systematic Reviews of Interventions version 6.2 (updated February 2021). Cochrane. 2021.

14. Chinn S. A simple method for converting an odds ratio to effect size for use in meta-analysis. Statistics in medicine. 2000;19(22):3127-31.

15. Viechtbauer W. Bias and efficiency of meta-analytic variance estimators in the random-effects model. 2005;30(3):261-93.

16. Knapp G, Hartung J. Improved tests for a random effects meta‐regression with a single covariate. 2003;22(17):2693-710.

17. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Statistics in medicine. 2002;21(11):1539-58.

18. Viechtbauer W, Cheung MW. Outlier and influence diagnostics for meta-analysis. Res Synth Methods. 2010;1(2):112-25.

19. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ (Clinical research ed). 1997;315(7109):629-34.

20. Sue Duval, Tweedie R. Trim and Fill: A Simple Funnel-Plot-Based Method of Testing and Adjusting for Publication Bias in Meta-Analysis. Biometrics. 2000.

21. Team RC. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/. 2020.

22. Barbeau A, Jasmin G, Duchastel Y. Biochemistry of Parkinson's disease. Neurology. 1963;13:56-8.

23. Schwab RS, Poryali A, Ames A. Normal serum magnesium levels in Parkinson's disease. Neurology. 1964;14:855-6.

24. Pall HS, Williams AC, Blake DR, Lunec J, Gutteridge JM, Hall M, et al. Raised cerebrospinal-fluid copper concentration in Parkinson's disease. Lancet (London, England). 1987;2(8553):238-41.

25. Ngim CH, Devathasan G. Epidemiologic study on the association between body burden mercury level and idiopathic Parkinson's disease. Neuroepidemiology. 1989;8(3):128-41.

26. Wechsler LS, Checkoway H, Franklin GM, Costa LG. A pilot study of occupational and environmental risk factors for Parkinson's disease. Neurotoxicology. 1991;12(3):387-92.

27. Abbott RA, Cox M, Markus H, Tomkins A. Diet, body size and micronutrient status in Parkinson's disease. European journal of clinical nutrition. 1992;46(12):879-84.

28. Gazzaniga GC, Ferraro B, Camerlingo M, Casto L, Viscardi M, Mamoli A. A case control study of CSF copper, iron and manganese in Parkinson disease. Italian journal of neurological sciences. 1992;13(3):239-43.

29. Jiménez-Jiménez FJ, Fernández-Calle P, Martínez-Vanaclocha M, Herrero E, Molina JA, Vázquez A, et al. Serum levels of zinc and copper in patients with Parkinson's disease. Journal of the neurological sciences. 1992;112(1-2):30-3.

30. Cabrera-Valdivia F, Jiménez-Jiménez FJ, Molina JA, Férnandez-Calle P, Vázquez A, Cañizares-Liébana F, et al. Peripheral iron metabolism in patients with Parkinson's disease. Journal of the neurological sciences. 1994;125(1):82-6.

31. Jiménez-Jiménez FJ, Molina JA, Aguilar MV, Arrieta FJ, Jorge-Santamaría A, Cabrera-Valdivia F, et al. Serum and urinary manganese levels in patients with Parkinson's disease. Acta neurologica Scandinavica. 1995a;91(5):317-20.

32. Jiménez-Jiménez FJ, Molina JA, Arrieta FJ, Aguilar MV, Cabrera-Valdivia F, Vázquez A, et al. Decreased serum selenium concentrations in patients with Parkinson's disease. European journal of neurology. 1995b;2(2):111-4.

33. Seidler A, Hellenbrand W, Robra BP, Vieregge P, Nischan P, Joerg J, et al. Possible environmental, occupational, and other etiologic factors for Parkinson's disease: A case-control study in Germany. Neurology. 1996;46(5):1275-84.

34. Gorell JM, Johnson CC, Rybicki BA, Peterson EL, Kortsha GX, Brown GG, et al. Occupational exposures to metals as risk factors for Parkinson's disease. Neurology. 1997;48(3):650-8.

35. Logroscino G, Marder K, Graziano J, Freyer G, Slavkovich V, LoIacono N, et al. Altered systemic iron metabolism in Parkinson's disease. Neurology. 1997;49(3):714-7.

36. Aguilar MV, Jiménez-Jiménez FJ, Molina JA, Meseguer I, Mateos-Vega CJ, González-Muñoz MJ, et al. Cerebrospinal fluid selenium and chromium levels in patients with Parkinson's disease. Journal of neural transmission (Vienna, Austria : 1996). 1998;105(10-12):1245-51.

37. Jiménez-Jiménez FJ, Molina JA, Aguilar MV, Meseguer I, Mateos-Vega CJ, González-Muñoz MJ, et al. Cerebrospinal fluid levels of transition metals in patients with Parkinson's disease. Journal of neural transmission (Vienna, Austria : 1996). 1998;105(4-5):497-505.

38. Smargiassi A, Mutti A, De Rosa A, De Palma G, Negrotti A, Calzetti S. A case-control study of occupational and environmental risk factors for Parkinson's disease in the Emilia-Romagna region of Italy. Neurotoxicology. 1998;19(4-5):709-12.

39. Boll MC, Sotelo J, Otero E, Alcaraz-Zubeldia M, Rios C. Reduced ferroxidase activity in the cerebrospinal fluid from patients with Parkinson's disease. Neuroscience letters. 1999;265(3):155-8.

40. Johnson CC, Gorell JM, Rybicki BA, Sanders K, Peterson EL. Adult nutrient intake as a risk factor for Parkinson's disease. Int J Epidemiol. 1999;28(6):1102-9.

41. Tórsdóttir G, Kristinsson J, Sveinbjörnsdóttir S, Snaedal J, Jóhannesson T. Copper, ceruloplasmin, superoxide dismutase and iron parameters in Parkinson's disease. Pharmacology & toxicology. 1999;85(5):239-43.

42. Kocatürk PA, Akbostanci MC, Tan F, Kavas GO. Superoxide dismutase activity and zinc and copper concentrations in Parkinson's disease. Pathophysiology : the official journal of the International Society for Pathophysiology. 2000;7(1):63-7.

43. Pals P, Van Everbroeck B, Grubben B, Viaene MK, Dom R, Van Der Linden C, et al. Case-control study of environmental risk factors for Parkinson's disease in Belgium. European Journal of Epidemiology. 2003;18(12):1133-42.

44. Powers KM, Smith-Weller T, Franklin GM, Longstreth WT, Jr., Swanson PD, Checkoway H. Parkinson's disease risks associated with dietary iron, manganese, and other nutrient intakes. Neurology. 2003;60(11):1761-6.

45. Bocca B, Alimonti A, Petrucci F, Violante N, Sancesario G, Forte G, et al. Quantification of trace elements by sector field inductively coupled plasma mass spectrometry in urine, serum, blood and cerebrospinal fluid of patients with Parkinson's disease. Spectrochimica Acta Part B: Atomic Spectroscopy. 2004;59(4):559-66.

46. Hegde ML, Shanmugavelu P, Vengamma B, Rao TS, Menon RB, Rao RV, et al. Serum trace element levels and the complexity of inter-element relations in patients with Parkinson's disease. Journal of trace elements in medicine and biology : organ of the Society for Minerals and Trace Elements (GMS). 2004;18(2):163-71.

47. Bocca B, Alimonti A, Senofonte O, Pino A, Violante N, Petrucci F, et al. Metal changes in CSF and peripheral compartments of parkinsonian patients. Journal of the neurological sciences. 2006;248(1-2):23-30.

48. Coon S, Stark A, Peterson E, Gloi A, Kortsha G, Pounds J, et al. Whole-body lifetime occupational lead exposure and risk of Parkinson's disease. Environmental health perspectives. 2006;114(12):1872-6.

49. Qureshi GA, Qureshi AA, Memon SA, Parvez SH. Impact of selenium, iron, copper and zinc in on/off Parkinson's patients on L-dopa therapy. Journal of neural transmission Supplementum. 2006(71):229-36.

50. Annanmaki T, Muuronen A, Murros K. Low plasma uric acid level in Parkinson's disease. Movement disorders : official journal of the Movement Disorder Society. 2007;22(8):1133-7.

51. Dick FD, De Palma G, Ahmadi A, Scott NW, Prescott GJ, Bennett J, et al. Environmental risk factors for Parkinson's disease and parkinsonism: The Geoparkinson study. Occupational and environmental medicine. 2007;64(10):666-72.

52. Finkelstein MM, Jerrett M. A study of the relationships between Parkinson's disease and markers of traffic-derived and environmental manganese air pollution in two Canadian cities. Environmental research. 2007;104(3):420-32.

53. Alimonti A, Bocca B, Pino A, Ruggieri F, Forte G, Sancesario G. Elemental profile of cerebrospinal fluid in patients with Parkinson's disease. Journal of trace elements in medicine and biology : organ of the Society for Minerals and Trace Elements (GMS). 2007a;21(4):234-41.

54. Alimonti A, Ristori G, Giubilei F, Stazi MA, Pino A, Visconti A, et al. Serum chemical elements and oxidative status in Alzheimer's disease, Parkinson disease and multiple sclerosis. Neurotoxicology. 2007b;28(3):450-6.

55. Bharucha KJ, Friedman JK, Vincent AS, Ross ED. Lower serum ceruloplasmin levels correlate with younger age of onset in Parkinson's disease. Journal of neurology. 2008;255(12):1957-62.

56. Boll MC, Alcaraz-Zubeldia M, Montes S, Rios C. Free copper, ferroxidase and SOD1 activities, lipid peroxidation and NO(x) content in the CSF. A different marker profile in four neurodegenerative diseases. Neurochemical research. 2008;33(9):1717-23.

57. Gellein K, Syversen T, Steinnes E, Nilsen TI, Dahl OP, Mitrovic S, et al. Trace elements in serum from patients with Parkinson's disease--a prospective case-control study: the Nord-Trondelag Health Study (HUNT). Brain research. 2008;1219:111-5.

58. Petersen MS, Halling J, Bech S, Wermuth L, Weihe P, Nielsen F, et al. Impact of dietary exposure to food contaminants on the risk of Parkinson's disease. Neurotoxicology. 2008;29(4):584-90.

59. Nikam S, Nikam P, Ahaley SK, Sontakke AV. Oxidative stress in Parkinson's disease. Indian journal of clinical biochemistry : IJCB. 2009;24(1):98-101.

60. Powers KM, Smith-Weller T, Franklin GM, Longstreth WT, Jr., Swanson PD, Checkoway H. Dietary fats, cholesterol and iron as risk factors for Parkinson's disease. Parkinsonism & related disorders. 2009;15(1):47-52.

61. Arnal N, Cristalli DO, de Alaniz MJ, Marra CA. Clinical utility of copper, ceruloplasmin, and metallothionein plasma determinations in human neurodegenerative patients and their first-degree relatives. Brain research. 2010;1319:118-30.

62. Baillet A, Chanteperdrix V, Trocmé C, Casez P, Garrel C, Besson G. The role of oxidative stress in amyotrophic lateral sclerosis and Parkinson's disease. Neurochemical research. 2010;35(10):1530-7.

63. Brewer GJ, Kanzer SH, Zimmerman EA, Molho ES, Celmins DF, Heckman SM, et al. Subclinical zinc deficiency in Alzheimer's disease and Parkinson's disease. American journal of Alzheimer's disease and other dementias. 2010;25(7):572-5.

64. Fukushima T, Tan X, Luo Y, Kanda H. Relationship between blood levels of heavy metals and Parkinson's disease in China. Neuroepidemiology. 2010;34(1):18-24.

65. Weisskopf MG, Weuve J, Nie H, Saint-Hilaire MH, Sudarsky L, Simon DK, et al. Association of cumulative lead exposure with Parkinson's disease. Environmental health perspectives. 2010;118(11):1609-13.

66. Fukushima T, Tan X, Luo Y, Kanda H. Serum vitamins and heavy metals in blood and urine, and the correlations among them in Parkinson's disease patients in China. Neuroepidemiology. 2011;36(4):240-4.

67. Hozumi I, Hasegawa T, Honda A, Ozawa K, Hayashi Y, Hashimoto K, et al. Patterns of levels of biological metals in CSF differ among neurodegenerative diseases. Journal of the neurological sciences. 2011;303(1-2):95-9.

68. Ling H, Bhidayasiri R. Reduced serum caeruloplasmin levels in non-wilsonian movement disorders. European neurology. 2011;66(3):123-7.

69. Miyake Y, Tanaka K, Fukushima W, Sasaki S, Kiyohara C, Tsuboi Y, et al. Lack of association of dairy food, calcium, and vitamin D intake with the risk of Parkinson's disease: a case-control study in Japan. Parkinsonism & related disorders. 2011a;17(2):112-6.

70. Miyake Y, Tanaka K, Fukushima W, Sasaki S, Kiyohara C, Tsuboi Y, et al. Dietary intake of metals and risk of Parkinson's disease: a case-control study in Japan. Journal of the neurological sciences. 2011b;306(1-2):98-102.

71. Farhoudi M, Taheraghdam A, Farid GA, Talebi M, Pashapou A, Majidi J, et al. Serum iron and ferritin level in idiopathic Parkinson. Pakistan journal of biological sciences : PJBS. 2012;15(22):1094-7.

72. Madenci G, Bilen S, Arli B, Saka M, Ak F. Serum iron, vitamin B12 and folic acid levels in Parkinson's disease. Neurochemical research. 2012;37(7):1436-41.

73. McIntosh KG, Cusack MJ, Vershinin A, Chen ZW, Zimmerman EA, Molho ES, et al. Evaluation of a prototype point-of-care instrument based on monochromatic x-ray fluorescence spectrometry: potential for monitoring trace element status of subjects with neurodegenerative disease. Journal of toxicology and environmental health Part A. 2012;75(21):1253-68.

74. Mariani S, Ventriglia M, Simonelli I, Donno S, Bucossi S, Vernieri F, et al. Fe and Cu do not differ in Parkinson's disease: a replication study plus meta-analysis. Neurobiology of aging. 2013;34(2):632-3.

75. Younes-Mhenni S, Aissi M, Mokni N, Boughammoura-Bouatay A, Chebel S, Frih-Ayed M, et al. Serum copper, zinc and selenium levels in Tunisian patients with Parkinson's disease. La Tunisie medicale. 2013;91(6):402-5.

76. Zhao HW, Lin J, Wang XB, Cheng X, Wang JY, Hu BL, et al. Assessing plasma levels of selenium, copper, iron and zinc in patients of Parkinson's disease. PloS one. 2013;8(12):e83060.

77. Kumudini N, Uma A, Devi YP, Naushad SM, Mridula R, Borgohain R, et al. Association of Parkinson's disease with altered serum levels of lead and transition metals among South Indian subjects. Indian journal of biochemistry & biophysics. 2014;51(2):121-6.

78. Arain MS, Afridi HI, Kazi TG, Talpur FN, Arain MB, Kazi A, et al. Correlation of aluminum and manganese concentration in scalp hair samples of patients having neurological disorders. Environmental monitoring and assessment. 2015;187(2):10.

79. Costa-Mallen P, Zabetian CP, Agarwal P, Hu SC, Yearout D, Samii A, et al. Haptoglobin phenotype modifies serum iron levels and the effect of smoking on Parkinson disease risk. Parkinsonism & related disorders. 2015;21(9):1087-92.

80. van der Mark M, Vermeulen R, Nijssen PCG, Mulleners WM, Sas AMG, van Laar T, et al. Occupational exposure to solvents, metals and welding fumes and risk of Parkinson's disease. Parkinsonism and Related Disorders. 2015;21(6):635-9.

81. Mariani S, Ventriglia M, Simonelli I, Bucossi S, Siotto M, Donno S, et al. Association between sex, systemic iron variation and probability of Parkinson's disease. The International journal of neuroscience. 2016;126(4):354-60.

82. Medeiros MS, Schumacher-Schuh A, Cardoso AM, Bochi GV, Baldissarelli J, Kegler A, et al. Iron and Oxidative Stress in Parkinson's Disease: An Observational Study of Injury Biomarkers. PloS one. 2016;11(1):e0146129.

83. Sanyal J, Ahmed SS, Ng HK, Naiya T, Ghosh E, Banerjee TK, et al. Metallomic Biomarkers in Cerebrospinal fluid and Serum in patients with Parkinson's disease in Indian population. Scientific reports. 2016;6:35097.

84. Schirinzi T, Martella G, D'Elia A, Di Lazzaro G, Imbriani P, Madeo G, et al. Outlining a population "at risk" of Parkinson's disease: Evidence from a case-control study. Parkinson's Disease. 2016;2016.

85. Stefano F, Cinzia N, Marco P, Marco G, Rita G, Augusto F, et al. Hair Microelement Profile as a Prognostic Tool in Parkinson's Disease. Toxics. 2016;4(4).

86. Verma AK, Keshari AK, Raj J, Kumari R, Kumar T, Sharma V, et al. Prolidase-Associated Trace Elements (Mn, Zn, Co, and Ni) in the Patients with Parkinson's Disease. Biological trace element research. 2016;171(1):48-53.

87. Zuo LJ, Yu SY, Hu Y, Wang F, Piao YS, Lian TH, et al. Serotonergic dysfunctions and abnormal iron metabolism: Relevant to mental fatigue of Parkinson disease. Scientific reports. 2016;6(1):19.

88. Costa-Mallen P, Gatenby C, Friend S, Maravilla KR, Hu SC, Cain KC, et al. Brain iron concentrations in regions of interest and relation with serum iron levels in Parkinson disease. Journal of the neurological sciences. 2017;378:38-44.

89. de Farias CC, Maes M, Bonifacio KL, Matsumoto AK, Bortolasci CC, Nogueira AS, et al. Parkinson's Disease is Accompanied by Intertwined Alterations in Iron Metabolism and Activated Immune-inflammatory and Oxidative Stress Pathways. CNS & neurological disorders drug targets. 2017;16(4):484-91.

90. Deng Q, Zhou X, Chen J, Pan M, Gao H, Zhou J, et al. Lower hemoglobin levels in patients with parkinson's disease are associated with disease severity and iron metabolism. Brain research. 2017;1655:145-51.

91. Gangania MK, Batra J, Kushwaha S, Agarwal R. Role of Iron and Copper in the Pathogenesis of Parkinson's Disease. Indian journal of clinical biochemistry : IJCB. 2017;32(3):353-6.

92. Hemmati-Dinarvand M, Taher-Aghdam AA, Mota A, Zununi Vahed S, Samadi N. Dysregulation of serum NADPH oxidase1 and ferritin levels provides insights into diagnosis of Parkinson's disease. Clinical biochemistry. 2017;50(18):1087-92.

93. Casjens S, Dydak U, Dharmadhikari S, Lotz A, Lehnert M, Quetscher C, et al. Association of exposure to manganese and iron with striatal and thalamic GABA and other neurometabolites — Neuroimaging results from the WELDOX II study. Neurotoxicology. 2018;64:60-7.

94. Dos Santos AB, Kohlmeier KA, Rocha ME, Barreto GE, Barreto JA, de Souza ACA, et al. Hair in Parkinson's disease patients exhibits differences in Calcium, Iron and Zinc concentrations measured by flame atomic absorption spectrometry - FAAS. Journal of trace elements in medicine and biology : organ of the Society for Minerals and Trace Elements (GMS). 2018;47:134-9.

95. Ilyechova EY, Miliukhina IV, Orlov IA, Muruzheva ZM, Puchkova LV, Karpenko MN. A low blood copper concentration is a co-morbidity burden factor in Parkinson's disease development. Neuroscience research. 2018;135:54-62.

96. Maass F, Michalke B, Leha A, Boerger M, Zerr I, Koch JC, et al. Elemental fingerprint as a cerebrospinal fluid biomarker for the diagnosis of Parkinson's disease. Journal of neurochemistry. 2018;145(4):342-51.

97. Xu W, Zhi Y, Yuan Y, Zhang B, Shen Y, Zhang H, et al. Correlations between abnormal iron metabolism and non-motor symptoms in Parkinson's disease. Journal of neural transmission (Vienna, Austria : 1996). 2018;125(7):1027-32.

98. Shen X, Yang H, Zhang D, Jiang H. Iron Concentration Does Not Differ in Blood but Tends to Decrease in Cerebrospinal Fluid in Parkinson's Disease. Front Neurosci. 2019;13:939.

99. Ajsuvakova OP, Tinkov AA, Willkommen D, Skalnaya AA, Danilov AB, Pilipovich AA, et al. Assessment of copper, iron, zinc and manganese status and speciation in patients with Parkinson's disease: A pilot study. Journal of trace elements in medicine and biology : organ of the Society for Minerals and Trace Elements (GMS). 2020;59:126423.

100. Belvisi D, Pellicciari R, Fabbrini A, Costanzo M, Pietracupa S, De Lucia M, et al. Risk factors of Parkinson disease Simultaneous assessment, interactions, and etiologic subtypes. Neurology. 2020;95(18):E2500-E8.

101. Kissani N, Naji Y, Mebrouk Y, Chraa M, Ghanima A, Reis J. Parkinsonism and chronic manganese exposure: Pilot study with clinical, environmental and experimental evidence. Clinical Parkinsonism and Related Disorders. 2020;3.

102. Lucchini RG, Guazzetti S, Renzetti S, Broberg K, Caci M, Covolo L, et al. Metal Exposure and SNCA rs356219 Polymorphism Associated With Parkinson Disease and Parkinsonism. Frontiers in Neurology. 2020;11.

103. Frigerio R, Elbaz A, Sanft KR, Peterson BJ, Bower JH, Ahlskog JE, et al. Education and occupations preceding Parkinson disease: a population-based case-control study. Neurology. 2005;65(10):1575-83.

104. Firestone JA, Lundin JI, Powers KM, Smith-Weller T, Franklin GM, Swanson PD, et al. Occupational factors and risk of Parkinson's disease: A population-based case-control study. American journal of industrial medicine. 2010;53(3):217-23.

105. Palacios N, Fitzgerald K, Roberts AL, Hart JE, Weisskopf MG, Schwarzschild MA, et al. A prospective analysis of airborne metal exposures and risk of Parkinson disease in the nurses' health study cohort. Environmental health perspectives. 2014;122(9):933-8.

106. Logroscino G, Gao X, Chen H, Wing A, Ascherio A. Dietary iron intake and risk of Parkinson's disease. American journal of epidemiology. 2008;168(12):1381-8.

107. Brouwer M, Koeman T, van den Brandt PA, Kromhout H, Schouten LJ, Peters S, et al. Occupational exposures and Parkinson's disease mortality in a prospective Dutch cohort. Occupational and environmental medicine. 2015;72(6):448-55.

108. Feldman AL, Johansson ALV, Nise G, Gatz M, Pedersen NL, Wirdefeldt K. Occupational exposure in Parkinsonian disorders: A 43-year prospective cohort study in men. Parkinsonism and Related Disorders. 2011;17(9):677-82.

109. Vinceti M, Ballotari P, Steinmaus C, Malagoli C, Luberto F, Malavolti M, et al. Long-term mortality patterns in a residential cohort exposed to inorganic selenium in drinking water. Environmental research. 2016;150:348-56.

110. Aschner M, Erikson KM, Herrero Hernández E, Tjalkens R. Manganese and its role in Parkinson's disease: from transport to neuropathology. Neuromolecular medicine. 2009;11(4):252-66.

111. Chang J, Kueon C, Kim J. Influence of lead on repetitive behavior and dopamine metabolism in a mouse model of iron overload. Toxicological research. 2014;30(4):267-76.

112. Cariccio VL, Samà A, Bramanti P, Mazzon E. Mercury Involvement in Neuronal Damage and in Neurodegenerative Diseases. Biological trace element research. 2019;187(2):341-56.

113. Ma L, Gholam Azad M, Dharmasivam M, Richardson V, Quinn RJ, Feng Y, et al. Parkinson's disease: Alterations in iron and redox biology as a key to unlock therapeutic strategies. Redox biology. 2021;41:101896.

114. Lingam I, Robertson NJ. Magnesium as a Neuroprotective Agent: A Review of Its Use in the Fetus, Term Infant with Neonatal Encephalopathy, and the Adult Stroke Patient. Developmental neuroscience. 2018;40(1):1-12.

115. Stelmashook EV, Isaev NK, Genrikhs EE, Amelkina GA, Khaspekov LG, Skrebitsky VG, et al. Role of zinc and copper ions in the pathogenetic mechanisms of Alzheimer's and Parkinson's diseases. Biochemistry Biokhimiia. 2014;79(5):391-6.

116. Ellwanger JH, Franke SI, Bordin DL, Prá D, Henriques JA. Biological functions of selenium and its potential influence on Parkinson's disease. Anais da Academia Brasileira de Ciencias. 2016;88(3 Suppl):1655-74.

117. Jackson D, White IR. When should meta-analysis avoid making hidden normality assumptions? Biom J. 2018;60(6):1040-58.

118. Guilarte TR. Manganese and Parkinson's disease: a critical review and new findings. Environmental health perspectives. 2010;118(8):1071-80.

119. Savitz DA, Forastiere F. Do pooled estimates from meta-analyses of observational epidemiology studies contribute to causal inference? Occupational and environmental medicine. 2021;78(9):621-2.

120. Wan X, Wang W, Liu J, Tong T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. BMC Med Res Methodol. 2014;14:135.

121. Walter SD, Yao X. Effect sizes can be calculated for studies reporting ranges for outcome variables in systematic reviews. Journal of clinical epidemiology. 2007;60(8):849-52.

Chapter 3

Pre-diagnostic Blood Metal Levels and the Risk of Parkinson's Disease: A Large European Prospective Cohort

Yujia Zhao, Anushree Ray, Karin Broberg, Maria Kippler, Christina M Lill, Paolo Vineis, Verena A Katzke, Miguel Rodriguez-Barranco, María-Dolores Chirlaque, Marcela Guevara, Jesús-Humberto Gómez, Johnni Hansen, Salvatore Panico, Lefkos T Middleton, Giovanna Masala, Valeria Pala, Ana Vinagre-Aragon, Maurizio Zibetti, Roel Vermeulen, Susan Peters

Corresponding author: Susan Peters

Published *in Movement Disorders*, September 2023

https://pubmed.ncbi.nlm.nih.gov/37675653/

Abstract

Background: Metals have been postulated as environmental concerns in the etiology of Parkinson's disease (PD), but metal levels are typically measured after diagnosis, which might be subject to reverse causality.

Objectives: To investigate the association between pre-diagnostic blood metal levels and PD risk.

Methods: A case-control study nested in a prospective European cohort, utilizing erythrocyte samples collected before PD diagnosis.

Results: Most assessed metals were not associated with PD risk. Cadmium has a suggestive negative association with PD (odds ratio (OR) and 95% confidence interval (CI) for the highest quartile, 0.70 (0.42-1.17)), which diminished among never smokers. Among current smokers only, lead was associated with decreased PD risk (0.06 (0.01-0.35)), while arsenic showed associations towards an increased PD risk (1.85 (0.45-7.93)).

Conclusions: We observe no strong evidence to support a role of metals in the development of PD. In particular smoking may confound the association with tobacco-derived metals.

Keywords: Parkinson's disease, metals, prospective exposure assessment, cohort study

Introduction

Parkinson's disease (PD) is the second most common neurodegenerative disease. Around 90% of PD cases are related to a variety of lifestyle and environmental factors (1). Metals have been implicated in the pathogenesis of PD for many years (2). For the general population, metal exposures usually result from contaminated food or drinking water, cigarette smoking, air pollution, dental amalgam fillings, medication and dietary supplements (3).

We recently evaluated the epidemiological evidence of associations between metal exposure and PD risk in a systematic review (2). We did not observe consistent associations with PD risk for most metals in meta-analyses. Notably, the research quality was greatly limited due to retrospective exposure assessment. Metal levels in human biofluids were mostly measured after disease diagnosis, and results were therefore possibly affected by reverse causality because PD patients tend to change smoking and diet habits when experiencing clinical manifestations (4).

To elucidate whether metal exposures represent genuine risk factors for PD, we assessed the association between PD risk and metal levels in blood samples collected several years before PD diagnosis, in a case-control study nested in the EPIC4PD cohort, a large prospective European study (5). Meanwhile, possible confounding of smoking was also explored for the effect of metals on PD, since smoking was reported to be inversely associated with PD risk (6) and cigarette smoke contains various metal species (7).

Methods

Study population

The EPIC4PD study is based on 220,494 subjects from the general population residing in seven countries, within the European Prospective Investigation into Cancer and Nutrition (EPIC) study (5, 8). Within EPIC4PD, 734 incident PD cases who received a diagnosis after the date of recruitment were identified through medical record linkage and neurologist validation (5). Here, we conducted a nested case-control study within the EPIC4PD cohort, including 362 incident PD cases for whom a baseline erythrocyte sample was available in the EPIC biobank. The reliability of the diagnoses was categorized into 'definite', 'very likely', 'probable' and 'possible' (Text S1) (5). One control per case matched by age at recruitment, sex and study center was selected using incidence density sampling.

Measurement of metal levels

Metal concentrations in erythrocytes were measured by inductively coupled plasma-mass spectrometry. Eleven elements were assessed: arsenic, cadmium, calcium, copper, iron, lead, magnesium, manganese, mercury, selenium and zinc. Details were described in Text S2.

Statistical analysis

Metal concentrations were compared among subjects with different smoking status. Correlations between metal levels and smoking intensity, represented by the reported number of cigarettes smoked per day at baseline when blood was collected, were tested by Spearman correlation (correlation coefficient, *rho*).

Conditional logistic regression for the matched case-control sets was applied to estimate the odds ratio (OR) and 95% confidence interval (CI) of PD incidence associated with quartile categories of metal levels (based on the distribution among controls, denoted as Q1-Q4). Considering the recognized inverse association between smoking and the risk of PD (6), smoking status at recruitment (never, former or current smoker, and unknown) was deemed as a potential confounder in the conditional analyses. Other possible confounding factors, including alcohol consumption, coffee drinking, seafood and vegetable intake, education, body mass index and physical activity, did not modify the risk estimated (all $p>0.1$) and were not included in the models.

We performed stratified analyses by sex and smoking status (current, non-current and never smokers) to test possible different effects. Two sensitivity analyses were conducted to test the robustness of our findings: i) limiting analyses to PD cases diagnosed after 8 years (median) since recruitment (timing for blood collection) to reduce possible reverse causality; and ii) limiting analyses to PD cases with definite and very likely diagnoses. All analyses were carried out in R 4.1.3 (9).

Results

For PD cases, the median period between recruitment and PD diagnosis was 7.8 years (Table S2). At baseline, 13% of the cases were smokers, compared with 16% among controls ($p=0.69$). Cadmium levels in current smokers were about 2.5 times higher than in never smokers (Table S3), and lead in current smokers was around 1.4 times higher compared with never smokers. Furthermore, cadmium and lead levels were both positively correlated with the number of daily smoked cigarettes (*rho*=0.50, 0.26, respectively).

Table 1. The association between pre-diagnostic blood metal concentrations and the risk of Parkinson's disease

Exposure category	PD cases, n	Controls, n	OR (95% CI)[1]	Smoking adjusted OR (95% CI)[2]
Arsenic				
Quartile 1	94	91	1.00 [Ref]	1.00 [Ref]
Quartile 2	97	90	1.02 (0.66–1.57)	1.02 (0.66–1.58)
Quartile 3	76	90	0.78 (0.48–1.27)	0.78 (0.48–1.26)
Quartile 4	95	91	0.97 (0.61–1.56)	0.96 (0.60–1.55)
p for trend, linear			0.46	0.47
Cadmium				
Quartile 1	118	91	1.00 [Ref]	1.00 [Ref]
Quartile 2	71	90	0.59 (0.39–0.91)	0.59 (0.38–0.90)
Quartile 3	90	90	0.75 (0.49–1.15)	0.75 (0.48–1.15)
Quartile 4	83	91	0.68 (0.44–1.04)	0.70 (0.42–1.17)
p for trend, linear			0.09	0.13
Calcium				
Quartile 1	80	91	1.00 [Ref]	1.00 [Ref]
Quartile 2	82	90	1.14 (0.70–1.87)	1.14 (0.70–1.88)
Quartile 3	114	90	1.62 (0.98–2.68)	1.60 (0.97–2.66)
Quartile 4	86	91	1.20 (0.63–2.29)	1.16 (0.61–2.23)
p for trend, linear			0.32	0.37
Copper				
Quartile 1	94	91	1.00 [Ref]	1.00 [Ref]
Quartile 2	77	90	0.85 (0.55–1.31)	0.85 (0.55–1.31)
Quartile 3	90	90	1.01 (0.62–1.64)	1.01 (0.62–1.65)
Quartile 4	101	91	1.20 (0.70–2.08)	1.20 (0.69–2.08)
p for trend, linear			0.84	0.83
Iron				
Quartile 1	97	91	1.00 [Ref]	1.00 [Ref]
Quartile 2	82	90	0.84 (0.54–1.32)	0.84 (0.54–1.32)
Quartile 3	100	90	1.02 (0.62–1.68)	1.00 (0.60–1.65)
Quartile 4	83	91	0.82 (0.48–1.41)	0.82 (0.48–1.42)
p for trend, linear			0.64	0.62
Lead				
Quartile 1	101	91	1.00 [Ref]	1.00 [Ref]
Quartile 2	92	90	0.90 (0.60–1.36)	0.93 (0.61–1.41)
Quartile 3	91	90	0.85 (0.56–1.31)	0.87 (0.56–1.34)
Quartile 4	78	91	0.65 (0.38–1.11)	0.68 (0.39–1.16)
p for trend, linear			0.09	0.12

(**Table 1** continued)

Magnesium				
Quartile 1	88	91	1.00 [Ref]	1.00 [Ref]
Quartile 2	92	90	1.06 (0.69–1.65)	1.05 (0.68–1.63)
Quartile 3	85	90	0.99 (0.63–1.57)	0.99 (0.62–1.56)
Quartile 4	97	91	1.12 (0.71–1.77)	1.12 (0.71–1.77)
p for trend, linear			0.99	0.99
Manganese				
Quartile 1	97	91	1.00 [Ref]	1.00 [Ref]
Quartile 2	99	90	1.02 (0.68–1.53)	1.01 (0.67–1.53)
Quartile 3	79	90	0.80 (0.52–1.24)	0.80 (0.51–1.25)
Quartile 4	87	91	0.87 (0.55–1.37)	0.86 (0.55–1.36)
p for trend, linear			0.60	0.57
Mercury				
Quartile 1	95	91	1.00 [Ref]	1.00 [Ref]
Quartile 2	82	90	0.87 (0.56–1.34)	0.88 (0.57–1.35)
Quartile 3	89	90	0.96 (0.60–1.54)	0.96 (0.60–1.54)
Quartile 4	96	91	1.10 (0.61–1.97)	1.11 (0.61–2.00)
p for trend, linear			0.95	0.98
Selenium				
Quartile 1	90	91	1.00 [Ref]	1.00 [Ref]
Quartile 2	103	90	1.14 (0.76–1.71)	1.12 (0.75–1.68)
Quartile 3	61	90	0.70 (0.45–1.09)	0.68 (0.43–1.07)
Quartile 4	108	91	1.28 (0.79–2.05)	1.25 (0.77–2.01)
p for trend, linear			0.94	0.88
Zinc				
Quartile 1	72	91	1.00 [Ref]	1.00 [Ref]
Quartile 2	98	90	1.43 (0.92–2.21)	1.41 (0.91–2.18)
Quartile 3	103	90	1.53 (0.97–2.42)	1.54 (0.98–2.45)
Quartile 4	89	91	1.35 (0.80–2.27)	1.35 (0.80–2.28)
p for trend, linear			0.51	0.54

PD, Parkinson's disease; OR, odds ratio; CI, confidence interval.

[1] Conditional logistic regression for the matched case-control sets.

[2] Conditional logistic regression for the matched case-control sets, adjusted by smoking status.

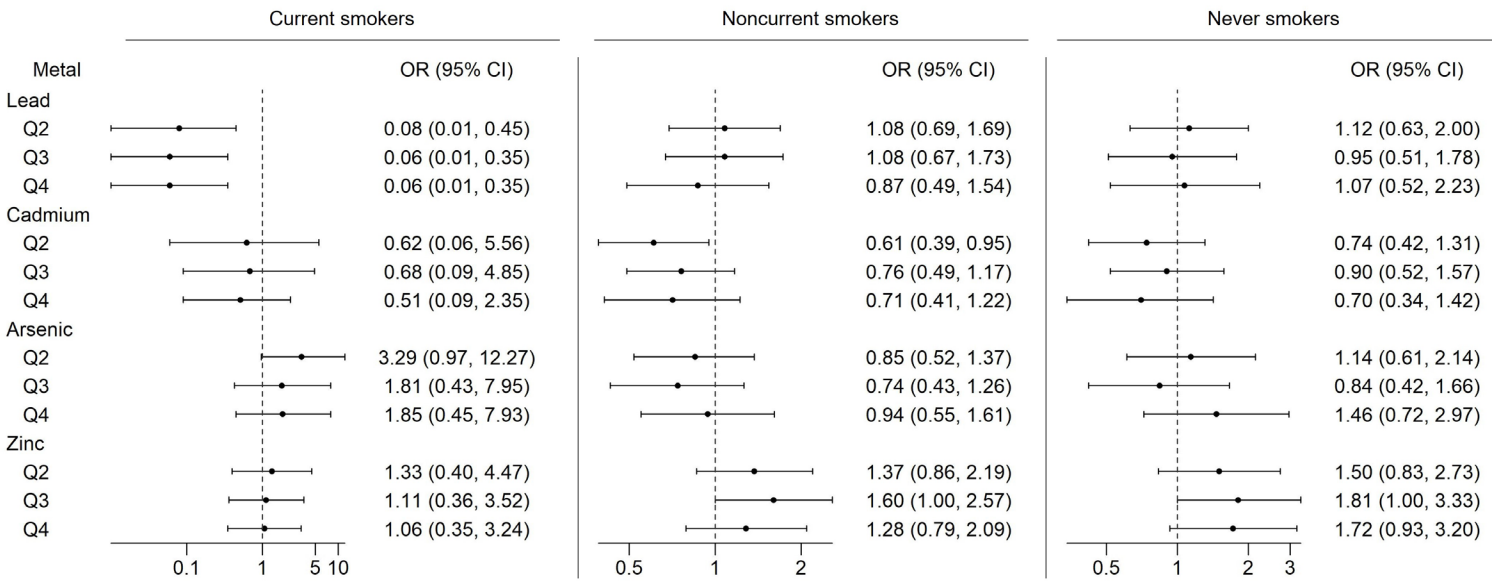


Figure 1. The association between metal concentrations and the risk of Parkinson's disease by smoking status

OR, odds ratio; CI, confidence interval, Q, quartile.

ORs were calculated by unconditional logistic regression, adjusted by age at recruitment, sex and country, quartile 1 as reference.

Our analyses did not demonstrate obvious associations between assessed metals and the risk of PD (Table 1). Cadmium was indicated to be associated with a decreased risk of PD (ORs and 95% CIs: Q2 0.59 (0.38–0.90), Q3 0.75 (0.48–1.15), Q4 0.70 (0.42–1.17)). However, the effects attenuated towards null when limited to never smokers (Figure 1). Zinc showed a borderline positive association in the third quartile (1.54 (0.98–2.45), *p*=0.064), which became most pronounced among never smokers (1.81 (1.00–3.33)) (Figure 1) and females (2.32 (1.16–4.63)) (Figure S4). No effects were observed for the remaining metals: calcium, copper, iron, magnesium, manganese, mercury, and selenium.

Exposure-response trends were not observed from linear (Table 1) nor spline regression for any of the assessed metals (Figure S2). Effect estimates from sensitivity analyses limiting to late-diagnosed cases and definite and very likely cases were similar to those of the main analyses, despite the widening of CIs due to the smaller sample size (Figure S5).

A few metals showed associations only among current smokers (Figure 1). Increased lead levels were found to be associated with a decreased risk of PD among current smokers (OR for highest quartile 0.06, 95% CI 0.01–0.35), also showing a clear linear trend (*p*=0.007). The inverse associations persisted after further controlling for the number of cigarettes (data not shown). In contrast, a positive association was suggested between arsenic and PD risk when limited to current smokers (ORs and 95% CIs: Q2 3.29 (0.97–12.3), Q3 1.81 (0.43–7.95), Q4 1.85 (0.45–7.93)).

Discussion

Our study is the first prospective study to investigate the role of metal levels in PD risk by assessing blood samples collected years prior to the clinical diagnosis. Our study did not provide robust evidence to verify the action of metals in the pathogenesis of PD.

Lead is a well-recognized toxicant primarily affecting the central nervous system (10). Two well-designed case-control studies found bone lead, which is a biomarker for cumulative lead exposure (11), to be associated with increased risk of PD (12, 13), indicating that long-term environmental lead exposure may be a risk factor for PD. In contrast, a strong inverse association between lead levels and PD risk was notable among current smokers in our study. A probable explanation is that lead is a surrogate measure or intermediate step of smoking in relation to PD, considering the positive correlation between lead levels and smoking intensity. This assumption is partly confirmed by our previous study on erythrocyte metal levels and the risk of amyotrophic lateral sclerosis (ALS) (14). In contrast to the observed decreased risk of

PD, smoking is associated with an increased risk of ALS (15), and a positive association between blood lead and ALS was found in current smokers.

Cadmium was also possibly subject to the influence of smoking. In terms of OR magnitude, the negative associations between cadmium and PD were stronger among current smokers than never smokers. Because cadmium source attributes to cigarette smoking more than lead (16), it is more reasonable to speculate the participation of smoking in the relation of cadmium to PD. Future studies should aim to explore the modification effect of smoking in the relation between tobacco-derived metals and neurodegenerations and to help elucidate the etiology of the disease.

In this study, subjects with higher zinc levels showed an elevated risk of PD, especially among never smokers and females, but no clear dose-response trend was observed. In our previous meta-analyses, zinc in the blood matrix were found to be lower in PD patients compared with controls, with a pooled standardized mean difference of -0.53 (95% CI:-0.92, -0.14) from 18 retrospective case-control studies (2). The discrepancy with results in our study may be due to different biospecimens measured (blood matrix or erythrocytes) and possible reverse causality in previous studies. The precise nature and underlying mechanisms of the effect of zinc in the development of PD requires further investigation.

We acknowledged some limitations in this study. First, metal levels in erythrocytes (with a life span of 120 days (17)) reflect recent exposures proximal to the time of blood sample collection. The biomonitoring at one time point could be inaccurate when metal exposures fluctuate over the lifetime. For example, unlike elevated levels in current smokers, cadmium concentrations in former smokers were similar as in never smokers (Table S3), suggesting one-time measurement cannot completely reflect past exposure. Second, blood samples were collected on average eight years prior to PD diagnosis, but PD prodromal phase could occur as early as 20 years before the onset of motor symptoms (18). Therefore, we cannot fully exclude that metal alterations were secondary to diet and smoking habit change related to PD symptoms. However, similar results were obtained when limiting analyses to those who were diagnosed >8 years after recruitment, indicating residual reverse causality does not seem to be substantial. Third, although a positive association for arsenic was suggestive among current smokers, it makes less sense to postulate that smoking plays a role in the impact of arsenic on PD. The arsenic levels measured in our study mostly reflected organic species (arsenobetaine) from seafood. The mechanism of arsenic in PD development warrants further exploration. Fourth, iron levels we measured were a crude estimation of iron status in humans. Besides iron content in erythrocytes, iron status also relies on serum-

based indicators, such as ferritin, transferrin saturation, and soluble transferrin receptor (19).

In conclusion, our study did not find strong evidence to support the risk of PD altered by metal exposures. Smoking may confound the association with lead and cadmium. To date, this is the first study to evaluate pre-diagnostic metal levels in blood in the development of PD, minimizing reverse causation. Further investigations are needed to gain a better understanding of the relationship between smoking, metals and PD. Furthermore, future studies of novel biomarkers of long-term metal exposure may provide more compelling evidence of the association between metals and PD.

References

1. Tysnes OB, Storstein A. Epidemiology of Parkinson's disease. J Neural Transm (Vienna). 2017;124(8):901-5.

2. Zhao Y, Ray A, Portengen L, Vermeulen R, Peters S. Metal exposure and risk of Parkinson's disease: a systematic review and meta-analysis. Am J Epidemiol. 2023.

3. Peters S, Visser A, Weisskopf M, Kioumourtzoglou M, Vermeulen R. Metals and neurodegeneration. In: Oxford textbook of neurological and neuropsychiatric epidemiology. Oxford University Press. 2020:417-24.

4. Thacker EL, O'Reilly EJ, Weisskopf MG, et al. Temporal relationship between cigarette smoking and risk of Parkinson disease. Neurology. 2007;68(10):764-8.

5. Gallo V, Brayne C, Forsgren L, et al. Parkinson's Disease Case Ascertainment in the EPIC Cohort: The NeuroEPIC4PD Study. Neurodegener Dis. 2015;15(6):331-8.

6. Gallo V, Vineis P, Cancellieri M, et al. Exploring causality of the association between smoking and Parkinson's disease. Int J Epidemiol. 2019;48(3):912-25.

7. Talhout R, Schulz T, Florek E, van Benthem J, Wester P, Opperhuizen A. Hazardous compounds in tobacco smoke. Int J Environ Res Public Health. 2011;8(2):613-28.

8. Riboli E, Hunt KJ, Slimani N, et al. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. Public health nutrition. 2002;5(6b):1113-24.

9. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. 2022.

10. Sanders T, Liu Y, Buchner V, Tchounwou PB. Neurotoxic effects and biomarkers of lead exposure: a review. Rev Environ Health. 2009;24(1):15-45.

11. Barry PS, Mossman DB. Lead concentrations in human tissues. British journal of industrial medicine. 1970;27(4):339-51.

12. Coon S, Stark A, Peterson E, et al. Whole-body lifetime occupational lead exposure and risk of Parkinson's disease. Environ Health Perspect. 2006;114(12):1872-6.

13. Weisskopf MG, Weuve J, Nie H, et al. Association of cumulative lead exposure with Parkinson's disease. Environ Health Perspect. 2010;118(11):1609-13.

14. Peters S, Broberg K, Gallo V, et al. Blood Metal Levels and Amyotrophic Lateral Sclerosis Risk: A Prospective Cohort. Ann Neurol. 2020.

15. Wang H, O'Reilly É J, Weisskopf MG, et al. Smoking and risk of amyotrophic lateral sclerosis: a pooled analysis of 5 prospective cohorts. Arch Neurol. 2011;68(2):207-13.

16. Bernhard D, Rossmann A, Wick G. Metals in cigarette smoke. IUBMB Life. 2005;57(12):805-9.

17. Thiagarajan P, Parker CJ, Prchal JT. How Do Red Blood Cells Die? Frontiers in physiology. 2021;12:655393.

18. Siderowf A, Lang AE. Premotor Parkinson's disease: concepts and definitions. Mov Disord. 2012;27(5):608-16.

19. Pfeiffer CM, Looker AC. Laboratory methodologies for indicators of iron status: strengths, limitations, and analytical challenges. Am J Clin Nutr. 2017;106(Suppl 6):1606s-14s.

Chapter 3 Supplemental materials

Text S1. PD case ascertainment in EPIC4PD

Text S2. Measurement of metal levels

Table S1. LOD and LOQ for metal measurement

Table S2. Characteristics of study participants

Table S3. Relationship between metals in blood and smoking

Table S4. Descriptive statistics of erythrocyte metal levels (ng/g) for Parkinson's cases and controls

Figure S1. Correlation matrix of metal concentrations

Figure S2. Natural spline for association between metals and risk of Parkinson's disease

Figure S3. The association between metal concentrations and the risk of Parkinson's disease by smoking status

Figure S4. The association between metal concentrations and the risk of Parkinson's disease by gender

Figure S5. The association between metal concentrations and the risk of Parkinson's disease in sensitivity analyses

Text S1. PD case ascertainment in EPIC4PD

In EPIC4PD, potential cases were first identified at each center through medical record linkage and then validated through clinical record review by movement disorder experts (1). The reliability of the diagnoses was determined by the quality of clinical data (rated as poor, good or excellent), as well as the degree of the confidence of the neurological expert (rated as low, medium or high) (1). Diagnoses were defined as 'definite' only when the confidence degree of the neurologist was high and the data quality was excellent; 'very likely' when the confidence degree was high, but data quality was either good or poor; 'probable' when the confidence degree was medium and data quality was either excellent or good; and diagnoses were defined as 'possible' in all remaining cases.

Reference:

1. Gallo V, Brayne C, Forsgren L, Barker RA, Petersson J, Hansson O, et al. Parkinson's Disease Case Ascertainment in the EPIC Cohort: The NeuroEPIC4PD Study. Neurodegener Dis. 2015;15(6):331-8.

Text S2. Measurement of metal levels

A blood sample was obtained from each participant at the baseline of the cohort. Plasma, erythrocytes and buffy coat were separated and stored in liquid nitrogen at -196 °C. Because metals are largely bound to erythrocytes in blood and the lifetime of erythrocytes is relatively long (about three months), metal concentrations in erythrocytes have been shown to be a relevant biomarker of ongoing exposure (1).

The erythrocyte samples were prepared by a direct alkali dilution method as described previously (2, 3). Metal concentrations were measured by inductively coupled plasma-mass spectrometry (ICP-MS) (Agilent 7700x, Agilent Technologies, Tokyo, Japan). Eleven elements (isotope) were assessed: arsenic (^{75}As), cadmium (^{111}Cd), calcium (^{44}Ca), copper (^{63}Cu), iron (^{56}Fe), lead (^{208}Pb), magnesium (^{24}Mg), manganese (^{55}Mn), mercury (^{202}Hg), selenium (^{78}Se) and zinc (^{66}Zn). The majority of the elements were measured in helium mode (^{75}As, ^{111}Cd, ^{63}Cu, ^{56}Fe, ^{24}Mg, ^{55}Mn, ^{66}Zn), ^{78}Se and ^{44}Ca measured in hydrogen mode, and ^{208}Pb and ^{202}Hg in standard mode (no gas). The elements germanium (^{72}Ge), rhodium (^{103}Rh), lutetium (^{175}Lu), and iridium (^{193}Ir) were included as internal standards.

The limit of detection (LOD) for each element was determined as 3×standard deviation (SD) of analyzed blanks (alkali solution) and as signal/noise=3 (Table S1). The limit of quantification (LOQ) was determined as 10×SD of analyzed blanks. As quality controls, two commercially available whole blood reference materials, Seronorm trace elements whole blood L1 and L2 (Lot no. 1702821 and 1406264, SERO, Billingstad, Norway), were analyzed. Blanks and reference materials were treated with the collected erythrocyte samples and analyzed in the beginning, middle, and end of each analysis.

In one sample, the mercury level was below the LOD and was replaced by LOD/$\sqrt{2}$. No metal values were between each respective LOD and LOQ.

References:

1. Nordberg G, Fowler B, Nordberg M. Handbook on the toxicology of metals. 4th ed. Cambridge, MA: Academic Press. 2014.

2. Peters S, Broberg K, Gallo V, Levi M, Kippler M, Vineis P, et al. Blood Metal Levels and Amyotrophic Lateral Sclerosis Risk: A Prospective Cohort. Ann Neurol. 2020.

3. Levi M, Hjelm C, Harari F, Vahter M. ICP-MS measurement of toxic and essential elements in human breast milk. A comparison of alkali dilution and acid digestion sample preparation methods. Clinical biochemistry. 2018;53:81-7.

Table S1. LOD and LOQ for metal measurement

Metal	Average LOD, ng/g	Average LOQ, ng/g
Arsenic	0.00279	0.01634
Cadmium	0.00171	0.00645
Calcium	5.29543	19.45424
Copper	0.08867	0.44666
Iron	13.18192	82.89144
Lead	0.01067	0.04123
Magnesium	2.20312	12.90307
Manganese	0.02933	0.05362
Mercury	0.00697	0.02386
Selenium	0.00575	0.02166
Zinc	4.07954	7.69266

LOD, limit of detection; LOQ, limit of quantification.

Table S2. Characteristics of study participants

Characteristic	PD cases n=362	Controls n=362
Age at recruitment[1], median (IQR)	60.9 (55.0–65.8)	60.6 (55.1–65.8)
Age at PD diagnosis, median (IQR)	68.7 (62.9–74.1)	–
Years between recruitment and PD diagnosis, median (IQR)	7.8 (4.5–11.0)	–
Reliability of diagnosis, n (%)		
Definite	46 (13%)	–
Very likely	147 (40%)	–
Probable	60 (17%)	–
Possible	109 (30%)	–
Sex[1], n (%)		
Male	203 (56%)	203 (56%)
Female	159 (44%)	159 (44%)
Country[1], n (%)		
Italy	58 (16%)	58 (16%)
Spain	97 (27%)	97 (27%)
UK	148 (40%)	148 (40%)
Netherlands	13 (4%)	13 (4%)
Germany	46 (13%)	46 (13%)
Smoking status at recruitment[2], n (%)		
Never smokers	183 (51%)	179 (49%)
Former smokers	121 (33%)	113 (31%)
Current smokers	48 (13%)	57 (16%)
Coffee consumption at recruitment[3] (ml/d)		
Nonconsumer, n (%)	49 (14%)	39 (11%)
Daily coffee, median (IQR)	190 (76–475)	190 (75–475)
Alcohol consumption at recruitment[4] (g/d)		
Nonconsumer, n (%)	78 (22%)	66 (18%)
Total alcohol, median (IQR)	9.3 (3.0–24.1)	10.5 (2.8–26.2)
BMI at recruitment (kg/m^2), median (IQR)	26.5 (24.1–29.2)	26.1 (23.9–29.0)
Higher education[5], n (%)	42 (12%)	47 (13%)
Physically active[6], n (%)	18 (4%)	27 (3%)

PD, Parkinson's disease; IQR, interquartile range; BMI, body mass index.

[1] Matching variables for PD case and control

[2] Smoking status was missing for 10 PD cases and 13 controls.

[3] Coffee consumption was missing for 3 PD cases and 4 controls.

[4] Alcohol consumption was missing for 3 PD cases and 4 controls.

[5] Education was missing for 26 PD cases and 17 controls.

[6] Combined index of occupational and recreational physical activity. Physically activity was missing for 13 PD cases and 10 controls.

Table S3. Relationship between metals in blood and smoking

Metal	Metal levels (ng/g)[1]			Correlation with smoking intensity[2]
	Current smokers n=105	Former smokers n=234	Never smokers n=362	
Arsenic	3.62 (3.05)	3.58 (2.74)	3.22 (2.97)	-0.19
Cadmium[3]	1.22 (2.03)	0.50 (1.62)	0.48 (1.61)	0.50
Calcium	10,678 (1.60)	12,142 (1.60)	11,262 (1.55)	-0.10
Copper	606 (1.24)	601 (1.20)	606 (1.22)	-0.04
Iron	873,466 (1.17)	861,206 (1.16)	853,763 (1.18)	-0.18
Lead[3]	92.0 (1.72)	71.5 (1.68)	67.2 (1.78)	0.26
Magnesium	42,384 (1.21)	41,358 (1.18)	41,098 (1.19)	-0.17
Manganese	13.3 (1.42)	13.6 (1.43)	14.41 (1.39)	-0.03
Mercury	4.70 (2.97)	3.82 (2.60)	4.09 (3.31)	-0.22
Selenium	125 (1.31)	123 (1.29)	128 (1.32)	-0.18
Zinc	9,732 (1.25)	9,804 (1.21)	9,594 (1.24)	-0.23

[1] Expressed as geometric mean (geometric standard deviation)

[2] Spearman correlation coefficients between metal levels and cigarettes consumed per day, which was only available for 78 current smokers

[3] From Dunn's test, metal levels among current smokers were significantly higher than these in former and never smokers ($p<0.05$).

Table S4. Descriptive statistics of erythrocyte metal levels (ng/g) for Parkinson's cases and controls

Metal	PD cases, n=362				Controls, n=362			
	AM	GM	GSD	Min–Max	AM	GM	GSD	Min–Max
Arsenic	5.57	3.34	2.90	0.14–42.8	5.98	3.51	2.90	0.21–54.5
Cadmium	0.67	0.54	1.86	0.14–4.76	0.72	0.58	1.83	0.18–5.60
Calcium	13,023	11,683	1.59	2,323–55,017	12,775	11,452	1.59	3,355–37,116
Copper	617	605	1.21	321–1,706	619	606	1.22	348–2,002
Iron	868,100	858,028	1.17	332,103–1316,044	871,939	861,553	1.17	500,487–1315,686
Lead	81.9	69.8	1.74	14.5–562	87.0	73.8	1.17	14.0–333.9
Magnesium	42,000	41,408	1.19	19,239–60,143	42,071	41,412	1.20	22,393–64,359
Manganese	14.9	13.9	1.42	2.79–42.4	14.9	14.1	1.39	3.87–36.9
Mercury[1]	6.92	4.07	3.05	0.05–37.7	6.88	4.09	2.85	0.08–69.5
Selenium	130	125	1.30	62.7–341	130	125	1.32	57.0–322
Zinc	9,910	9,714	1.22	4,117–18,817	9,855	9,632	1.24	4,869–16,526

PD, Parkinson's disease; AM, arithmetic mean; GM, geometric mean; GSD, geometric standard deviation.

[1] One value for mercury was below limit of detection.

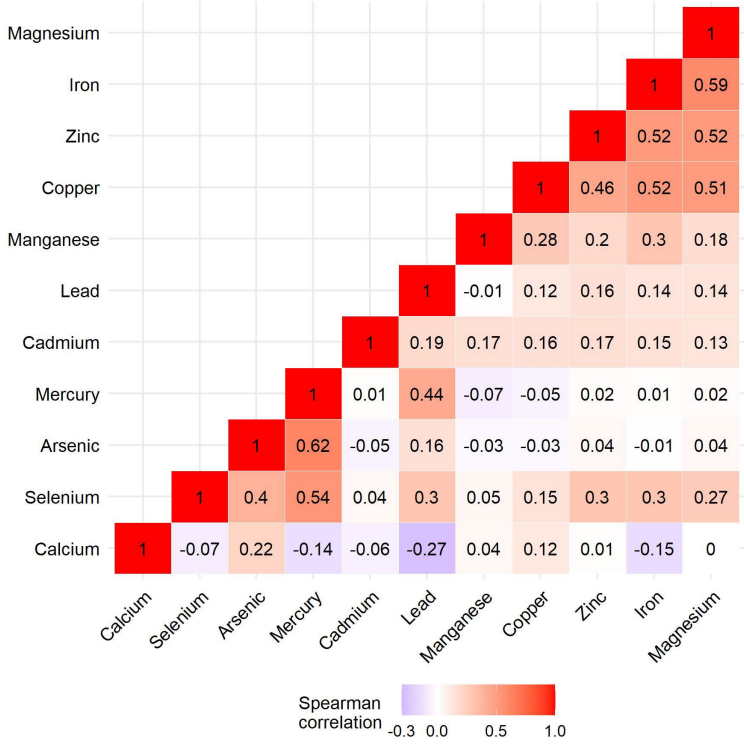
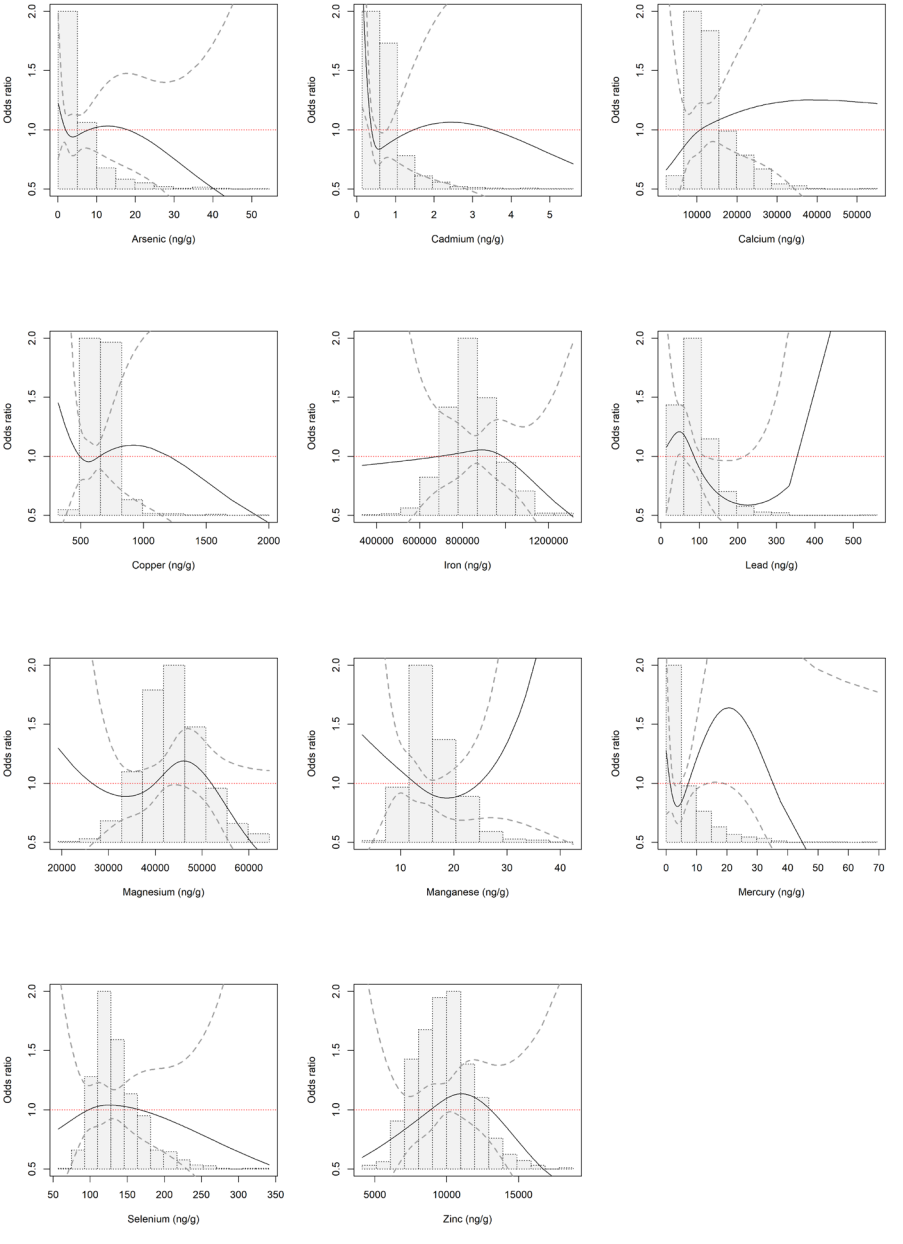


Figure S1. Correlation matrix of metal concentrations

Values in tiles were Spearman correlation coefficients.

Figure S2. Natural spline for association between metals and risk of Parkinson's disease

Histogram shows the metal distribution; solid line represents odds ratios, estimated by conditional logistic regression of natural spline with 3 degrees of freedom on metal levels; dashed lines are the 95% confidence intervals; Y-axis is truncated at 0.5 and 2.0 for all metals; odds ratios are expressed relative to minimum levels of metal.



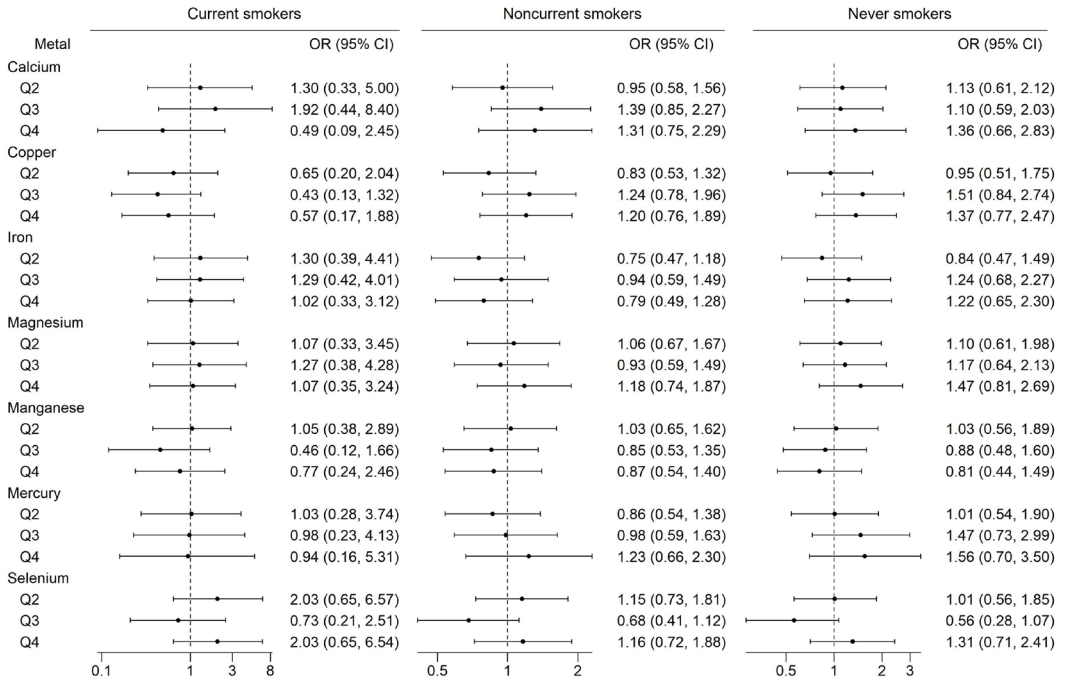


Figure S3. The association between metal concentrations and the risk of Parkinson's disease by smoking status

OR, odds ratio; CI, confidence interval, Q, quartile.

ORs were calculated by unconditional logistic regression, adjusted by age at recruitment, sex and country, quartile 1 as reference. Noncurrent smokers included former and never smokers.

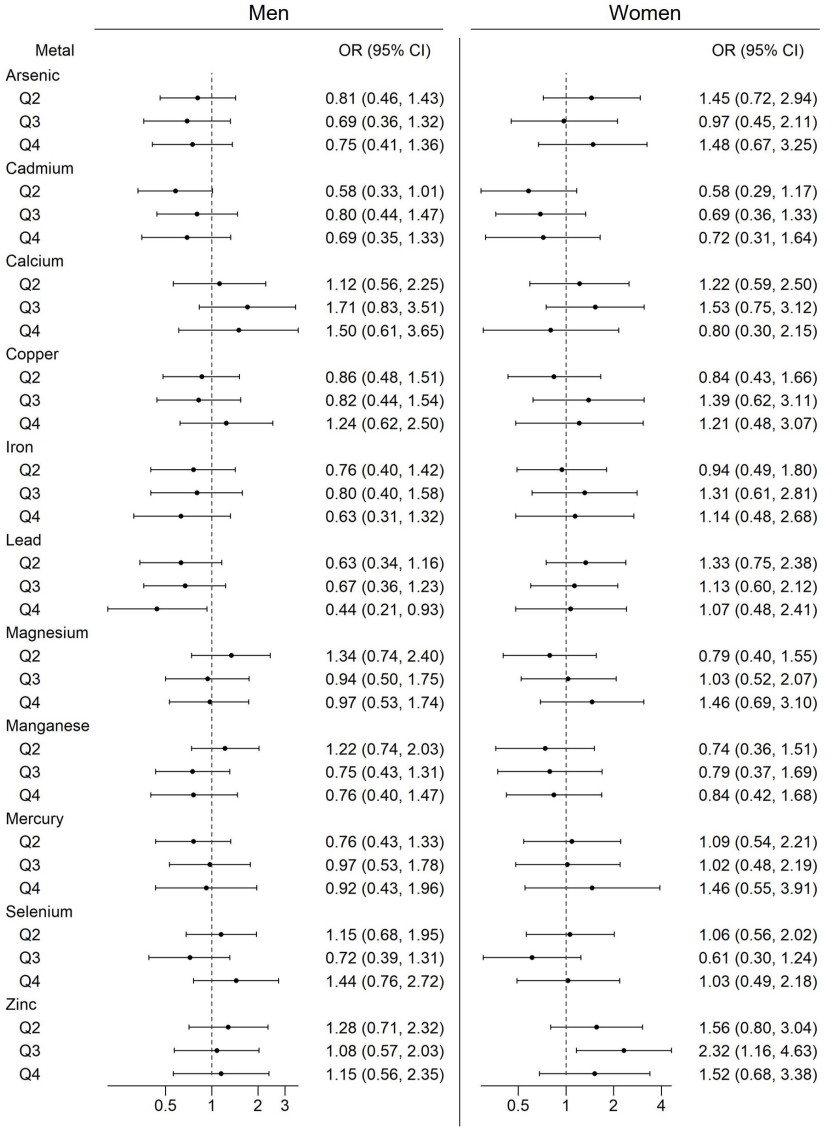


Figure S4. The association between metal concentrations and the risk of Parkinson's disease by gender

OR, odds ratio; CI, confidence interval, Q, quartile.

ORs were calculated by conditional logistic regression for the matched case-control sets, adjusted by smoking status.

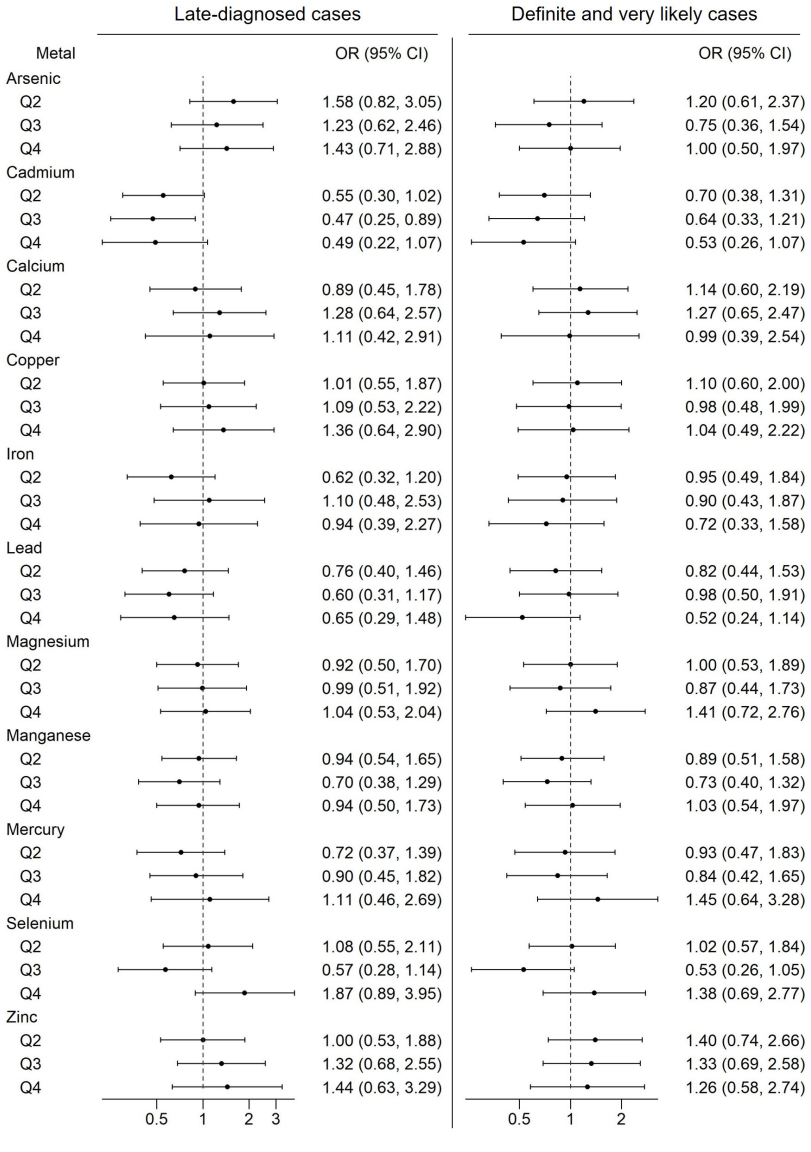


Figure S5. The association between metal concentrations and the risk of Parkinson's disease in sensitivity analyses

OR, odds ratio; CI, confidence interval, Q, quartile.

ORs were calculated by conditional logistic regression for the matched case-control sets, adjusted by smoking status. Late–diagnosed cases were those diagnosed as Parkinson's disease after 8 years since recruitment.

Chapter 4

Dioxin(-like) Related Biological Effects Through Integrated Chemical-wide and Metabolome-wide Analyses

Yujia Zhao, Jeroen Meijer, Douglas I. Walker, Juni Kim, Lützen Portengen, Dean P. Jones, Fatemeh Saberi Hosnijeh, Jelle Vlaanderen, Roel Vermeulen

Corresponding author: Roel Vermeulen

Published on *Environmental Science & Technology*, December 2023

https://pubmed.ncbi.nlm.nih.gov/38149779/

Abstract

Dioxin(-like) exposures are linked to adverse health effects, including cancer. However, metabolic alterations induced by these chemicals remain largely unknown. Beyond known dioxin(-like) compounds, we leveraged a chemical-wide approach to assess chlorinated co-exposures and parent compound products (termed dioxin(-like) related compounds) among 137 occupational workers. Endogenous metabolites were profiled by untargeted metabolomics, namely reversed-phase chromatography with negative electrospray ionization (C18-negative) and hydrophilic interaction liquid chromatography with positive electrospray ionization (HILIC-positive). We performed a metabolome-wide association study to select dioxin(-like) associated metabolic features using a 20% false discovery rate threshold. Metabolic features were then characterized by pathway enrichment analyses. No significant features associated with polychlorinated dibenzo-*p*-dioxins (PCDDs), a subgroup of known dioxin(-like) compounds. However, 3,110 C18-negative and 2,894 HILIC-positive features were associated with at least one of the PCDD-related compounds. Abundant metabolic changes were also observed for polychlorinated dibenzofurans (PCDF)-related and polychlorinated biphenyls (PCB)-related compounds. These metabolic features were primarily enriched in pathways of amino acids, lipids and fatty acids, carbohydrates, cofactors and nucleotides. Our study highlights the potential of chemical-wide analysis for comprehensive exposure assessment beyond targeted chemicals. Coupled with advanced endogenous metabolomics, this approach allows for in-depth exploration of metabolic alterations induced by environmental chemicals.

Keywords: Dioxin(-like) exposures, chemical-wide association study, metabolome-wide association study, occupational population, biological pathways, exposome

Introduction

Dioxin(-like) compounds rank among the most notorious anthropogenic environmental toxicants and have been extensively studied over the past four decades (1). This chemical category includes three structurally related subclasses: polychlorinated dibenzo-*p*-dioxins (PCDDs), dioxin-like polychlorinated dibenzofurans (PCDFs) and dioxin-like polychlorinated biphenyls (PCBs) (2). The risk assessment of dioxin(-like) compounds, like many other exposures, primarily focuses on individual chemicals, particularly the most toxic chemical, 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD). TCDD is classified as a "known human carcinogen" and associated with an increased risk of all cancers combined (3). Furthermore, TCDD has been implicated in toxicities concerning the immune, nervous, endocrine and reproductive systems (4).

A one-by-one assessment of the biological impact of chemicals may overlook important biological perturbations. In a metabolome-wide association study (MWAS) by Walker et al. on trichloroethylene (TCE), it was shown that most of the observed biological effects associated stronger with unknown metabolic products of TCE, as opposed to with TCE itself or prior known metabolites (5). This challenges the conventional practice of assessing chemical toxicity by focusing solely on parent compounds and known metabolites (6). An alternative strategy could be to first comprehensively map exposures to known compounds, co-exposures (e.g., unrecognized chemicals with analogous properties) and their metabolites, followed by associating these with biological changes. This integrated chemical-wide and metabolome-wide approach could yield a more comprehensive evaluation of biological effects.

We illustrate here an example of a chemical-wide and metabolome-wide investigation (Figure 1) through i) an exhaustive targeted analysis of dioxin(-like) compounds, ii) connecting these targeted dioxin(-like) compounds to associated chlorinated compounds characterized using untargeted gas chromatography with high-resolution mass spectrometry (GC-HRMS), thus encompassing a thorough representation of dioxin(-like) exposures, and iii) linking targeted and related dioxin(-like) compounds with biological changes assessed through metabolomics and targeted immunological phenotyping.

For these research goals, we used a highly unique sub-population of the Dutch herbicide cohort, recognized as one of the most informative epidemiological studies in dioxin research (7). The cohort comprised workers of two chlorophenoxy herbicide producing factories (8, 9). One factory (factory A) experienced high TCDD exposure due to a reactor vessel explosion in 1963. Even decades later, TCDD levels in blood of ex-factory A workers remained substantially higher than the general population (4ppt vs. below

detection limit). Likewise, levels of dioxin-like PCDFs and PCBs generally exceeded background levels as reported in monitoring data (Table S1). This occupational cohort provides a distinctive opportunity to investigate health and biological effects associated with dioxin(-like) exposures.

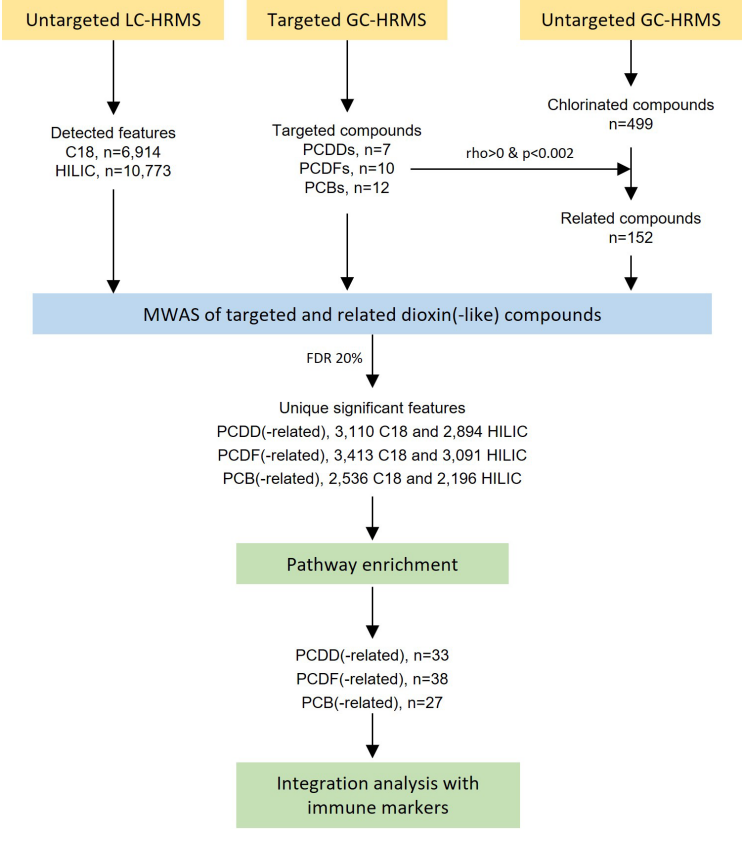


Figure 1. Workflow of the chemical-wide and metabolome-wide association analyses

Abbreviation: LC-HRMS, liquid chromatography with orbitrap high-resolution mass spectrometry; GC-HRMS, gas chromatography with high-resolution-mass-spectrometry; C18, C18-negative mode; HILIC, HILIC-positive mode; MWAS, metabolome-wide association study; FDR, false discovery rate.

Methods

Study population

The subjects involved in this study were drawn from the Dutch herbicide cohort. Details have been described elsewhere (8-10). Briefly, this cohort comprised workers from two factories (denoted as factory A and factory B) engaged in manufacturing chlorophenoxy herbicides during the 1950s-1980s in the Netherlands. Factory A's primary products were 2,4,5-trichlorophenoxyacetic acid (2,4,5-T) and 2,4,5-trichlorophenol (2,4,5-TCP), which included potential contamination with TCDD and other PCDDs. In March 1963, an explosion occurred within an autoclave for the synthesis of 2,4,5-TCP, releasing its contents, including TCDD. Individuals working in production departments at factory A or present during the accident and subsequent clean-up were exposed to high levels of TCDD. Conversely, Factory B produced 4-chloro-2-methylphenoxy propanoic acid (MCPA), 4-chloro-2-methylphenoxy propanoic acid (MCPP), and in smaller amounts, 2,4-dichlorophenoxyacetic acid (2,4-D). While potential by-products in factory B included PCDDs (mainly with 2 to 3 chlorine atoms) and dioxin-like PCDFs and PCBs, the presence of TCDD was unlikely.

Workers ever employed in factory A (n=1,167) and factory B (n=1,143) were enrolled in the cohort. During the third follow-up period (2007-2008), participants were selected for blood collection based on a stratified sampling strategy that considered their exposure status to chlorophenoxy herbicides, chlorophenols, and associated contaminants. The study enrolled 82 workers from factory A, half of whom had worked within production departments or participated in accident-related cleaning-up, alongside a randomly selected sample of 70 workers from factory B. All study subjects were male and completed a questionnaire covering basic information, anthropometric parameters (height and weight) and lifestyle habits (smoking status and alcohol consumption). Plasma was separated and stored at -80 °C.

Exposure assessment for dioxin(-like) compounds

Measurement of targeted dioxin(-like) compounds

Previously identified dioxin(-like) compounds in plasma were quantified by targeted gas chromatography with high-resolution mass spectrometry (GC-HRMS) in the Centers for Disease Control and Prevention, USA. Targeted dioxin(-like) compounds covered seven PCDDs (TCDD, 12378D, 123478D, 123678D, 123789D, OCDD), ten dioxin-like PCDFs (2378F, 12378F, 23478F, 123478F, 123678F, 123789F, 234678F, 1234678F, 1234789F, OCDF), twelve dioxin-like PCBs (PCB77, PCB81, PCB126, PCB169, PCB105, PCB114, PCB118, PCB123, PCB156, PCB157, PCB167, PCB189) (Table S1). Concentrations of these targeted dioxin(-like) compounds were adjusted for total lipids and reported as

parts per trillion (ppt). Values below detection limits were imputed using a maximum likelihood method (11).

Given TCDD's protracted half-life in humans, its presence in biofluids persists for decades post initial exposure. We previously developed a predictive model to back-extrapolate plasma TCDD levels at time of last exposure ($TCDD_{max}$). This prediction integrated measured TCDD levels within a one-compartment first-order kinetic model, where TCDD's half-life was established at 7.1 years ($t_{1/2}$) (12).

$$TCDD_{max} = background + (measured\ TCDD - background) * exp(ln(2) * lag / t_{1/2})$$

Lag periods for factory A workers were determined by their occupational history and involvement in the clean-up following the 1963 accident (12). Factory B workers were not assigned a lag period; the measured TCDD levels were taken as the $TCDD_{max}$. The average TCDD concentration detected in factory B served as the background level in the model.

Measurement of dioxin(-like) related compounds

We characterized all possible dioxin(-like) exposures using untargeted GC-HRMS in Rollins School of Public Health, Emory University, USA. Plasma samples were prepared and analyzed using methods described previously (13). Plasma samples were extracted using 4:1 hexane:ethyl acetate and analyzed in duplicate using a Themo Scientific 1310 gas chromatograph connected to a Thermo Scientific Q Exactive GC Orbitrap GC-MS/MS mass spectrometer. The GC-HRMS was operated in full-scan mode over a mass-to-charge (*m/z*) range of 85-850 and 60,000 resolution. Uniquely detected metabolic features consisting of *m/z*, retention time and ion abundance were extracted and aligned using extensible computational mass spectrometry (XCMS) software (14). To identify unique mass spectra, we performed data-driven clustering algorithm using *RamClustR* (15), which aggregates feature intensities based on correlation and retention-time grouping, and provides a weight-averaged intensity for each group of features corresponding to an individual compound. After *m/z* clustering, 11,004 unique mass spectra, referred to as chemical features, were identified from the untargeted GC-HRMS data.

To identify additional compounds related to targeted dioxin(-like) compounds, mass spectra corresponding to each chemical feature were evaluated for chlorinated isotopic patterns by linking monoisotopic masses to M+2, M+4, M+6 and M+8 isotopic envelopes using the R package *nontarget* (16). Compounds showing a significant positive correlation (Spearman's rank correlation coefficient >0 with a *p*-value below 0.002; corresponding to a 20% false discovery threshold) with any of the 29 targeted dioxin(-like) compounds were designated as dioxin(-like) related compounds. This

criterion was chosen to mitigate the impact of false positives from multiple testing (Text S1). The relationship among targeted dioxin(-like) compounds and those identified as dioxin(-like) related were depicted by a correlation-based network. Node clustering was identified using a multilevel community detection algorithm implemented in the *igraph* package (17). Finally, the network and clustering were visualized using Cytoscape software (18).

Throughout this paper, chlorinated compounds that were correlated with at least one targeted PCDD are called 'PCDD-related compounds'. Similarly, chlorinated compounds correlated with at least one PCDF or PCB are called 'PCDF-related compounds' or 'PCB-related compounds', respectively.

High-resolution metabolomics

Untargeted metabolomic profiling in plasma was conducted using liquid chromatography coupled with Orbitrap high-resolution mass spectrometry (LC-HRMS) (Dionex Ultimate 3000, Q-Exactive HF, Thermo Scientific) in Emory University, as previously described (19). Two complimentary LC columns were used to maximize coverage, including reversed-phase with negative electrospray ionization (C18-negative) and hydrophilic interaction liquid chromatography with positive electrospray ionization (HILIC-positive) (20). Plasma samples were processed by adding two volumes of acetonitrile to precipitate proteins, and triplicate analyses were conducted in each mode. The HRMS was operated in full scan mode at 120,000 resolution over a *m/z* range 85-1,275. Raw data files were extracted and aligned using *apLCMS* (21) with modifications by *xMSanalyzer* (22). In total, 10,477 and 16,605 metabolite features were detected for C18-negative and HILIC-positive mode, respectively. Before data analysis, metabolite features were batch-corrected using *ComBat* (23) and averaged, followed by removing features with coefficient of variation among technical replicates $\geqslant$100% and detected in <60% of the study subjects. Remaining missing values were imputed using a left-censored quantile regression approach, implemented in *imputeLCMD* (24). After imputation, 6,914 C18-negative and 10,773 HILIC-positive LC-HRMS features were retained for subsequent analyses.

MWAS of dioxin(-like) exposures

Targeted and related dioxin(-like) compounds and metabolic features were naturally log-transformed for analyses. In the MWAS, we used the linear regression framework as implemented in *Omics* R package (25), by regressing metabolic features one-by-one on a specific exposure compound (either a known or related dioxin(-like) compound). These models were adjusted for age (continuous variable), factory (categorical variable), body mass index (BMI; kg/m^2, continuous variable) (model 1). To account for multiple comparisons, the Benjamini-Hochberg (BH) procedure (26) was applied for the MWAS

of each exposure, and a false discovery rate (FDR) threshold of 20% was adopted to identify metabolite features associated with the exposure. Separate FDR procedures were applied to metabolic features detected by the C18-negative and HILIC-positive modes.

In this study, dioxin(-like) exposures were primarily originated from occupational activities and minimally associated with lifestyle factors such as smoking and alcohol consumption. Moreover, we conducted an expanded analysis to incorporate smoking status and alcohol intake as additional covariates (model 2). However, corresponding exposure-feature coefficients in model 2 exhibited minimal deviations, less than 3% on average compared to those in model 1 (data not shown). Consequently, smoking and alcohol consumption were less likely to act as confounders in this study, leading us to select model 1 for subsequent analyses.

We categorized metabolic features into three subclasses based on their associations with PCDD(-related) compounds, PCDF(-related) compounds, and PCB(-related) compounds under FDR 20%. Subsequently, separate pathway enrichments were performed for each subclass of metabolic features.

Biological pathway enrichment and metabolite annotation

To characterize metabolic features associated with dioxin(-like) exposures, we first matched the significant features to an internal compound database, confirmed by authentic reference standards, denoting a level 1 confidence (27). Features without matches with authentic standards were annotated using *xMSannotator* (28). This tool categorizes annotations into different confidence tiers using a multistage clustering algorithm, based on database matches. We searched against Human Metabolome Database (HMDB) (29) with mass tolerance of ±5 ppm and retention time tolerance of ±5 seconds. The adduct were ‘M+H’, ‘M+2H’, ‘M+ACH+2H’, ‘M+Na’, ‘M+ACN+H’, ‘M+ACN+Na’, ‘2M+H’, ‘M+H+H2O’ for positive mode, "M-H","M-H2O-H","M+Na-2H","M+Cl","M+Hac-H","2M-H" for negative mode. Annotations suggested with high confidence level by *xMSannotator* were presented as level 4 confidence annotations according to Schymanski et al (27).

Metabolic pathways associated with dioxin(-like) exposures were identified using Mummichog (version 1.0.10) with a mass tolerance of ±5 ppm (30). Associated metabolic pathways were identified using a pathway significance threshold <0.05 as well as the presence of at least four metabolites associated with the exposures.

Integration of metabolic pathways and immune phenotypic measures

To explore the potential mechanisms underlying dioxin(-like) exposure toxicity, we used a network-based integration approach to evaluate the relationship between the

metabolic pathways associated with dioxin(-like) exposures and immune phenotypic endpoints. Immune markers, including cytokines and growth factors (n=21), hematologic parameters (i.e., cell counts) (n=23), humoral immunity markers (immunoglobulins (Ig) and complement factors (C)) (n=7), lymphoma makers (n=3), were previously measured for factory A workers (31-34) (Table S3).

For factory A workers, principal component analysis (PCA) was conducted on intensities of significant metabolic features within each enriched Mummichog-identified pathway. Subsequently, first principal component (PC1) scores were computed as a summary measure for each respective pathway. These PC1 scores, representing metabolic pathways, were subjected to partial least squares (PLS) regression via the *xMWAS* package (35) to explore potential associations with all immune markers. In PLS regression, the association score between variables from two matrices approximates their correlation coefficient, determined by PLS components and regression coefficients (36). The resultant pairwise associations, marked by an |association score| exceeding 0.3 and *p*-value below 0.05, were used to build a network to visualize the connections between the pathways and biomarkers. A multilevel community detection method was applied to uncover clustering of pathways and biomarkers (37). The network and identified communities were visualized using Cytoscape (18).

Results

Study population and dioxin(-like) exposures

After excluding workers with a diagnosis of cancer (except for skin cancer) (6 from factory A and 9 from factory B), 76 workers from factory A and 61 workers from factory B were retained in the analyses. Workers in factory A were older compared to those in factory B (69.0 vs. 58.8 years, $p<0.001$) (Table 1). No significant differences were observed in BMI, alcohol intake and smoking status between the two factories.

As expected, the concentrations of PCDDs were markedly higher in factory A workers compared to factory B workers (all $p<0.05$) (Table S1). Notably, the difference in TCDD levels between the two factories was substantial (median 4.35 vs. 0.30 ppt, $p<0.0001$). Levels of dioxin-like PCDFs and PCBs were comparable across both factories. We observed moderate correlations among PCDDs, and some high correlations among PCBs ($r_s>0.9$), while most PCDFs were only weakly correlated (Figure S3).

Of the 499 suspected chlorinated compounds detected by untargeted GC-HRMS, 152 were identified as possible dioxin(-like) related compounds. Specifically, 109 chlorinated compounds correlated to at least one PCDD, 136 to at least one PCDF, and 58 to at least one PCB (Table S4). Because of the correlated property of targeted compounds, the overlap among three categories of dioxin(-like) related compounds

was considerable (Figure S4). Our analysis of the network integrating targeted and related dioxin(-like) compounds identified six densely interconnected communities (Figure 2, Table S5). All PCBs except PCB169 were grouped in Community 1, while five PCDDs were in Community 2. TCDD was clustered with 23478F and 234678F in Community 3.

MWAS of dioxin(-like) exposures

The numbers of metabolic features significantly associated with each targeted and its related dioxin(-like) compounds at a 20% FDR threshold are given in Table 2. While no feature was significantly associated with any targeted PCDD, we found that the PCDD-related compounds contributed substantially to metabolic alterations. Specifically, 3,110 C18-negative and 2,894 HILIC-positive features were found to be associated with at least one of the PCDD-related compounds. This phenomenon of enriched metabolic changes is held for PCDFs and PCBs, alongside their corresponding related compounds. Predicted maximum levels of TCDD were not associated with any metabolic feature (Figure S5).

Table 1. Characteristics of study participants

	Factory A (n=76)	Factory B (n=61)	p-value[a]
Age (years), mean (SD)	69.0 (7.7)	58.8 (9.0)	<0.001
Body mass index (kg/m^2)[b], mean (SD)	26.9 (3.0)	27.1 (3.6)	0.726
Alcohol intake (units/week), mean (SD)	13.2 (13.6)	13.7 (15.1)	0.835
Smoking status, n (%)			0.594
Never-smokers	12 (15.8%)	13 (21.3%)	
Former smokers	46 (60.5%)	32 (52.5%)	
Current smokers	18 (23.7%)	16 (26.2%)	

Abbreviations: SD, standard deviation.

[a] p values from t-test for continuous variables and chi-square test for categorical variable, subjects from factory A vs. factory B.

[b] Body mass index (BMI) was calculated as the weight in kilograms divided by the square of the height in meters.

Table 2. The number of metabolic features significantly associated with dioxin(-like) exposures

Congener	C18-negative features[a]			HILIC-positive features[a]		
	Targeted compound	Related compounds[b]	Total[c]	Targeted compound	Related compounds[b]	Total[c]
PCDDs						
TCDD	0	2528	2528	0	776	776
12378D	0	1789	1789	0	998	998
123478D	0	2147	2147	0	2163	2163
123678D	0	2058	2058	0	1099	1099
123789D	0	1010	1010	0	1604	1604
1234678D	0	125	125	0	46	46
OCDD	0	413	413	0	158	158
Total[c]	0	3110	3110	0	2894	2894
Dioxin-like PCDFs						
2378F	2	NA	2	5	NA	5
12378F	0	267	267	0	3	3
23478F	0	3240	3240	0	2839	2839
123478F	0	1572	1572	0	682	682
123678F	1	902	903	0	156	156
123789F	0	0	0	0	0	0
234678F	0	1688	1688	1	727	727
1234678F	0	4	4	0	302	302
1234789F	0	NA	0	0	NA	0
OCDF	1	NA	1	1	NA	1
Total[c]	4	3411	3413	7	3086	3091
Dioxin-like PCBs						
PCB77	0	NA	0	143	NA	143
PCB81	0	NA	0	0	NA	0
PCB126	237	503	692	42	77	112
PCB169	0	2153	2153	0	2053	2053
PCB105	151	1182	1277	12	124	127
PCB114	0	1314	1314	0	740	740
PCB118	319	1182	1360	6	124	126
PCB123	100	292	360	9	67	71
PCB156	0	1270	1270	0	182	182
PCB157	0	1270	1270	0	182	182
PCB167	346	1540	1662	0	195	195
PCB189	0	1272	1272	0	183	183
Total[c]	507	2347	2536	192	2057	2196

Abbreviation: NA, not applicable (due to no chlorinated compound identified as related compound for the corresponding targeted dioxin(-like) compound).

[a] The number of features associated with targeted or related dioxin(-like) compound under FDR 20%, adjusted by age, BMI and factory.

[b] Features significantly associated with at least one related compound, which were significantly correlated with the specific targeted dioxin(-like) compound.

[c] Total number of unique features.

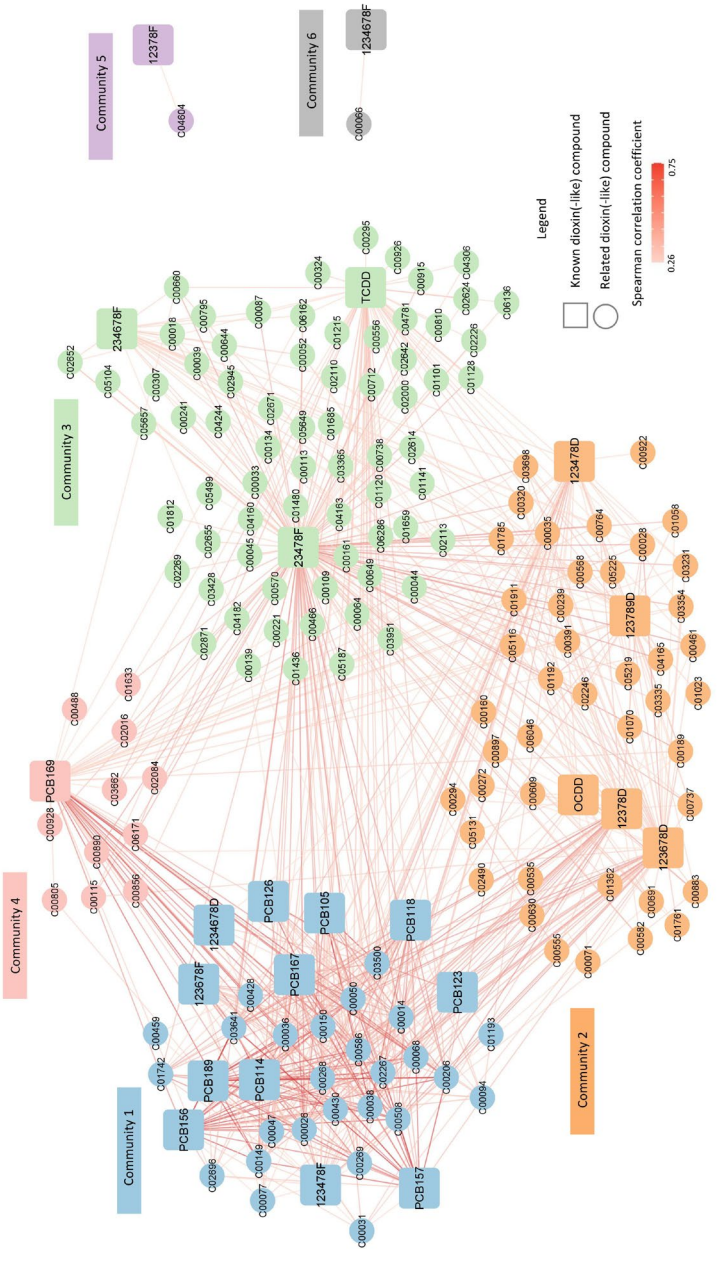


Figure 2. Network of targeted and related dioxin(-like) compounds

Network was generated based upon Spearman correlations between targeted dioxin(-like) compounds and their related compounds. Community detection was used to identify closely related nodes, which were indicated by node color. Correlation magnitude was indicated by edge color.

Annotations of metabolic features in response to dioxin(-like) exposures

Among the metabolic features showing significant associations with dioxin(-like) exposures, level 1 annotations included 21 amino acids, 9 fatty acids, 4 cofactors, D-glucose, cholesterol, uric acid and xanthine (Table S7). Further annotations at level 4 confidence included 7 androstane steroids and 8 metabolites from glycerolipids and glycerophospholipids.

In terms of metabolic pathway annotations, there were 33 pathways enriched from the metabolic features linked with PCDD(-related) compounds, 38 pathways with PCDF(-related) compounds, and 27 pathways with PCB(-related) compounds (Table 3). Among these, 21 pathways showed shared enrichment across all three subclasses, including 7 lipid pathways (de novo fatty acid biosynthesis, fatty acid activation, fatty acid metabolism, linoleate metabolism, phytanic acid peroxisomal oxidation, omega-3 fatty acid metabolism, phosphatidylinositol phosphate metabolism), 6 amino acid pathways (alanine and aspartate, arginine and proline, aspartate and asparagine, histidine, lysine, urea cycle/amino group metabolism), 3 carbohydrate pathways (aminosugar, butanoate, pentose and glucoronate interconversion), purine metabolism, vitamin B6 metabolism, glutathione metabolism, drug metabolism, xenobiotics metabolism.

Integration of metabolic pathways with immune phenotypic endpoints

The integration analysis involving metabolic pathways and immune endpoints was conducted only among factory A workers. Since the pathways largely overlapped for the three subclasses of dioxin(-like) exposures, integration analysis was first done involving pathways associated with all exposures, then moved on to analyses including pathways associated with PCDD(-related), PCDF(-related) and PCB(-related) compounds, separately.

In the network encompassing all pathways, every pathway (represented by PC1 scores) was associated with at least one immune marker (Figure 3). Community detection revealed the presence of three communities. Community 1 consisted of subsets of T and B lymphocytes, alongside complement factors (C3, C4) and a lymphoma marker, soluble B-cell activation marker 27 (sCD27). Pathways associated with this community mainly included various amino acid pathways, cofactor metabolism pathway (vitamin B3, B6, porphyrin), xenobiotics metabolism pathways, as well as pathways of purine and tricarboxylic acid (TCA) cycle.

Community 2 predominantly included various cytokines and growth factors (mainly on interleukins), hematologic parameters (red blood cells, hemoglobin, hematocrit, monocytes, B cells, T helper cells) and lymphoma markers (soluble CD30 (sCD30) and interleukin 1 receptor antagonist (IL1RA)). This community also exhibited associations

Table 3. Enriched biological pathways associated with dioxin(-like) exposures

Pathway	Overlap size[a]	Pathway size[b]	%[c]	PCDD(-related)		PCDF(-related)		PCB(-related)	
				C18	HILIC	C18	HILIC	C18	HILIC
Lipid metabolism, n=13									
Carnitine shuttle	12	14	86						
De novo fatty acid biosynthesis	17	17	100						
Fatty acid activation	15	15	100						
Fatty acid metabolism	9	14	64						
Glycerophospholipid metabolism	26	35	74						
Glycosphingolipid biosynthesis - globoseries	5	5	100						
Glycosphingolipid metabolism	19	25	76						
Leukotriene metabolism	28	34	82						
Linoleate metabolism	16	19	84						
Omega-3 fatty acid metabolism	6	6	100						
Phosphatidylinositol phosphate metabolism	14	24	58						
Phytanic acid peroxisomal oxidation	7	7	100						
Squalene and cholesterol biosynthesis	19	24	79						
Amino acid metabolism, n=12									
Alanine and aspartate metabolism	14	20	70						
Arginine and proline metabolism	22	32	69						
Aspartate and asparagine metabolism	37	57	65						
Beta-alanine metabolism	9	12	75						
Glutamate metabolism	9	12	75						
Glycine, serine, alanine and threonine metabolism	24	40	60						
Histidine metabolism	13	19	68						
Lysine metabolism	18	21	86						
Methionine and cysteine metabolism	21	41	51						
Tyrosine metabolism	40	80	50						
Urea cycle/amino group metabolism	22	39	56						
Valine, leucine and isoleucine degradation	20	24	83						

Carbohydrate metabolism, n=8				
Aminosugars metabolism	16	28	57	
Ascorbate and aldarate metabolism	8	14	57	
Butanoate metabolism	21	23	91	
Pentose and glucuronate Interconversions	13	14	93	
Pentose phosphate pathway	25	33	76	
Propanoate metabolism	12	15	80	
Pyruvate metabolism	13	16	81	
TCA cycle	10	15	67	
Metabolism of cofactors and vitamins, n=5				
Porphyrin metabolism	9	14	64	
Vitamin B1 metabolism	8	8	100	
Vitamin B3 metabolism	11	18	61	
Vitamin B6 metabolism	5	7	71	
Vitamin B9 metabolism	7	8	88	
Nucleotide metabolism, n=2				
Purine metabolism	38	51	75	
Pyrimidine metabolism	39	47	83	
Xenobiotics biodegradation and metabolism, n=2				
Drug metabolism - other enzymes	12	20	60	
Xenobiotics metabolism	28	56	50	
Glycan biosynthesis and metabolism, n=1				
Proteoglycan biosynthesis	4	4	100	
Metabolism of other amino acids, n=1				
Glutathione Metabolism	6	9	67	

[a] The average number of significant putative metabolites that were associated with dioxin(-like) exposures among each metabolic pathway.

[b] The average number of metabolites detected in each metabolic pathway.

[c] The percentage of overlap size to pathway size.

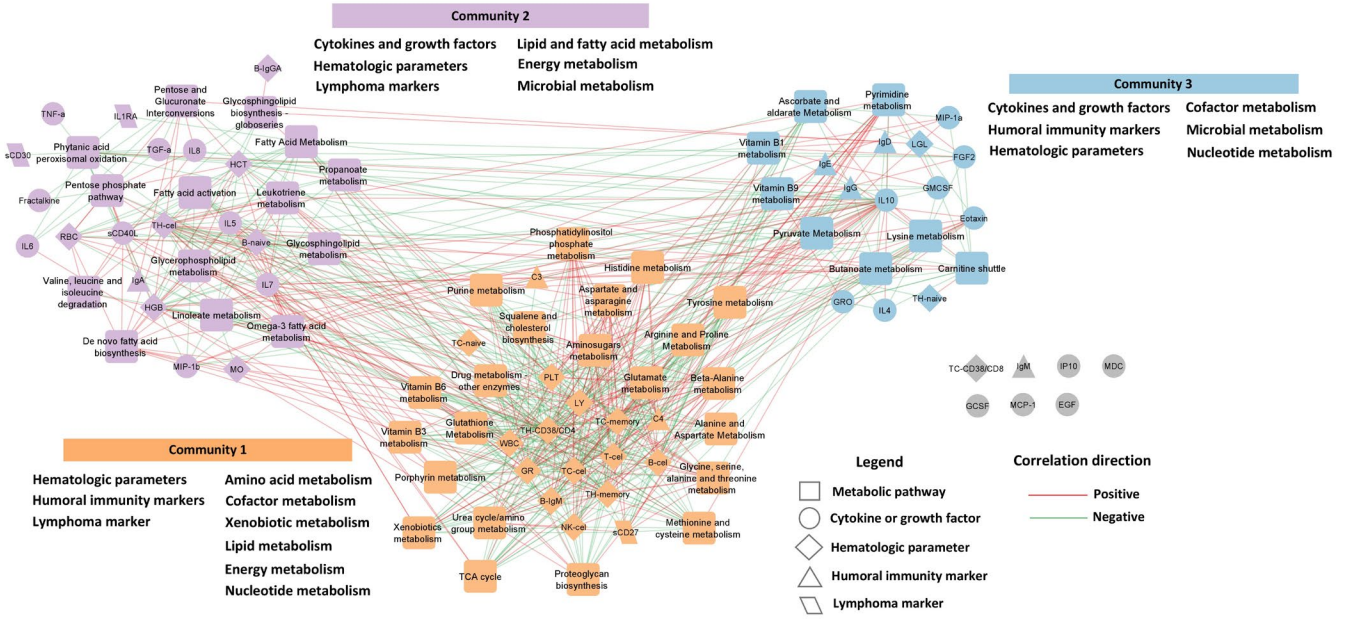


Figure 3. Network analysis of dioxin(-like) related pathways and immune markers

Abbreviation: IL, interleukin; GMCSF, granulocyte-macrophage colony-stimulating factor; GCSF, granulocyte colony-stimulating factor; TNF-α, tumor necrosis factor alpha; EGF, epidermal growth factor; FGF2, fibroblast growth factor 2; GRO, melanoma growth stimulatory activity/growth-related oncogene; IP10, interferon gamma-induced protein 10; MCP-1, monocyte chemotactic protein-1; MDC, macrophage derived chemokine; MIP-1α, macrophage inflammatory protein-1 alpha; MIP-1β, macrophage inflammatory protein-1 beta; sCD40L, soluble CD40 ligand; TGF-α, transforming growth factor alpha; sCD30, soluble CD30; sCD27, soluble CD27; IL1RA, interleukin 1 receptor antagonist; RBC, red blood cells; HGB, hemoglobin; HCT, hematocrit; PLT, platelet counts; MO, monocytes; GR, granulocytes; LY, lymphocytes; B-cel, B cells; B-naïve, naïve B cells; B-IgM, IgM+ memory B cells; B-IgG, IgG/IgA+ memory B cells; T-cel, T cells; TH-cell, T helper cells; TH-CD38/CD4, CD38/CD4 cells; TH-naïve, naïve CD4 cells; TH-memory, memory CD4 cells; TC-cel, cytotoxic T cells; TC-CD38/CD8, CD38/CD8 cells; TC-naïve, naïve CD8 cells; TC-memory, memory CD8 cells; LGL, large granular lymphocytes; NK-cel, natural killer cells.

with many lipid and fatty acid pathways, glucose metabolism pathways (pentose phosphate, pentose and glucuronate interconversions) and propanoate pathway.

A smaller cluster, Community 3, included cytokines and growth factors, immunoglobins (IgD, IgE, IgG), hematologic parameters (naïve CD4 cells, large granular lymphocytes (LGL)), clustered with pathways of cofactors (vitamin B1, B9), pyruvate and butanoate.

For networks specifically for each dioxin(-like) subclass, three communities were identified for both networks for PCDD(-related) and PCB(-related) compounds, and compositions were similar to those in the network for all exposures (Figure S8A, C). Whereas, in the case of PCDF(-related) compounds, the network yielded four communities (Figure S8B).

Discussion

In this study we employed a pioneering approach that integrates chemical-wide and metabolome-wide analyses. The rationale for using a chemical-wide analyses is that important biological insights might be missed by neglecting associated chemicals or related metabolites. This work was motivated by our previous observation in a MWAS on TCE conducted by Walker et al. (5). In that study, it was observed that most exposure-related biological effects exhibited stronger associations with previously unidentified metabolic products of TCE, rather than with TCE itself or recognized precursor metabolites. Similarly, in the present study, by including dioxin(-like) related compounds, we have obtained a much richer insight into the associated biological responses. This finding challenges the prevailing paradigm of evaluating the toxic effects of chemicals solely by examining parent compounds and potentially recognized metabolites (6).

It is important to note that applying this chemical-wide approach to different chemicals in other studies may not be straightforward. We acknowledge the uniqueness of the TCE and dioxin(-like) instances, wherein the identification of halogenated signals in untargeted HRMS analysis and prior knowledge of occupational exposure to parent compounds enabled us to identify related compounds and their metabolic products. However, considering the progress in HRMS data annotation capabilities through authentic standards and various in-silico tools, the feasibility of this approach is likely to extend to other chemical-classes in the future. Our findings in this study underscore the potential benefits of this strategy, suggesting that it could lead to an enhanced assessment of toxicological effects.

The activation of the aryl hydrocarbon receptor (AhR) stands as a well-established mechanism of action for dioxin(-like) compounds (38). While numerous studies have

investigated AhR-linked gene and protein expression (38), the underlying metabolic mechanism within the human body has received limited exploration. In an earlier study, we used nuclear magnetic resonance spectroscopy for metabolomic analysis on the same study population (39). Like our current investigation, this earlier analysis yielded few signals associated with the targeted TCDD (no. of features=27, p<0.05; none survived after multiple testing correction). Jeanneret et al. identified 24 metabolites as putative biomarkers of dioxin exposures among TCDD-exposed workers (40, 41). Liang et al., comparing HRMS for a high and low TCDD exposed group, identified 20 metabolites strongly correlated to the summed toxicological equivalent quantity scores of 17 congeners of 2,3,7,8-substituted dioxins (42). In our enriched analyses including dioxin(-like) related compounds, we identified over 7,000 HRMS signals, underscoring the potency of integrating a chemical-wide and metabolome-wide analyses.

Oxidative stress has been identified as a key mechanism underlying the toxicity of dioxin(-like) compounds (38), with biomolecules such as DNA, proteins and lipids becoming targets of free radical attacks (43). In our study, pathway enrichment analysis strongly suggests effects related to oxidative stress. The results point towards disruption in nucleotide metabolism, highlighted by the observation that PCDD(-related) and PCB(-related) compounds were linked to the purine pathway, while PCDF(-related) compounds were associated with both purine and pyrimidine pathways. Seven amino acids susceptible to oxidative damage, annotated at level 1 (methionine, cysteine, lysine, proline, threonine, histidine, tyrosine), were found to be associated with at least one of the three subclasses of dioxin(-like) exposures, and these associations were confirmed through enrichment analysis. Perturbations in lipids, encompassing membrane lipids (pathways of phospholipids, glycolipids and cholesterol) and long-chain polyunsaturated fatty acids (omega-3 fatty acids pathway and linoleic acid with level 1 annotation), provide additional support for oxidation-induced lipid peroxidation. Conversely, reductions in the antioxidants, specifically glutathione and ascorbic acid pathways, further support heightened oxidative stress.

Carcinogenesis stands as the most severe outcome of dioxin(-like) toxicity, with TCDD, PCB126 and 23478F being classified as human carcinogens (44). In our study, pathways involving pyruvate (a glycolysis product), pentose phosphate and the TCA cycle exhibited associations with both PCDD(-related) and PCDF(-related) compounds. Fatty acid pathways, encompassing biosynthesis, transport, activation and degradation, demonstrated relations to all three subclasses of dioxin(-like) exposures. These aberrations in bioenergetic synthesis and fatty acid metabolism are in line with microenvironmental shifts in human malignancies.

Our study also presents novel evidence of dysregulations within several metabolic pathways associated with dioxin(-like) exposures. Particularly noteworthy are microbiome-related pathways involving butanoate and propanoate and cofactor metabolisms including porphyrin and B vitamins. Animal studies have indicated that exposure to TCDD and PCB126 can induce alterations in gut microbial composition (45, 46). Additionally, PCB126 has been linked to elevated gut inflammation (46), while TCDD administration exhibited a mitigating effect on gut inflammation (47). Alteration of cofactors in response to dioxin(-like) compounds remains unexplored in experimental studies. The interplay between environmental dioxin(-like) compounds, the microbiome and cofactors calls for further investigation.

Adverse immunological effects have been extensively documented in experimental studies (48). However, human data remains inconclusive. To investigate the immune toxicity of dioxin(-like) exposures, we performed integrative network analysis, connecting perturbed metabolic pathways and phenotypic measures of immune responses from samples of highly TCDD-exposed workers. In the network incorporating all dioxin(-like) related pathways, distinct subsets of lymphocytes were grouped in the same community and linked to antioxidant pathways involving methionine, cysteine, and glutathione. Previous studies have shown that TCDD can suppress the differentiation of $CD4^+$ T cell into effector cells (49) and potently inhibit IgM production (50). Our findings suggest oxidative stress could potentially underlie immune toxicity. As expected, relevant measures of cytokines and growth factors clustered together and exhibited enrichment with two inflammation-related pathways, linoleate and leukotriene. Additionally, pathways related to fatty acid metabolism, bioenergy production and gut microbiome were clustered with B-cell activation markers shown to be predictive of lymphoma risk. This highlights the potential role of immune responses in dioxin(-like)-induced carcinogenesis and microbiome dysbiosis.

We acknowledge several limitations in our study. First, this study adopts a cross-sectional design. Consequently, we cannot infer the temporal sequence of exposure and health outcomes. Nonetheless, due to the protracted elimination of dioxin(-like) compounds, the measured levels effectively represent historical exposures. Second, the workers in factory A were, on average, 10 years older than workers in factory B, and it is possible that other unmeasured factors differed between factories. However, in subgroup analyses by factory, the associations for dioxin(-like) compounds and metabolic features remained highly consistent with those in the main analysis, which included workers from both factories (Figure S9). Therefore, we conclude that characteristics specific to each factory did not substantially impact the effects of dioxin(-like) exposures in our presented analyses. Third, over the course of an extended

35-year follow-up period in the Dutch herbicide cohort, 27% of participants had died (567 out of 2106 workers), and 5% were lost to follow-up (109 out of 2106) (10). This attrition may introduce the 'healthy worker effect', which may result in underestimating the adverse effects attributed to dioxin(-like) exposures. Fourth, precise annotations and absolute quantification of dioxin(-like) related compounds continue to pose challenges. This limitation also impedes ascertaining these toxic chemicals' origin, whether they originate from the environment or from endogenous metabolic modification. Therefore, future confirmatory studies are necessary. Lastly, the study assessed targeted and untargeted dioxin(-like) exposures in two separate laboratories, without accounting for potential measurement variations between different analytical pipelines. Additionally, the untargeted compounds were not normalized for lipid content. In a sensitivity analysis of MWAS on untargeted compounds, we further adjusted for total lipid levels measured at the time of the targeted measurement. The resulting altered features were similar to those generated in this study (data not shown). This suggests that lipid content did not considerably bias our findings.

We employed a pioneering approach that integrates chemical-wide and metabolome-wide analyses. This innovative approach substantially broadens the ability to evaluate the biological effects of chemical exposures, encompassing not only the traditionally recognized dioxin(-like) compounds but also all relevant compounds representing co-exposures and exposure metabolites. The results from the MWAS align with the existing understanding of dioxin(-like) toxicities, highlighting perturbations in metabolic pathways linked to amino acids, lipids and fatty acids, carbohydrates and nucleotides. Importantly, our study offers new perspectives regarding the mechanisms of action of dioxin(-like) compounds, such as altered activities of the gut microbiome.

References

1. White SS, Birnbaum LS. An overview of the effects of dioxins and dioxin-like compounds on vertebrates, as documented in human and ecological epidemiology. J Environ Sci Health C Environ Carcinog Ecotoxicol Rev. 2009;27(4):197-211.

2. Van den Berg M, Birnbaum LS, Denison M, et al. The 2005 World Health Organization reevaluation of human and Mammalian toxic equivalency factors for dioxins and dioxin-like compounds. Toxicological sciences : an official journal of the Society of Toxicology. 2006;93(2):223-41.

3. IARC. Polychlorinated Dibenzo-Para-Dioxins and Polychlorinated Dibenzofurans. IARC monographs on the evaluation of carcinogenic risks to humans. 1997;69:1-631.

4. Council NR. Health Risks from Dioxin and Related Compounds: Evaluation of the EPA Reassessment. Washington, DC: The National Academies Press; 2006. 268 p.

5. Walker DI, Uppal K, Zhang L, et al. High-resolution metabolomics of occupational exposure to trichloroethylene. Int J Epidemiol. 2016;45(5):1517-27.

6. Vermeulen R, Schymanski EL, Barabasi AL, Miller GW. The exposome and health: Where chemistry meets biology. Science. 2020;367(6476):392-6.

7. Boffetta P, Mundt KA, Adami HO, Cole P, Mandel JS. TCDD and cancer: a critical review of epidemiologic studies. Crit Rev Toxicol. 2011;41(7):622-36.

8. Bueno de Mesquita HB, Doornbos G, Van der Kuip DA, Kogevinas M, Winkelmann R. Occupational exposure to phenoxy herbicides and chlorophenols and cancer mortality in The Netherlands. American journal of industrial medicine. 1993;23(2):289-300.

9. Hooiveld M, Heederik DJ, Kogevinas M, et al. Second follow-up of a Dutch cohort occupationally exposed to phenoxy herbicides, chlorophenols, and contaminants. American journal of epidemiology. 1998;147(9):891-901.

10. Boers D, Portengen L, Bueno-de-Mesquita HB, Heederik D, Vermeulen R. Cause-specific mortality of Dutch chlorophenoxy herbicide manufacturing workers. Occup Environ Med. 2010;67(1):24-31.

11. Lubin JH, Colt JS, Camann D, et al. Epidemiologic evaluation of measurement data in the presence of detection limits. Environmental health perspectives. 2004;112(17):1691-6.

12. Boers D, Toxic RAo, Agents I, IRAS D. Exposure and risk assessment of dioxins and PCBs in humans: Universiteit Utrecht; 2010.

13. Hu X, Walker DI, Liang Y, et al. A scalable workflow to characterize the human exposome. Nat Commun. 2021;12(1):5575.

14. Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Analytical chemistry. 2006;78(3):779-87.

15. Broeckling CD, Afsar FA, Neumann S, Ben-Hur A, Prenni JE. RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. Analytical chemistry. 2014;86(14):6812-7.

16. Loos M. nontarget: Detecting, Combining and Filtering Isotope, Adduct and Homologue Series Relations in High-Resolution Mass Spectrometry (HRMS) Data. R package version 1.7 ed2015.

17. Csárdi G, Nepusz T, Traag V, et al. igraph: Network Analysis and Visualization in R. https://CRAN.R-project.org/package=igraph. Access 2023-11-17.

18. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome research. 2003;13(11):2498-504.

19. Liu KH, Nellis M, Uppal K, et al. Reference Standardization for Quantification and Harmonization of Large-Scale Metabolomics. Analytical chemistry. 2020.

20. Liu KH, Walker DI, Uppal K, et al. High-Resolution Metabolomics Assessment of Military Personnel: Evaluating Analytical Strategies for Chemical Detection. Journal of occupational and environmental medicine. 2016;58(8 Suppl 1):S53-61.

21. Yu T, Park Y, Li S, Jones DP. Hybrid feature detection and information accumulation using high-resolution LC-MS metabolomics data. Journal of proteome research. 2013;12(3):1419-27.

22. Uppal K, Soltow QA, Strobel FH, et al. xMSanalyzer: automated pipeline for improved feature detection and downstream analysis of large-scale, non-targeted metabolomics data. BMC bioinformatics. 2013;14:15.

23. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics (Oxford, England). 2007;8(1):118-27.

24. Lazer C. imputeLCMD: A Collection of Methods for Left-Censored Missing Data Imputation. R package version 2.1. https://CRAN.R-project.org/package=imputeLCMD. Access 2022-06-10.

25. Campanella G. omics: '--omics' Data Analysis Toolbox. R package version 0.1-5. 2016.

26. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological). 1995;57(1):289-300.

27. Schymanski EL, Jeon J, Gulde R, et al. Identifying small molecules via high resolution mass spectrometry: communicating confidence. Environ Sci Technol. 2014;48(4):2097-8.

28. Uppal K, Walker DI, Jones DP. xMSannotator: An R Package for Network-Based Annotation of High-Resolution Metabolomics Data. Analytical chemistry. 2017;89(2):1063-7.

29. Wishart DS, Guo A, Oler E, et al. HMDB 5.0: the Human Metabolome Database for 2022. Nucleic acids research. 2022;50(D1):D622-d31.

30. Li S, Park Y, Duraisingham S, et al. Predicting network activity from high throughput metabolomics. PLoS computational biology. 2013;9(7):e1003123.

31. Saberi Hosnijeh F, Boers D, Portengen L, Bueno-de-Mesquita HB, Heederik D, Vermeulen R. Long-term effects on humoral immunity among workers exposed to 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD). Occup Environ Med. 2011;68(6):419-24.

32. Saberi Hosnijeh F, Boers D, Portengen L, Bueno-de-Mesquita HB, Heederik D, Vermeulen R. Plasma Cytokine Concentrations in Workers Exposed to 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD). Front Oncol. 2012;2:37.

33. Saberi Hosnijeh F, Lenters V, Boers D, et al. Changes in lymphocyte subsets in workers exposed to 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD). Occup Environ Med. 2012;69(11):781-6.

34. Saberi Hosnijeh F, Portengen L, Bueno-de-Mesquita HB, Heederik D, Vermeulen R. Circulating soluble CD27 and CD30 in workers exposed to 2,3,7,8-Tetrachlorodibenzo-p-dioxin (TCDD). Cancer Epidemiol Biomarkers Prev. 2013;22(12):2420-4.

35. Uppal K, Ma C, Go YM, Jones DP, Wren J. xMWAS: a data-driven integration and differential network analysis tool. Bioinformatics. 2018;34(4):701-2.

36. González I, Cao KA, Davis MJ, Déjean S. Visualising associations between paired 'omics' data sets. BioData mining. 2012;5(1):19.

37. Blondel V, Guillaume J, Lambiott R, Lefebvre E. Fast unfolding of communities in large networks. J Stat Mech-Theory. 2008;P10008.

38. Grishanova AY, Perepechaeva ML. Aryl Hydrocarbon Receptor in Oxidative Stress as a Double Agent and Its Biological and Therapeutic Significance. Int J Mol Sci. 2022;23(12):6719.

39. Saberi Hosnijeh F, Pechlivanis A, Keun HC, et al. Serum metabolomic pertubations among workers exposed to 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD). Environ Mol Mutagen. 2013;54(7):558-65.

40. Jeanneret F, Boccard J, Badoud F, et al. Human urinary biomarkers of dioxin exposure: analysis by metabolomics and biologically driven data dimensionality reduction. Toxicology letters. 2014;230(2):234-43.

41. Jeanneret F, Tonoli D, Hochstrasser D, et al. Evaluation and identification of dioxin exposure biomarkers in human urine by high-resolution metabolomics, multivariate analysis and in vitro synthesis. Toxicology letters. 2016;240(1):22-31.

42. Liang Y, Tang Z, Jiang Y, et al. Serum metabolic changes associated with dioxin exposure in a Chinese male cohort. Environ Int. 2020;143:105984.

43. Lobo V, Patil A, Phatak A, Chandra N. Free radicals, antioxidants and functional foods: Impact on human health. Pharmacogn Rev. 2010;4(8):118-26.

44. IARC. Chemical agents and related occupations. IARC monographs on the evaluation of carcinogenic risks to humans. 2012;100(Pt F):9-562.

45. Lefever DE, Xu J, Chen Y, Huang G, Tamas N, Guo TL. TCDD modulation of gut microbiome correlated with liver and immune toxicity in streptozotocin (STZ)-induced hyperglycemic mice. Toxicology and applied pharmacology. 2016;304:48-58.

46. Petriello MC, Hoffman JB, Vsevolozhskaya O, Morris AJ, Hennig B. Dioxin-like PCB 126 increases intestinal inflammation and disrupts gut microbiota and metabolic homeostasis. Environmental pollution (Barking, Essex : 1987). 2018;242(Pt A):1022-32.

47. Benson JM, Shepherd DM. Aryl hydrocarbon receptor activation by TCDD reduces inflammation associated with Crohn's disease. Toxicological sciences : an official journal of the Society of Toxicology. 2011;120(1):68-78.

48. Esser C. Dioxins and the Immune System. Encyclopedia of Immunotoxicology2014. p. 1-7.

49. Marshall NB, Kerkvliet NI. Dioxin and immune regulation: emerging role of aryl hydrocarbon receptor in the generation of regulatory T cells. Ann N Y Acad Sci. 2010;1183:25-37.

50. North CM, Crawford RB, Lu H, Kaminski NE. Simultaneous in vivo time course and dose response evaluation for TCDD-induced impairment of the LPS-stimulated primary IgM response. Toxicological sciences : an official journal of the Society of Toxicology. 2009;112(1):123-32.

Chapter 4 Supplemental materials

Text S1. Identification of dioxin(-like) related compounds

Approximately 60% of the pairwise Spearman correlations between the 499 chlorinated compounds detected by untargeted GC-HRMS and the 29 targeted dioxin(-like) compounds exhibited positive relationships (8,427 out of 14,471 correlations) (Figure S1). When considering the *p*-value cutoff as 0.05, the average proportion of negative correlations across all targeted dioxin(-like) compounds was 40% (Table S2). Notably, compounds displaying negative correlations to targeted compounds are unlikely to be co-exposures or metabolites of parental compounds. As such, the percentage of negatively correlated significant outcomes can be regarded as empirically derived from false discovery rates. For the *p*-value cutoff of 0.002, the average proportion of negative correlations was around 20%.

Applying the criteria of correlation exceeding 0 and *p*-value below 0.002, we identified 152 chlorinated compounds that exhibited correlations with 23 targeted dioxin(-like) compounds (Figure S2). These correlation coefficients ranged from 0.26 to 0.75, with a median of 0.33. Among these, 106 chlorinated compounds were correlated to more than one targeted dioxin(-like) compound, and 45 compounds were related to more than five targeted compounds.

Table S1. Descriptive statistics of targeted dioxin(-like) compounds

Compound	Factory A (n=76)[a]	Factory B (n=61)[a]	p-value[b]	Reference level[a,c]
PCDDs				
TCDD	4.35 (1.60-5.60)	0.30 (0.21-0.56)	<0.0001	<LOD
12378D	7.55 (6.50-9.10)	6.40 (5.30-7.00)	0.016	<LOD
123478D	4.50 (3.60-5.40)	1.90 (1.11-3.10)	<0.001	<LOD
123678D	30.0 (27.2-33.1)	20.7 (15.6-25.7)	<0.001	19.8 (17.8-21.6)
123789D	3.70 (2.70-4.60)	0.62 (0.42-1.06)	<0.0001	<LOD
1234678D	15.6 (12.5-18.1)	12.5 (9.6-14.8)	0.041	23.2 (21.1-25.6)
OCDD	216 (192-269)	171 (139-212)	0.036	<LOD
Dioxin-like PCDFs				
2378F	0.02 (0.01-0.03)	0.02 (0.01-0.04)	0.523	<LOD
12378F	0.14 (0.10-0.19)	0.14 (0.12-0.21)	0.342	<LOD
23478F	22.9 (20.5-27.5)	13.1 (11.6-15.4)	<0.0001	<LOD
123478F	4.75 (4.10-5.70)	4.50 (3.70-4.90)	0.091	<LOD
123678F	5.65 (5.00-6.20)	4.90 (4.20-5.70)	0.059	<LOD
123789F	0.39 (0.39-0.40)	0.39 (0.39-0.40)	0.936	<LOD
234678F	0.30 (0.14-0.55)	0.11 (0.07-0.17)	0.008	<LOD
1234678F	4.30 (3.70-5.20)	4.10 (3.20-5.90)	0.990	<LOD
1234789F	0.10 (0.07-0.15)	0.11 (0.07-0.16)	0.827	<LOD
OCDF	0.68 (0.38-1.80)	0.49 (0.31-0.72)	0.512	<LOD
Dioxin-like PCBs				
PCB77	23.7 (20.4-26.6)	23.6 (20.0-29.2)	0.940	NA
PCB81	4.95 (3.80-6.50)	6.80 (6.00-7.90)	<0.001	<LOD
PCB126	38.8 (35.0-46.2)	37.0 (29.4-42.3)	0.276	<LOD
PCB169	83.3 (78.2-90.1)	63.1 (54.9-74.0)	<0.0001	<LOD
PCB105	1.55 (1.20-1.80)	1.40 (1.10-1.90)	0.661	0.98 (0.85-1.08)
PCB114	1.00 (0.80-1.10)	0.80 (0.60-1.10)	0.379	NA
PCB118	11.0 (9.2-13.0)	10.0 (6.7-12.9)	0.557	4.71 (4.16-5.06)
PCB123	0.23 (0.20-0.30)	0.21 (0.15-0.27)	0.479	NA
PCB156	13.9 (12.6-15.9)	12.4 (10.0-14.6)	0.076	3.10 (2.80-3.80)
PCB157	2.50 (2.30-3.00)	2.30 (1.70-2.80)	0.177	0.75 (0.70-0.90)
PCB167	3.40 (2.90-3.90)	2.80 (2.00-3.80)	0.296	0.50 (0.44-0.60)
PCB189	2.30 (2.10-2.50)	1.90 (1.70-2.20)	0.008	<LOD
$TCDD_{max}$[d]	98.9 (19.2-341.6)	0.30 (0.21-0.56)	<0.0001	

Abbreviations: PCDD, pentachloro dibenzo-p-dioxin; TCDD, 2,3,7,8-tetrachloro dibenzo-p-dioxin; 12378D, 1,2,3,7,8-PCDD; 123478D, 1,2,3,4,7,8-PCDD; 123678D, 1,2,3,6,7,8-PCDD; 123789D, 1,2,3,7,8,9-PCDD; 1234678D, 1,2,3,4,6,7,8-PCDD; OCDD, octachloro dibenzo-p-dioxin; PCDF,

pentachloro dibenzofuran; 2378F, 2,3,7,8- tetrachloro dibenzofuran; 12378F, 1,2,3,7,8- PCDF; 23478F, 2,3,4,7,8-PCDF; 123478F, 1,2,3,4,7,8-hexachloro dibenzofuran (HCDF); 123678F, 1,2,3,6,7,8-HCDF; 123789F, 1,2,3,7,8,9-HCDF; 234678F, 2,3,4,6,7,8-HCDF; 1234678F, 1,2,3,4,6,7,8-heptachloro dibenzofuran (HpCDF); 1234789F, 1,2,3,4,5,8,9-HpCDF; OCDF, octachloro dibenzofuran; PCB, polychlorinated biphenyls; PCB77, 3,3',4,4'-tetrachlorobiphynyl (tetraCB); PCB81, 3,4,4',5-tetraCB; PCB126, 3,3',4,4',5-pentachlorobiphenyl (pentaCB); PCB169, 3,3',4,4',5,5'-hexachlorobiphenyl (hexaCB); PCB105, 2,3,3',4,4'-pentaCB; PCB114, 2,3,4,4',5-pentaCB; PCB118, 2,3',4,4',5-pentaCB; PCB123, 2',3,4,4',5-pentraCB; PCB156, 2,3,3',4,4',5-hexaCB; PCB157, 2,3,3',4,4',5'-hexaCB; PCB167, 2,3',4,4',5,5'-hexaCB; PCB189, 2,3,3',4,4',5,5'-heptaCB; LOD, limit of detection; NA, not available.

[a] Parts per trillion, lipid adjusted; data are presented as median (95% confidence interval)

[b] p-value from Mann-Whitney-Wilcoxon test, concentrations between factory A vs. factory B

[c] Concentrations of dioxin(-like) compounds from US general males in 2003-2004, reported by National Health and Nutrition Examination Survey.

[d] Estimated maximum levels of TCDD

Table S2. The number of significant correlations between targeted dioxin(-like) compounds and chlorinated compounds under different cutoffs of *p*-value

Targeted compound	*p*-value<0.05		*p*-value<0.01		*p*-value<0.002		*p*-value<0.001		*p*-value<0.00005	
	No.[a]	%[b]	No.[a]	%[b]	No.[a]	%[b]	No.[a]	%[b]	No.[a]	%[b]
PCDDs										
TCDD	198/31	16	121/25	21	74/21	28	61/17	28	18/9	50
12378D	170/31	18	97/14	14	52/12	23	45/11	24	21/4	19
123478D	171/31	18	94/17	18	51/13	25	36/10	28	11/2	18
123678D	192/27	14	124/20	16	73/11	15	59/8	14	23/4	17
123789D	163/20	12	73/6	8	35/5	14	26/4	15	3/1	33
1234678D	111/17	15	36/5	14	6/1	17	4/0	0	1/0	0
OCDD	134/17	13	50/8	16	18/1	6	10/1	0	1/0	0
Dioxin-like PCDFs										
2378F	24/18	75	3/2	67	none	NA	none	NA	none	NA
12378F	48/31	65	6/2	33	1/0	0	none	NA	none	NA
23478F	263/37	14	202/23	11	149/19	13	137/19	14	85/13	15
123478F	139/33	24	69/15	22	28/3	11	20/2	10	7/0	0
123678F	115/57	50	52/23	44	34/15	44	27/11	41	11/2	18
123789F	8/6	75	2/2	100	none	NA	none	NA	none	NA
234678F	130/13	10	63/3	5	28/1	4	17/1	6	none	NA
1234678F	43/18	42	8/3	38	2/1	50	1/0	0	none	NA
1234789F	37/31	84	6/6	100	1/1	100	1/1	100	none	NA
OCDF	1/1	100	none	NA	none	NA	none	NA	none	NA
Dioxin-like PCBs										
PCB77	6/6	100	none	NA	none	NA	none	NA	none	NA
PCB81	136/128	94	56/53	95	16/16	100	10/10	100	none	NA

PCB126	84/52	62	28/11	39	15/3	20	13/2	15	4/0	0
PCB169	206/28	14	138/25	18	70/18	26	51/13	25	19/1	5
PCB105	64/33	52	28/8	29	23/3	13	22/2	9	14/0	0
PCB114	105/18	17	48/6	12	29/0	0	25/0	0	19/0	0
PCB118	76/40	53	39/11	28	23/3	13	22/2	9	16/0	0
PCB123	30/5	17	18/1	6	8/0	0	8/0	0	3/0	0
PCB156	119/20	17	51/9	18	29/5	17	24/1	4	21/0	0
PCB157	90/16	18	41/8	20	27/4	15	23/1	4	20/0	0
PCB167	113/54	48	44/16	36	26/2	8	25/1	4	20/0	0
PCB189	103/19	18	39/9	23	25/1	4	23/0	0	19/0	0
Average %[c]	40		32		22		19		9	

Abbreviation: NA, not applicable.

p-values were calculated from tests of pairwise Spearman correlations between known dioxin(-like) compounds and chlorinated compounds.

[a] No. of all significant associations under the p-value cutoff / No. of negative and significant associations under the p-value cutoff.

[b] The percentage of negative associations among all significant associations.

[c] The average percentage of negative associations among all known dioxin(-like) compounds.

Table S3. List of immune phenotypic markers

Full name	Acronym
Cytokines and growth factors	
Interleukin 4	IL4
Interleukin 5	IL5
Interleukin 6	IL6
Interleukin 7	IL7
Interleukin 8	IL8
Interleukin 10	IL10
Granulocyte-macrophage colony-stimulating factor	GMCSF
Granulocyte colony-stimulating factor	GCSF
Tumor necrosis factor alpha	TNF-a
Epidermal growth factor	EGF
Eotaxin	Eotaxin
Fibroblast growth factor 2	FGF2
Fractalkine	Fractalkine
Melanoma growth stimulatory activity/growth-related oncogene	GRO
Interferon gamma-induced protein 10	IP10
Monocyte chemotactic protein-1	MCP-1
Macrophage derived chemokine	MDC
Macrophage inflammatory protein-1 alpha	MIP-1a
Macrophage inflammatory protein-1 beta	MIP-1b
Transforming growth factor alpha	TGF-a
Soluble CD40 ligand	sCD40L
Hematologic parameters	
Red blood cells	RBC
Hemoglobin	HGB
Hematocrit	HCT
Platelet count	PLT
White blood cells	WBC
Monocytes	MO
Granulocytes	GR
Lymphocytes	LY
B cells	B-cel
Naïve B cells	B-naive
IgM+ memory B cells	B-IgM
IgG/IgA+ memory B cells	B-IgGA
T cells	T-cel
T helper cells	TH-cel
CD38/CD4 cells	TH-CD38/CD4
Naïve CD4 cells	TH-naive
Memory CD4 cells	TH-memory

Cytotoxic T cells	TC-cel
CD38/CD8 T cells	TC-CD38/CD8
Naïve CD8 cells	TC-naive
Memory CD8 cells	TC-memory
Large granular lymphocytes	LGL
Natural killer cells	NK-cel
Humoral immunity markers	
Immunoglobulin A	IgA
Immunoglobulin D	IgD
Immunoglobulin E	IgE
Immunoglobulin G	IgG
Immunoglobulin M	IgM
Complement factor 3	C3
Complement factor 4	C4
Lymphoma markers	
Soluble CD30	sCD30
Soluble CD27	sCD27
Interleukin 1 receptor antagonist	IL1RA

Table S4. The number of dioxin(-like) related compounds in relation to each targeted compound

Targeted compound	n. of chlorinated related compounds[a]
PCDDs	
TCDD	53
12378D	40
123478D	38
123678D	62
123789D	30
1234678D	5
OCDD	17
Total	109
Dioxin-like PCDFs	
2378F	0
12378F	1
23478F	130
123478F	25
123678F	19
123789F	0
234678F	27
1234678F	1
1234789F	0
OCDF	0
Total	136
Dioxin-like PCBs	
PCB77	0
PCB81	0
PCB126	12
PCB169	52
PCB105	20
PCB114	29
PCB118	20
PCB123	8
PCB156	24
PCB157	23
PCB167	24
PCB189	24
Total	58

[a] Chlorinated compounds were considered in relation to targeted dioxin(-like) compounds when they exhibited a correlation coefficient exceeding 0 and *p*-value below 0.002 with at least one targeted compound.

Table S5. Targeted and related dioxin(-like) compounds in each community from network analysis

Targeted compounds	Related compounds
Community 1	
1234678D	C00014, C00026, C00036, C00038, C00047,
123478F, 123678F	C00050, C00068, C00149, C00150, C00206,
PCB126, PCB105, PCB114, PCB118,	C00268, C00269, C00508, C00586, C02267,
PCB123, PCB156, PCB157, PCB167,	C03500, C03641, C00031, C00428, C00430,
PCB189	C01742, C02696, C00077, C00094, C00459, C01193
Community 2	
12378D, 123478D, 123678D, 123789D, OCDD	C00028, C00239, C00320, C00461, C00764, C03231, C04165, C05116, C05225, C00035, C00189, C00272, C00294, C00391, C00535, C00555, C00582, C00691, C00737, C01785, C01911, C00922, C01192, C03354, C05219, C00071, C00568, C00609, C00630, C00883, C01058, C01362, C02246, C03335, C03698, C06046, C01023, C01070, C01761, C00160, C00897, C02490, C05131
Community 3	
TCDD 23478F, 234678F	C00087, C00295, C00324, C00556, C00660, C00712, C00810, C00915, C00926, C01101, C01128, C01215, C02000, C02113, C02226, C02624, C02642, C02652, C04306, C04781, C06136, C06162, C00052, C05187, C00018, C00039, C00241, C00307, C05104, C00644, C00738, C01659, C02110, C02614, C05649, C00033, C00044, C00045, C00064, C00109, C00113, C00134, C00139, C00161, C00221, C00466, C00570, C00649, C00795, C01120, C01141, C01436, C01480, C01685, C01812, C02269, C02655, C02671, C02871, C03365, C03428, C03951, C04160, C04163, C04182, C04244, C05499, C05657, C06286, C02945
Community 4	
PCB169	C00115, C01633, C02016, C03662, C06171, C00890, C02084, C00856, C00928, C00488, C00805
Community 5	
12378F	C04604
Community 6	
1234678F	C00066

Table S6. Dioxin(-like) related compounds with abundant associated metabolic features (n>300)

Related compound	Community[a]	Correlated known compounds[b]	No. of features[c]
C18-negative			
C00295	3	TCDD	1179
C05187	3	123478D, 23478F, 123478F	952
C00459	1	123478F, PCB167	732
C02671	3	23478F, 234678F	679
C00430	1	12378D, 123678D, 23478F, 123678F, PCB169, PCB105, PCB114, PCB118 PCB156, PCB157, PCB167, PCB189	585
C00047	1	TCDD, 12378D, 123478D, 123678D, 23478F, PCB169, PCB105, PCB114, PCB118, PCB156, PCB157, PCB167, PCB189	567
C00044	3	23478F	565
C00764	2	TCDD, 12378D, 123478D, 123678D, 23478F, 234678F, PCB169	556
C00094	1	123478D, 123678D, 23478F, PCB169, PCB105, PCB114, PCB118, PCB156, PCB157, PCB167, PCB189	518
C02113	3	TCDD, 23478F	512
C01480	3	23478F	493
C00660	3	TCDD, 123678D, 23478F, 234678F, PCB169	458
C04182	3	23478F	431
C02945	3	234678F	411
C00795	3	23478F, 234678F	391
C00915	3	TCDD, 23478F	363
C00052	3	12378D, 123678D, 23478F, 234678F	352
C00139	3	23478F, 123478F	343
C00582	2	12378D, 23478F, 234678F	338
C00269	1	TCDD, 12378D, 123478D, 123678D, 23478F, 123478F, 123678F, PCB126, PCB169, PCB105, PCB114, PCB118, PCB156, PCB157, PCB167, PCB189	327
C00691	2	12378D, 123678D, OCDD, 23478F, 123478F, 234678F	310
C02016	4	TCDD, 23478F, PCB169	309
HILIC-positive			
C00922	2	123478D	1464
C02084	4	123789D, 23478F, PCB169	1186
C00221	3	23478F	1158
C00035	2	12378D, 123478D, 123678D, 23478F, 123478F, PCB169, PCB114	565
C00066	6	1234678F	302

[a] Community from community detection of the network between known and related dioxin(-like) related compounds (Supplemental table 5).

[b] Under the criteria of Spearman correlation coefficient exceeding 0 and *p*-value below 0.002.

[c] The number of significantly associated features under FDR 20%, adjusted by age, BMI and factory.

Table S7. Annotation of metabolic features associated with dioxin(-like) exposures

HMDB ID	Name	m/z	Retention time	Confidence level	Mode	Adduct	PCDD (-related)[a]	PCDF (-related)[a]	PCB (-related)[a]
Amino acid									
HMDB00001	1-methylhistidine	170.0924	82.3	1	HILIC	M+H	1	2	1
HMDB00043	betaine	116.0717	240.4	1	C18	M-H	2	2	1
HMDB00064	creatine	130.0605	151.2	1	C18	M-H	6	9	2
HMDB00112	aminobutyrate	104.0706	61.8	1	HILIC	M+H	1	0	0
HMDB00123	glycine	120.0032	67.8	1	HILIC	M+2Na-H	0	0	1
HMDB00148	glutamic acid	151.0627	25	1	C18	M-H	1	5	1
		153.0772	70.4	1	HILIC	M+H	2	3	1
HMDB00158	tyrosine	181.0638	30.2	1	C18	M-H	1	1	0
		183.0783	57.8	1	HILIC	M+H	0	1	0
		183.0804	86.3	1	HILIC	M+H	17	20	7
		183.0804	29.7	1	HILIC	M+H	2	2	0
		182.0812	53.9	1	HILIC	M+H	1	2	1
HMDB00162	proline	114.056	289.6	1	C18	M-H	1	0	0
		116.0706	75	1	HILIC	M+H	1	0	0
HMDB00167	threonine	120.0655	25.7	1	HILIC	M+H	0	1	0
		120.0655	68	1	HILIC	M+H	2	2	1
HMDB00168	asparagine	131.0445	292.4	1	C18	M-H	11	16	4
		131.0445	182.8	1	C18	M-H	7	11	3
		133.0603	106.7	1	HILIC	M+H	1	1	1
HMDB00177	histidine	154.0623	24.4	1	C18	M-H	0	1	0
		156.0768	82.2	1	HILIC	M+H	2	2	1
HMDB00182	lysine	145.0983	32.4	1	C18	M-H	1	2	0
		147.1128	87.1	1	HILIC	M+H	3	3	2
HMDB00192	cystine	241.0232	197.4	1	C18	M-H	2	2	1
		241.0233	34.2	1	C18	M-H	2	5	2
		243.0378	185.9	1	HILIC	M+H	0	1	0
HMDB00214	ornithine	133.0971	91.2	1	HILIC	M+H	2	2	1

HMDB00562	creatinine	112.0516	285.2	1	C18	M-H	9	14	4
		114.0662	38.7	1	HILIC	M+H	1	2	1
HMDB00574	cysteine	122.027	114.1	1	HILIC	M+H	0	1	0
HMDB00641	glutamine	145.0619	23.2	1	C18	M-H	0	0	1
		154.0862	92.3	1	HILIC	M+H	0	1	0
		154.0863	225.5	1	HILIC	M+H	2	8	1
		154.0863	29.1	1	HILIC	M+H	20	21	4
HMDB00696	methionine	155.0374	187.3	1	C18	M-H	1	1	0
HMDB00883	valine	116.0717	240.4	1	C18	M-H	2	2	1
HMDB00904	citrulline	174.0885	22.9	1	C18	M-H	1	1	1
		176.103	80.6	1	HILIC	M+H	2	2	1
HMDB00929	tryptophan	203.0827	27	1	C18	M-H	0	1	1
HMDB06029	N-acetylglutamine	247.0934	28.1	2	C18	M+Hac-H	0	1	0
		233.0779	26.8	2	C18	M-H+HCOOH	0	1	0
		187.0724	27.8	2	C18	M-H	1	2	1
Androstane steroid									
HMDB00369	3b17b-dihydroxyetiocholane	351.2541	236	2	C18	M+Hac-H	8	11	6
		291.2328	233.1	2	C18	M-H	8	10	6
HMDB00383	3a17a-dihydroxy-5b-androstane	351.2541	236	2	C18	M+Hac-H	8	11	6
		291.2328	233.1	2	C18	M-H	8	10	6
HMDB00412	3b17a-dihydroxy-5a-androstane	351.2541	236	2	C18	M+Hac-H	8	11	6
		291.2328	233.1	2	C18	M-H	8	10	6
HMDB00458	5a-androstane-3a17a-diol	351.2541	236	2	C18	M+Hac-H	8	11	6
		291.2328	233.1	2	C18	M-H	8	10	6
HMDB00493	5a-androstane-3b17b-diol	351.2541	236	2	C18	M+Hac-H	8	11	6
		291.2328	233.1	2	C18	M-H	8	10	6
HMDB00551	etiocholanediol	351.2541	236	2	C18	M+Hac-H	8	11	6
		291.2328	233.1	2	C18	M-H	8	10	6
HMDB00554	dihydroandrosterone	351.2541	236	2	C18	M+Hac-H	8	11	6
		291.2328	233.1	2	C18	M-H	8	10	6

Carbohydrate									
HMDB00122	D-glucose	209.0728	57.7	1	HILIC	M+Na	0	1	0
Cholestane steroid									
HMDB00067	cholesterol	369.3511	26.2	1	HILIC	M+H-H2O	4	5	3
Cofactor									
HMDB00244	riboflavin	377.1454	43.7	1	HILIC	M+H	1	2	0
HMDB01431	pyridoxamine	169.0948	84.8	1	HILIC	M+H	1	1	0
HMDB01488	niacin	122.0247	289.1	1	C18	M-H	9	11	6
		124.0394	170.1	1	HILIC	M+H	2	1	1
		124.0394	12.4	1	HILIC	M+H	0	1	0
		124.0411	100.6	1	HILIC	M+H	2	1	1
HMDB01545	pyridoxal	166.0511	109.4	1	C18	M-H	1	0	0
Dipeptide									
HMDB11173	glycylhydroxyproline	247.0934	28.1	2	C18	M+Hac-H	0	1	0
		233.0779	26.8	2	C18	M-H+HCOOH	0	1	0
		187.0724	27.8	2	C18	M-H	1	2	1
Fatty acid									
HMDB00201	Acetylcarnitine	204.1231	49.5	1	HILIC	M+H	2	1	1
HMDB00220	Palmitic acid	255.2328	251.8	1	C18	M-H	9	14	8
		257.2476	245.8	1	HILIC	M+H	1	1	0
		257.2477	19.5	1	HILIC	M+H	7	16	3
HMDB00222	Palmitoylcarnitine	400.3499	92.7	1	HILIC	M+H	2	2	0
		400.3499	254.2	1	HILIC	M+H	3	5	3
HMDB00511	Decanoic acid	171.139	246.9	1	C18	M-H	3	3	4
		171.1392	118.7	1	C18	M-H	6	5	2
HMDB00638	Lauric acid	199.1704	270	1	C18	M-H	2	1	1
HMDB00673	Linoleic acid	279.2327	238.7	1	C18	M-H	8	13	6
HMDB00827	Stearic acid	283.264	276.9	1	C18	M-H	8	12	6
		285.2788	25.8	1	HILIC	M+H	4	6	2
HMDB01388	Alpha-Linolenic acid	277.2172	224.1	1	C18	M-H	8	11	6
HMDB02212	Arachidic acid	311.2956	290.3	1	C18	M-H	7	7	6

Glycerolipid									
HMDB11131	MG(18:0/0:0/0:0)	376.3419	23.4	2	HILIC	M+NH4	3	4	2
		381.2973	28.2	2	HILIC	M+Na	3	4	2
		359.3152	25.8	2	HILIC	M+H	8	12	4
HMDB11533	MG(0:0/16:0/0:0)	353.2662	22.2	2	HILIC	M+Na	1	2	2
		331.2839	26.1	2	HILIC	M+H	7	8	3
		376.3419	23.4	2	HILIC	M+NH4	3	4	2
		381.2973	28.2	2	HILIC	M+Na	3	4	2
		359.3152	25.8	2	HILIC	M+H	8	12	4
HMDB11564	MG(16:0/0:0/0:0)	353.2662	22.2	2	HILIC	M+Na	1	2	2
		331.2839	26.1	2	HILIC	M+H	7	8	3
Glycerophospholipid									
HMDB00495	Androstanediol	351.2541	236	2	C18	M+Hac-H	8	11	6
		291.2328	233.1	2	C18	M-H	8	10	6
HMDB11483	LysoPE(0:0/20:2)	564.3326	221.9	2	C18	M+Hac-H	0	1	1
		504.3092	223.5	2	C18	M-H	0	2	1
HMDB11499	LysoPE(0:0/24:6)	612.3294	211.7	2	C18	M+Hac-H	0	0	1
		552.3092	211.6	2	C18	M-H	0	0	4
HMDB11513	LysoPE(20:2/0:0)	564.3326	221.9	2	C18	M+Hac-H	0	1	1
		504.3092	223.5	2	C18	M-H	0	2	1
HMDB11529	LysoPE(24:6/0:0)	612.3294	211.7	2	C18	M+Hac-H	0	0	1
		552.3092	211.6	2	C18	M-H	0	0	4
Purine or its derivative									
HMDB00289	Uric acid	167.0211	24.4	1	C18	M-H	7	10	4
HMDB00292	Xanthine	151.0262	25.1	1	C18	M-H	38	51	17

[a] The number of known and related dioxin(-like) compounds associated with the specific metabolite.

Table S8. Correlations of the same pathways across different dioxin(-like) subclasses

Pathway	PCDD vs. PCDF	PCDD vs. PCB	PCDF vs. PCB
De novo fatty acid biosynthesis	1.00	1.00	1.00
Aspartate and asparagine metabolism	0.99	0.99	0.99
Fatty acid activation	1.00	1.00	1.00
Arginine and Proline Metabolism	0.98	-0.98	-0.98
Pentose and Glucuronate Interconversions	0.98	0.97	0.98
Butanoate metabolism	0.93	-0.97	-0.88
Lysine metabolism	-0.97	0.95	-0.95
Linoleate metabolism	0.99	-1.00	-0.99
Phytanic acid peroxisomal oxidation	0.98	0.98	1.00
Purine metabolism	-1.00	0.99	-0.98
Histidine metabolism	0.99	-0.96	-0.96
Fatty Acid Metabolism	1.00	0.55	0.56
Omega-3 fatty acid metabolism	1.00	1.00	1.00
Drug metabolism - other enzymes	0.99	-0.99	-1.00
Vitamin B6 metabolism	1.00	1.00	1.00
Aminosugars metabolism	0.97	0.99	0.99
Phosphatidylinositol phosphate metabolism	0.99	-0.98	-0.98
Glutathione Metabolism	0.89	-0.97	-0.93
Xenobiotics metabolism	0.99	-0.99	-0.99
Urea cycle/amino group metabolism	0.98	-0.73	-0.74
Alanine and Aspartate Metabolism	-0.86	-0.62	0.82

PC1 scores for each pathway were calculated based on selected features in the corresponding pathway. For the same pathway, included features could be slightly different across the three dioxin(-like) subclasses. Correlation coefficients of the same pathway in different exposure subclasses were presented.

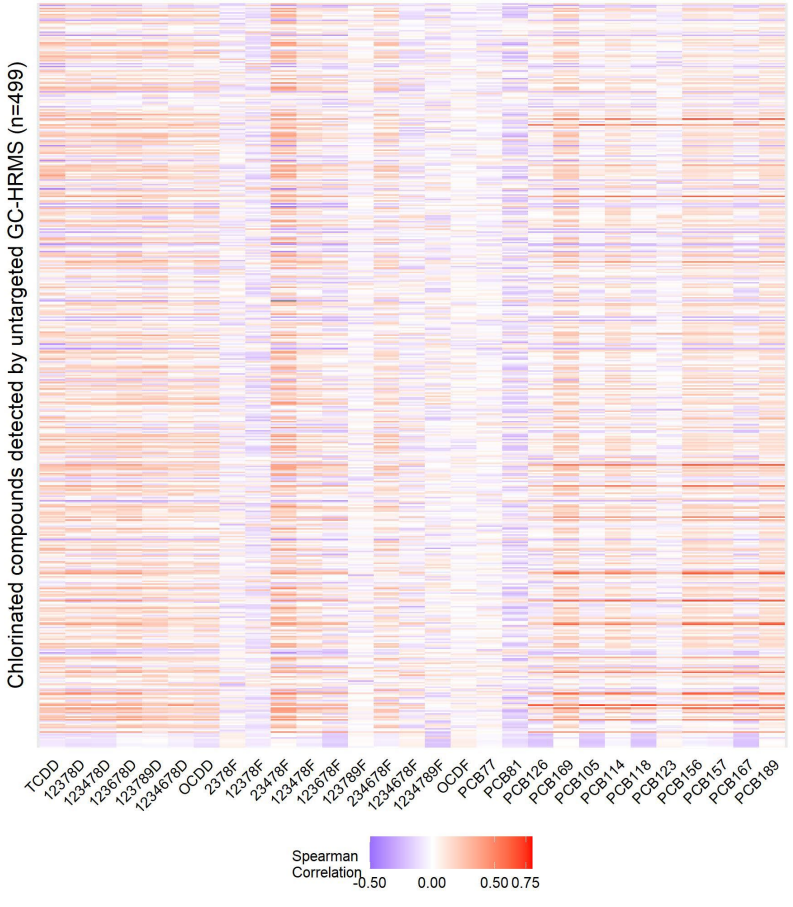


Figure S1. Correlation heatmap of targeted dioxin(-like) compounds and all detected chlorinated compounds

Tile color showed magnitude of pairwise Spearman correlation coefficients of the targeted dioxin(-like) compounds (n=29) and all chlorinated compounds detected by untargeted GC-HRMS (n=499).

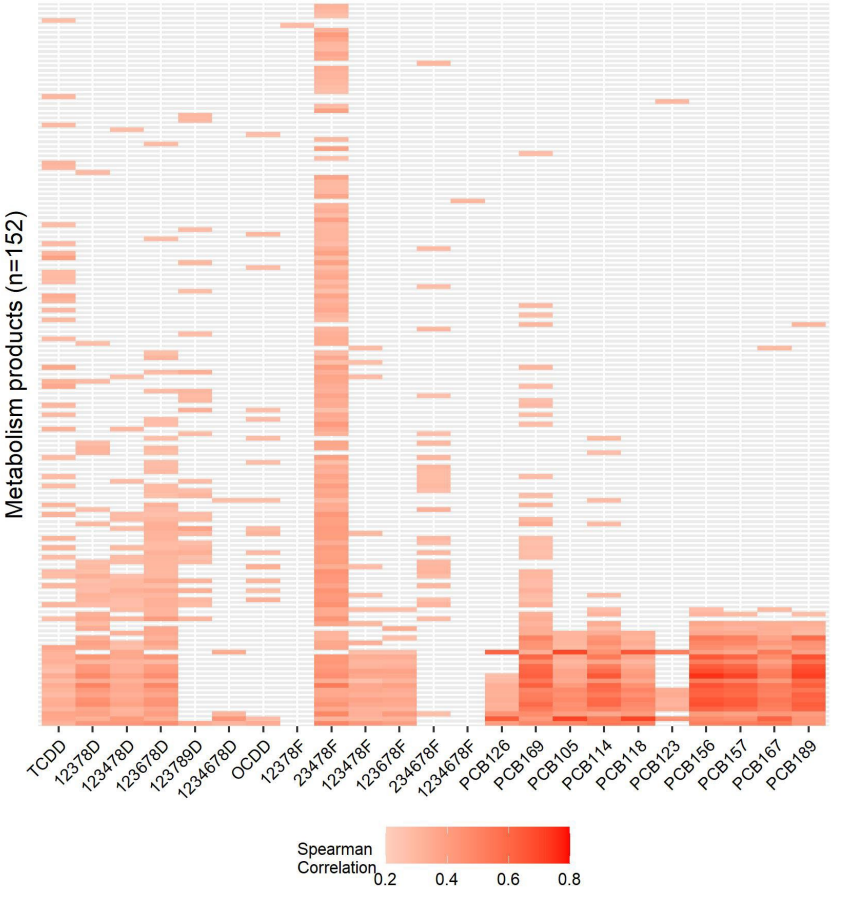


Figure S2. Correlations between targeted dioxin(-like) compounds and their related compounds

Chlorinated compounds (n=152) were identified as possible related dioxin(-like) compounds, under the criteria of pairwise correlation coefficient exceeding 0 and *p*-value below 0.002. Tile color showed magnitude of correlation coefficients; gray color indicated the correlation was either non-significant (*p*-value≥0.002) or in negative direction, or both.

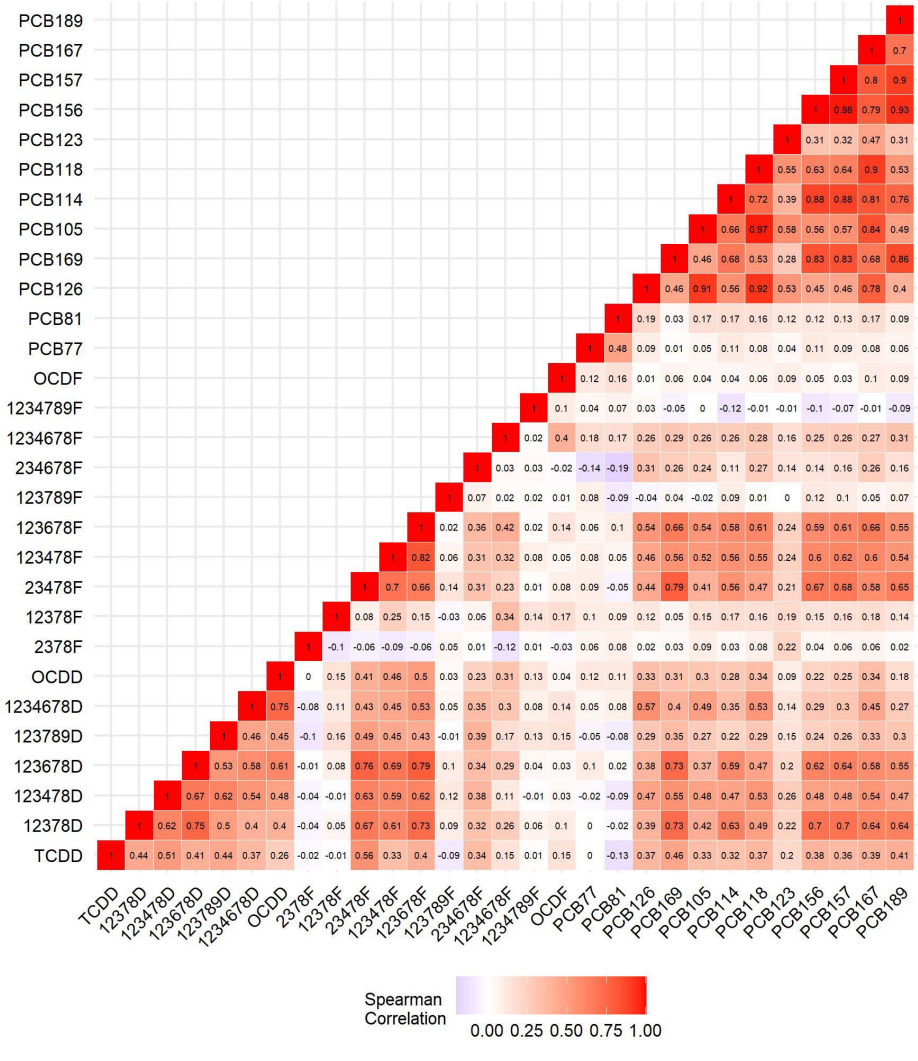


Figure S3. Correlation matrix of targeted dioxin(-like) compounds

Values in tiles were pairwise Spearman correlation coefficients among targeted dioxin(-like) compounds (n=29).

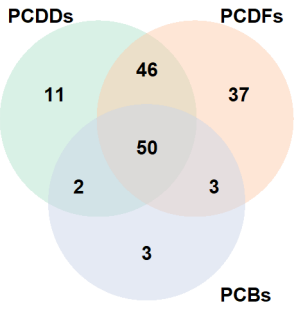


Figure S4. Venn diagram of PCDD-related, PCDF-related and PCB-related compounds

Dioxin(-like) related compounds were grouped by being related to PCDDs, PCDFs and PCBs. The respective numbers were 109, 136, 58.

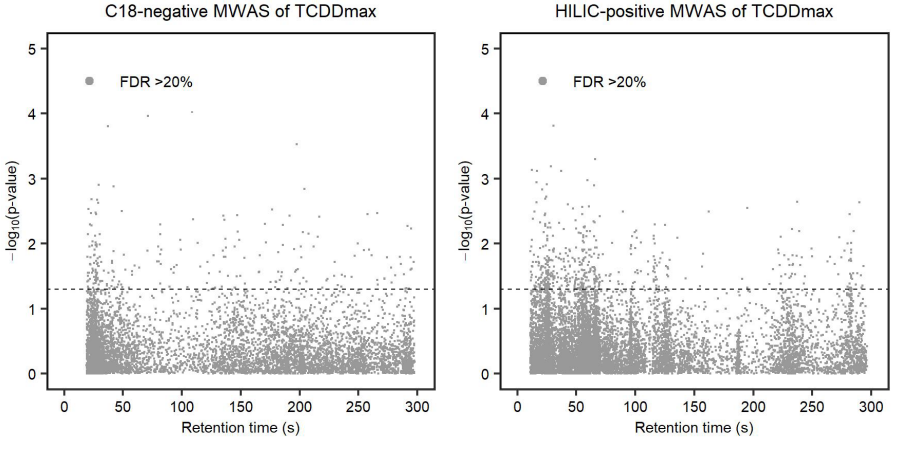


Figure S5. Manhattan plots of MWAS of maximum TCDD levels

Each dot represented a metabolic feature (in total 6,914 for C18-negative mode, 10,733 for HILIC-positive mode) and was presented by -$\log_{10}$(*p*-value), measuring the strength of association, against retention time of the feature. Dashed line represented raw *p*-value of 0.05. No feature was deemed significant under the threshold of false discovery rate 20%.

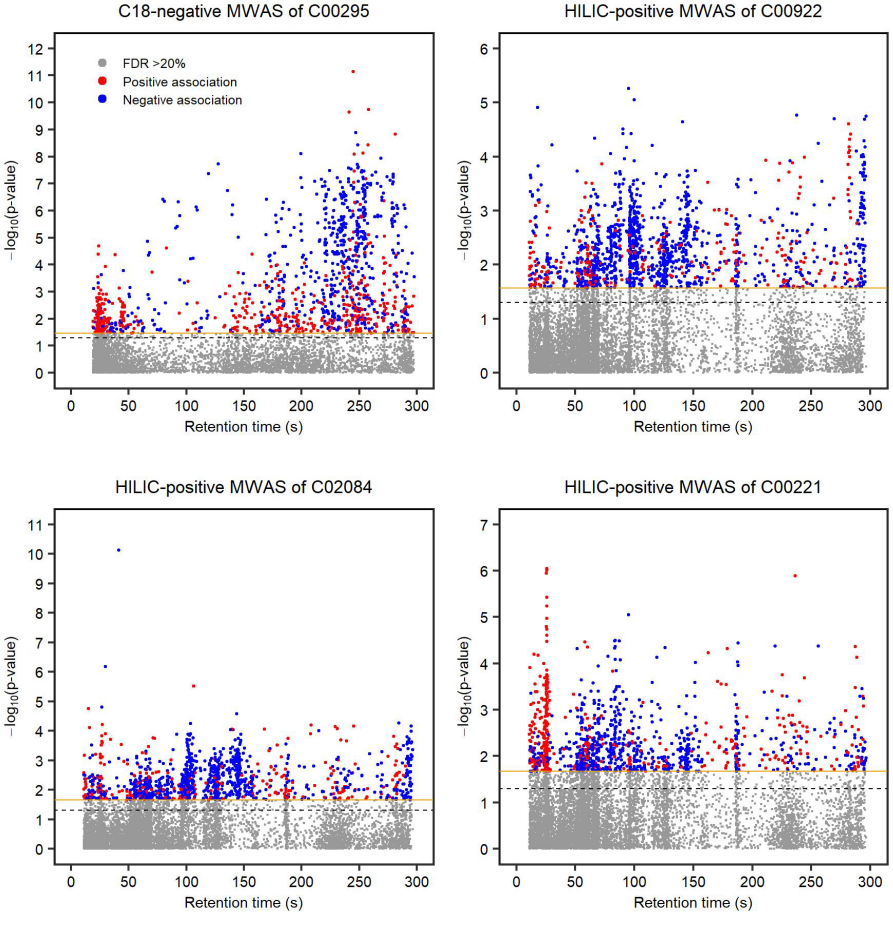


Figure S6. Manhattan plots of MWAS of certain dioxin(-like) related compounds

These four dioxin(-like) related compounds were significantly associated with >1000 metabolic features. Each dot represented a metabolic feature (in total 6,914 for C18-negative mode, 10,733 for HILIC-positive mode) and was presented by -$\log_{10}$(*p*-value), measuring the strength of association, against retention time of the feature. Significant features were colored according to the association direction. Dashed line represented raw *p*-value of 0.05; solid line represented FDR 20%.

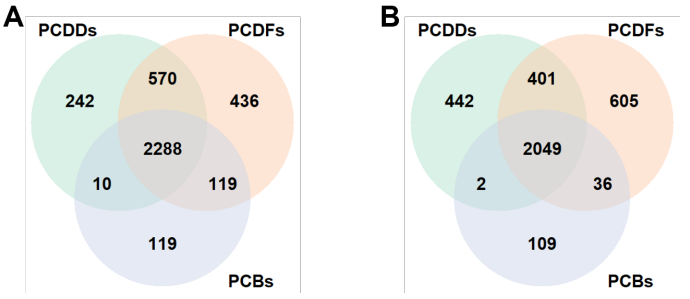
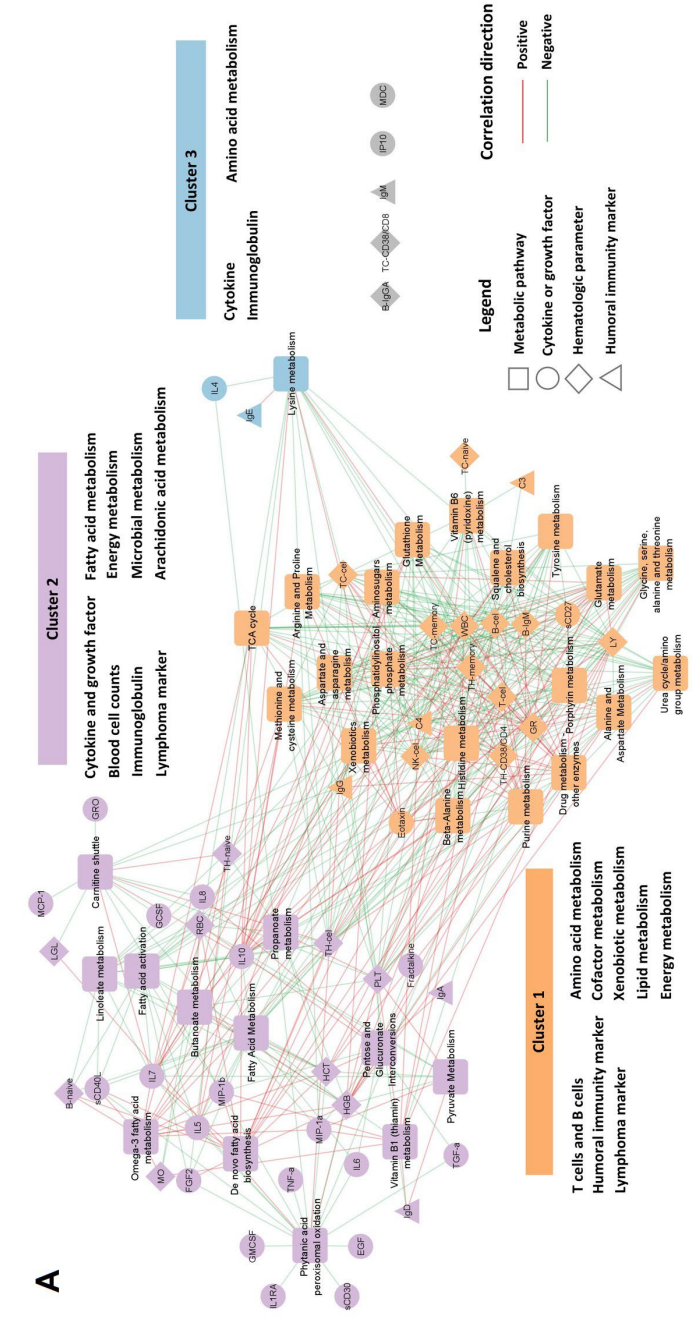


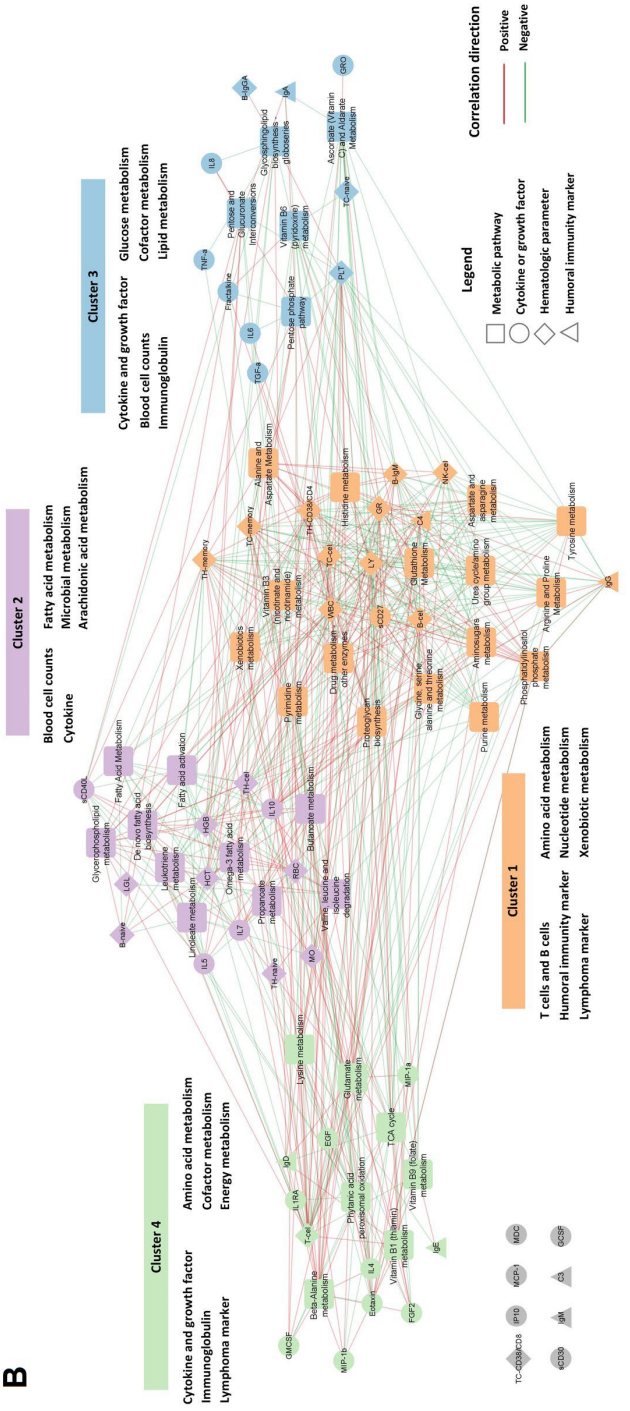
Figure S7. Venn diagram of metabolic features associated with PCDD(-related), PCDF(-related) and PCB(-related) compounds

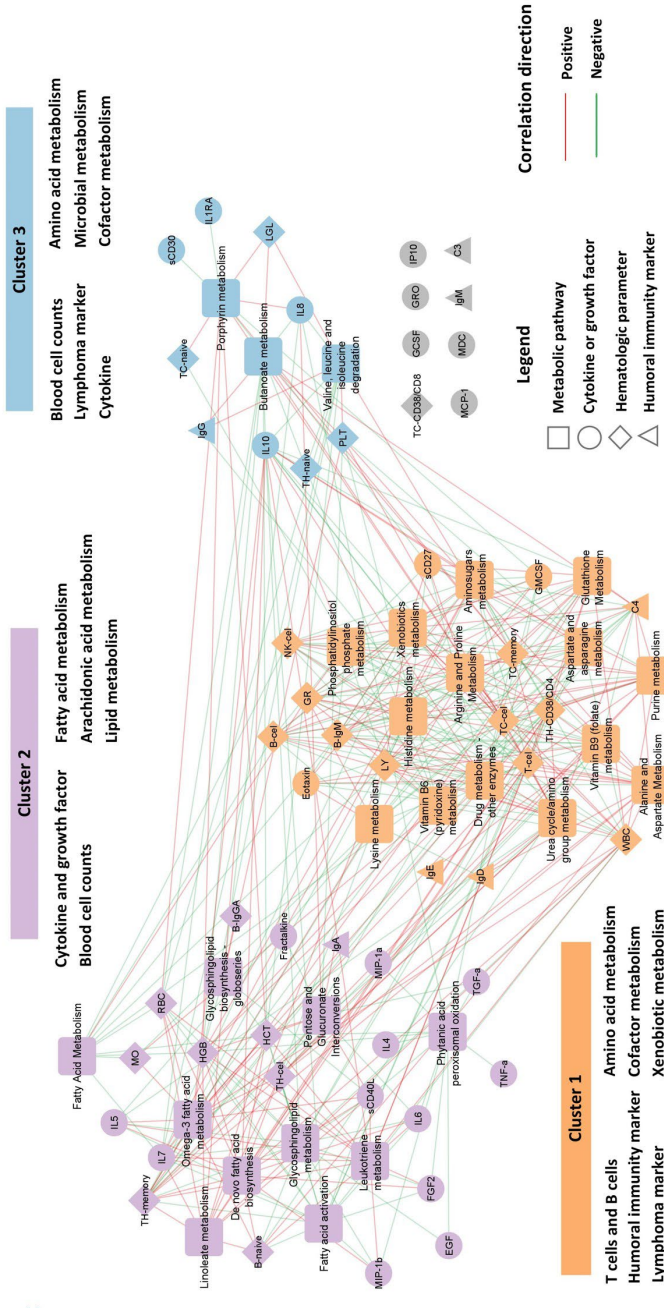
C18-negative (A) and HILIC-positive (B) features were identified under FDR 20% of MWAS with at least one of known and related dioxin(-like) compounds. Metabolic features were categorized by being associated with PCDD(-related), PCDF(-related) and PCB(-related) compounds.

Figure S8. Network analysis of dioxin(-like) related pathways with immune markers, separately by PCDD(-related) (A), PCDF(related) (B), and PCB(-related) compounds (C)

Abbreviation: IL, interleukin; GMCSF, granulocyte-macrophage colony-stimulating factor; GCSF, granulocyte colony-stimulating factor; TNF-α, tumor necrosis factor alpha; EGF, epidermal growth factor; FGF2, fibroblast growth factor 2; GRO, melanoma growth stimulatory activity/growth-related oncogene; IP10, interferon gamma-induced protein 10; MCP-1, monocyte chemotactic protein-1; MDC, macrophage derived chemokine; MIP-1α, macrophage inflammatory protein-1 alpha; MIP-1β, macrophage inflammatory protein-1 beta; sCD40L, soluble CD40 ligand; TGF-α, transforming growth factor alpha; sCD30, soluble CD30; sCD27, soluble CD27; IL1RA, interleukin 1 receptor antagonist; RBC, red blood cells; HGB, hemoglobin; HCT, hematocrit; PLT, platelet counts; MO, monocytes; GR, granulocytes; LY, lymphocytes; B-cel, B cells; B-naïve, naïve B cells; B-IgM, IgM+ memory B cells; B-IgG, IgG/IgA+ memory B cells; T-cel, T cells; TH-cell, T helper cells; TH-CD38/CD4, CD38/CD4 cells; TH-naïve, naïve CD4 cells; TH-memory, memory CD4 cells; TC-cel, cytotoxic T cells; TC-CD38/CD8, CD38/CD8 cells; TC-naïve, naïve CD8 cells; TC-memory, memory CD8 cells; LGL, large granular lymphocytes; NK-cel, natural killer cells.


A
Cluster 2
Cytokine and growth factor
Blood cell counts
Immunoglobulin
Lymphoma marker
Fatty acid metabolism
Energy metabolism
Microbial metabolism
Arachidonic acid metabolism
Cluster 3
Cytokine
Immunoglobulin
Amino acid metabolism
Legend
Metabolic pathway
Cytokine or growth factor
Hematologic parameter
Humoral immunity marker
Correlation direction
Positive
Negative
Cluster 1
T cells and B cells
Humoral immunity marker
Lymphoma marker
Amino acid metabolism
Cofactor metabolism
Xenobiotic metabolism
Lipid metabolism
Energy metabolism
Lysine metabolism
TCA cycle
Arginine and Proline Metabolism
Methionine and cysteine metabolism
Aspartate and asparagine metabolism
Phosphatidylinositol phosphate metabolism
Aminosugars metabolism
Glutathione Metabolism
Vitamin B6 (pyridoxine) metabolism
Squalene and cholesterol biosynthesis
Tyrosine metabolism
Glutamate metabolism
Glycine, serine, alanine and threonine metabolism
Urea cycle/amino group metabolism
Alanine and Aspartate Metabolism
Porphyrin metabolism
Drug metabolism other enzymes
Purine metabolism
Beta-Alanine metabolism
Histidine metabolism
Xenobiotics metabolism
Linoleate metabolism
Fatty acid activation
Butanoate metabolism
Fatty Acid Metabolism
Propanoate metabolism
Pentose and Glucuronate Interconversions
Pyruvate Metabolism
Carnitine shuttle
Omega-3 fatty acid metabolism
De novo fatty acid biosynthesis
Phytanic acid peroxisomal oxidation
Vitamin B1 (thiamin) metabolism
B-naive
TH-naive
TH-cell
TC-cell
TC-naive
TC-memory
B-cell
B-naive
B-IgM
B-IgGA
TH-memory
TH-CD38HLADR
TH-CD38HCD4
T-cell
NK-cell
WBC
RBC
HCT
HGB
PLT
IgA
IgG
IgM
IgD
C3
C4
GR
Eotaxin
Fractalkine
IL4
IL5
IL6
IL7
IL8
IL10
IL1RA
IFNg
TNFa
TGFa
FGF2
EGF
GMCSF
GCSF
MIP1a
MIP1b
MCP1
MO
CD40
CD30
CD27
LCL
GRO
LY
TC-CD38HCD8
IP10
MDC


B
Cluster 1
T cells and B cells
Humoral immunity marker
Lymphoma marker
Amino acid metabolism
Nucleotide metabolism
Xenobiotic metabolism
Cluster 2
Blood cell counts
Cytokine
Fatty acid metabolism
Microbial metabolism
Arachidonic acid metabolism
Cluster 3
Cytokine and growth factor
Blood cell counts
Immunoglobulin
Glucose metabolism
Cofactor metabolism
Lipid metabolism
Cluster 4
Cytokine and growth factor
Immunoglobulin
Lymphoma marker
Amino acid metabolism
Cofactor metabolism
Energy metabolism
Legend
Metabolic pathway
Cytokine or growth factor
Hematologic parameter
Humoral immunity marker
Correlation direction
Positive
Negative


C
Cluster 1
T cells and B cells
Humoral immunity marker
Lymphoma marker
Amino acid metabolism
Cofactor metabolism
Xenobiotic metabolism
Cluster 2
Cytokine and growth factor
Blood cell counts
Fatty acid metabolism
Arachidonic acid metabolism
Lipid metabolism
Cluster 3
Blood cell counts
Lymphoma marker
Cytokine
Amino acid metabolism
Microbial metabolism
Cofactor metabolism
Legend
Metabolic pathway
Cytokine or growth factor
Hematologic parameter
Humoral immunity marker
Correlation direction
Positive
Negative

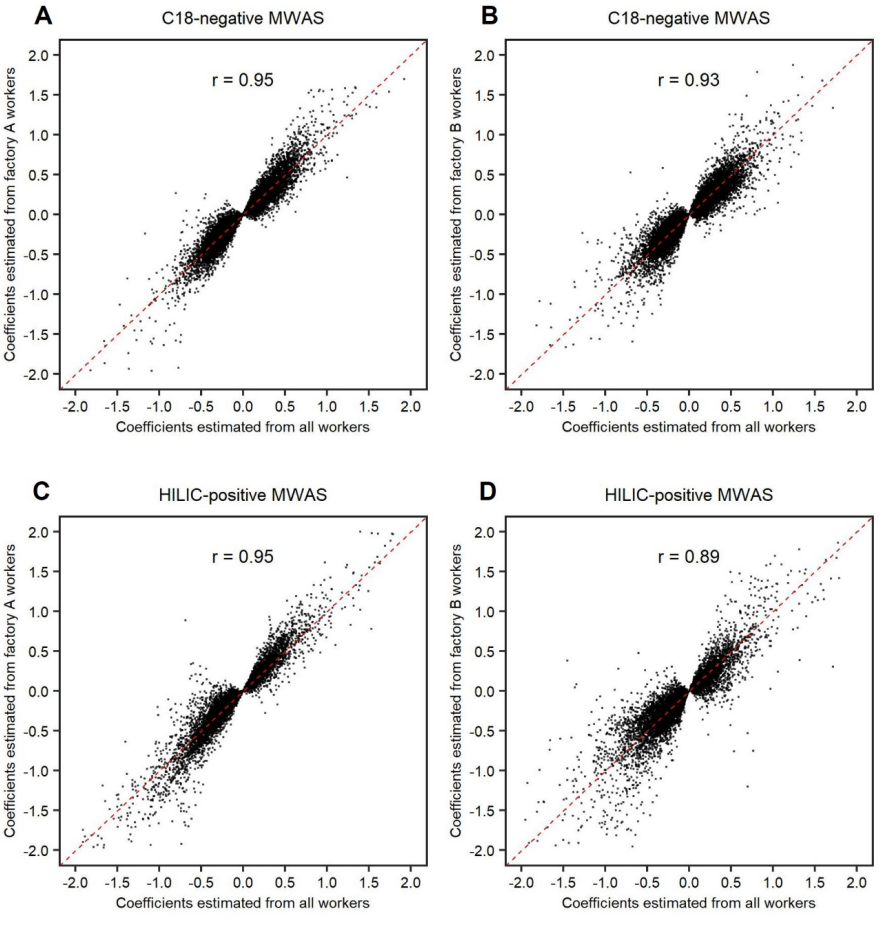


Figure S9. Comparison of main analyses and subgroup analyses on factory

Each dot represented one significant association (one targeted/related dioxin(-like) compound – one metabolic feature) detected in the main analysis of all workers (in total 22,557 C18-negative MWAS, 11,642 from HILIC-positive MWAS). They were displayed by coefficients from the main analysis against the corresponding values from subgroup analysis of factory A or factory B workers (adjusted by age and BMI). The dotted line is the line of identity.

Chapter 5

Association of Coffee Consumption and Prediagnostic Caffeine Metabolites With Incident Parkinson's Disease in a Population-based Cohort

Yujia Zhao[#], Yunjia Lai[#], Hilde Konijnenberg, José María Huerta, Ana Vinagre-Aragon, Jara Anna Sabin, Johnni Hansen, Dafina Petrova, Carlotta Sacerdote, Raul Zamora-Ros, Valeria Pala, Alicia K. Heath, Salvatore Panico, Marcela Guevara, Giovanna Masala, Christina M. Lill, Gary W. Miller, Susan Peters, Roel Vermeulen

[#] Joint first authors

Corresponding author: Roel Vermeulen

Published on *Neurology*

https://pubmed.ncbi.nlm.nih.gov/38513162/

Abstract

Background and Objective: Inverse associations between caffeine intake and Parkinson's disease (PD) have been frequently implicated in human studies. However, no studies have quantified biomarkers of caffeine intake years before PD onset and investigated if and which caffeine metabolites are related to PD.

Methods: Associations between self-reported total coffee consumption and future PD risk were examined in the EPIC4PD study, a prospective population-based cohort including six European countries. PD cases were identified through medical records and reviewed by expert neurologists. Hazard ratios (HRs) and 95% confidence intervals (CIs) for coffee consumption and PD incidence were estimated using Cox proportional hazards models. A case-control study nested within the EPIC4PD was conducted, recruiting incident PD cases and matching each case with a control by age, sex, study center and fasting status at blood collection. Caffeine metabolites were quantified by high-resolution mass-spectrometry in baseline collected plasma samples. Using conditional logistic regression models, odds ratios (ORs) and 95% CIs were estimated for caffeine metabolites and PD risk.

Results: In the EPIC4PD cohort (comprising 184,024 individuals), the multivariable-adjusted HR comparing the highest coffee intake to nonconsumers was 0.63 (95%CI 0.46-0.88, *p*-value 0.006). In the nested case-control study, which included 351 PD incident cases and 351 matched controls, prediagnostic caffeine and its primary metabolites, paraxanthine and theophylline, were inversely associated with PD risk. The ORs were 0.80 (95%CI 0.67-0.95, *p*-value 0.009), 0.82 (95%CI 0.69-0.96, *p*-value 0.015), and 0.78 (95%CI 0.65-0.93, *p*-value 0.005), respectively. Adjusting for smoking and alcohol consumption did not substantially change these results.

Discussion: This study demonstrates that the neuroprotection of coffee on PD is attributed to caffeine and its metabolites by detailed quantification of plasma caffeine and its metabolites years prior to diagnosis.

Introduction

Parkinson's disease (PD) is the most common motor neurodegenerative disorder for which there is no effective prevention or curative treatment available so far. Coffee consumption has been associated with a reduced risk of PD in several prospective cohorts during the past twenty years (1-5). The protective effect was also present for caffeine from non-coffee sources, such as tea, cola beverages, and chocolate (2-4). In contrast, the effect was not observed for decaffeinated coffee (5), suggesting that the inverse association between coffee consumption and PD is largely due to caffeine and its metabolites, rather than other bioactive compounds in coffee. However, these findings were based on food questionnaire data rather than on measuring caffeine or its metabolites in pre-disease biological samples.

Some exploratory case-control studies have indicated that blood concentrations of caffeine and its major metabolites in humans, namely paraxanthine and theophylline, were reduced in prevalent PD patients as compared to healthy individuals (6-8). Following these observations, clinical trials have been initiated to investigate whether caffeine or its metabolites could slow the progression of PD. Unfortunately, these studies have shown no benefit of caffeine and its metabolites on symptom attenuation and progression in PD (9, 10). However, no studies to date have prospectively investigated the role of caffeine levels in prediagnostic samples to investigate whether caffeine and its metabolites could be protective in a prodromal state of the disease. This research question can only be investigated in very large cohorts with baseline blood samples and long follow-up available, such as in the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort. The EPIC cohort comprises more than half million participants across Europe that have been followed up for >20 years and for which baseline blood samples were collected and ascertained in a highly standardized fashion (11). During the long follow-up, several hundred participants have been diagnosed with PD (12).

Coffee is the most widely consumed psychoactive beverage in the world. Unraveling the biological action of caffeine on PD not only carries important public health implications, but also enhances our understanding of PD etiology and fosters potential prevention strategies. In this study, we aimed to investigate the relationship between caffeine and future PD risk prospectively in the EPIC cohort, utilizing self-reported coffee consumption and direct measurement of prediagnostic caffeine and its metabolites.

Methods

The EPIC4PD cohort

The EPIC is an ongoing prospective cohort study designed to explore the relationship between nutrition and non-communicable diseases (11). Baseline recruitment was conducted between 1992 and 2000 across 23 centers in 10 European countries. The EPIC cohort comprises 519,978 participants (366,521 women and 153,457 men), mostly aged 35-70 years at recruitment. At enrollment, comprehensive dietary habits and lifestyle data were collected using questionnaires. Moreover, anthropometric measurements were conducted, and blood samples were obtained (11).

To prospectively investigate the association between prediagnostic risk factors and the incidence of PD, a sub-study known as EPIC4PD was initiated with the EPIC cohort (12). The inclusion of study centers in EPIC4PD depended on the availability of neurologists for PD ascertainment. Ultimately, the EPIC4PD was based on a source population of 192,980 subjects from six countries, including Sweden (Umeå and Malmö), the UK (Cambridge), the Netherlands (Utrecht), Germany (Heidelberg), Spain (Navarra, San Sebastián and Murcia), and Italy (Turin, Varese, Florence and Naples). The Naples and Utrecht cohorts exclusively comprised women, while all the other cohorts included participants of both sexes. To date, follow-up for the EPIC4PD is 98.5% complete, and the median duration follow-up of the entire population is 12.8 years (maximum 20.8 years) (12).

Case ascertainment and study population

In brief, potential PD cases were identified through record linkage and validated by experts in movement disorders through clinical records (12). Reliability of diagnoses was determined by the quality of clinical data (rated as ‘poor’, ‘good’ or ‘excellent’), as well as the confidence degree of the neurologist expert on the basis of their final judgement (rated as ‘low’, ‘medium’ or ‘high’). Diagnoses were defined as ‘definite’ only when the confidence degree of the neurologist was high and the data quality was excellent; ‘very likely’ when the confidence degree was high, while data quality was either good or poor; ‘probable’ when the confidence degree was medium and data quality was either excellent or good; and diagnoses were defined as ‘possible’ in all remaining cases. A total of 786 PD cases was ascertained. Cases who received a diagnosis after the date of recruitment were defined as incident cases (n=639) (12).

Our study consisted of two parts, including a prospective cohort study (EPIC4PD) and a nested case-control study, to interrogate the links of coffee consumption and caffeine and its metabolites with PD risk, respectively. For the EPIC4PD, several exclusion criteria were applied: prevalent PD cases and cases without date of diagnosis (n=147);

participants with PD-like conditions (multiple system atrophy, progressive supranuclear palsy, vascular parkinsonism, dementia with Lewy bodies, essential tremor, PD with essential tremor, unclassifiable parkinsonism) (n=214); those with missing information on coffee consumption and smoking status at recruitment (n=8,484); those with extreme coffee consumption (>2,500 ml/day) (n=111), to exclude possible bias related to caffeine addiction.

Incident PD cases within the EPIC4PD study were considered for inclusion in the nested case-control study, provided that a plasma sample was accessible. Subjects from Sweden were excluded due to the unavailability of plasma samples. For each PD case, one control was selected by incidence density sampling matched for age at recruitment, sex, study center and fasting status at blood collection.

Dietary and lifestyle data

Dietary intake was assessed by a dietary questionnaire that had been developed and validated in each participating country. A face-to-face dietary interview was applied in Spain, while self-administered questionnaires were used in other countries (11). To increase comparability across the study centers, a standardized 24-hour diet recall was collected, as a reference calibration method, from a stratified random sample of 36,900 subjects from the entire EPIC cohort (13). Total coffee consumption was available for all countries. Caffeinated coffee consumption was available for almost all centers except for Naples and Umeå. Information regarding decaffeinated coffee consumption was collected from participants in Germany, Italy (excluding Naples), the Netherlands, and the UK (14). Participants reported the number of cups of coffee consumed per month, week or day. Daily coffee consumption (in milliliters, ml) was then calculated using the typical sizes of cups for each center (14).

Participants also completed questionnaires on lifestyle including smoking and alcohol consumption, education level, and physical activity. Female participants additionally reported menopausal status and hormone usage. Height and weight were measured following standardized protocols, and body mass index (BMI) was subsequently calculated (11).

Caffeine metabolite measurement

In the nested case-control study, plasma samples for the subjects were sourced from the cohort biobank at the International Agency for Research on Cancer (IARC). These samples were collected between 1992 and 1998, with on average 8 years prior to the diagnosis of PD in cases. To profile circulating caffeine metabolites in plasms, we performed untargeted metabolomics analysis using a liquid chromatography-high resolution mass spectrometry (LC-HRMS)-based platform as previously described (15,

16) (Text S1). To maximize detection of polar and nonpolar metabolites, two complementary analyses were performed, namely hydrophilic interaction liquid chromatography (HILIC)-ESI(+) and reverse-phase chromatography (RPC)-ESI(-), termed as 'HILpos' and 'C18neg', both operated in full scan mode at 120,000 mass resolution with a mass-to-charge (*m/z*) range of 85-1,275. Raw data files were extracted and aligned using *apLCMS* R package (17) and further processed through *xMSanalyzer* (18) and corrected for batch effects (*ComBat*). Uniquely detected peaks consisting of *m/z*, retention time (RT) and ion abundance, were referred to as metabolite features. In total, 9,435 features for HILpos and 8,439 for C18neg were yielded.

Structural annotation of compounds of caffeine metabolism was implemented through an integrated cheminformatic strategy. We retrieved a complete set of 22 structures from caffeine metabolism referencing the Kyoto Encyclopedia of Genes and Genomes (KEGG) (19), alongside 1,3,7-trimethyldihydrourate, a novel caffeine metabolite recently discovered through our *in vitro* exposomic platform (20). We first built internal RT-*m/z* libraries respectively for HILpos and C18neg modes, including caffeine and its major metabolites. Meanwhile, to expand the coverage, we leveraged *in silico* cheminformatic analyses for annotating all plausible metabolites involved in caffeine metabolism. Using accurate *m/z*, isotopic ratios, and RT, we annotated detected peaks based on formula prediction (21, 22) and RT estimation by XGBoost algorithm (23), and manually curated based on extensive bioanalytical inferences and expert consultation. Annotation confidence was assigned as Level 1 for features matched with our in-house library, and Level 2 for features with predicted parameters but not validated by authentic chemical standards (24).

A total of 15 features were successfully annotated, corresponding to 12 unique caffeine metabolites. Three metabolites, 5-acetylamino-6-amino-3-methyluracil (AAMU), 1-methylxanthine and 3,6,8-trimethylallantoin, were detected in both HILpos and C18neg modes.

Statistical analysis

In the EPIC4PD, coffee consumers were binned into quartiles based on the distribution in each country (country-specific quartiles), to account for heterogeneity of consumed volume and concentration of coffee between countries (14). Hazard ratios (HRs) and their corresponding 95% confidence intervals (CIs) for PD risk were estimated using Cox proportional hazards models, with age as the underlying time variable. The entry time for all participants was defined as age at recruitment, and the exit time was either the age at diagnosis for PD cases or the last date when follow-up was deemed complete for participants without PD. We also performed analysis using coffee overall, non-country specific, quartiles (based on data from all countries combined). Exposure-response

effect of coffee intake on PD was examined by entering the categorical value of the quartiles (0 for nonconsumers and 1-4 for coffee quartiles) into the model as a continuous term. PD risk was also estimated per 100 ml/d coffee intake. To assess the impact of coffee consumption in the population, the population preventable fraction (PPF) was calculated. PPF is defined as the proportion of PD cases that could be prevented within the population if coffee intake were intervened upon (formula in Text S2).

Age, male gender and smoking are well recognized risk factors for PD (25) (26), and they might influence coffee-consumption habits. Moreover, there might be systematic differences in data collection among study centers. Thus, the main analyses were adjusted for age at recruitment, sex, study center, and smoking status at recruitment (never, former and current smoker). In additional analyses, a set of confounders were also considered, including BMI, alcohol consumption (nonconsumer, 0.1-5, 5-15, 15-30, 30-60, ≥60 g/day], physical activity (inactive, moderately inactive, moderately active, active and not specified), education level (none, primary school, secondary/technical school, longer education and not specified). Age (in years) and BMI (in kg/m^2) were included in the Cox models as continuous variables, while categorical variables were represented using dummy codes in the models ('male' and 'Italy' as reference for sex and study center respectively, lowest level as reference for alcohol consumption, physical activity and education level). None of variables in additional analyses considerably modified the risk estimates (Table S1), and they were thus not included in the final models.

The main analyses were stratified by sex and smoking status to account for possible effect modifications. In the subgroup analysis of women, menopausal status (premenopausal, postmenopausal, perimenopausal and ovariectomy) and history of using hormone therapy (ever used or not) were further adjusted. Possible interactions between sex or smoking status and coffee intake were tested using the likelihood ratio test based on models with and without the interaction terms.

Heterogeneity across countries was investigated using a meta-analytic approach based on HRs of coffee consumers compared with nonconsumers in each country. The I^2 statistic was used to illustrate the proportion of observed variance that reflects true variance among countries rather than sampling error (27). I^2 values of 25%, 50% and 75% represent low, moderate and high levels of heterogeneity. Sensitivity analyses were performed limiting to 'definite' and 'very likely' PD cases (n=314). To rule out possible reverse causality, our analyses were limited to PD cases diagnosed after eight years (median) since recruitment into the cohort. Additionally, we further conducted

analyses after exclusion of cases by consecutive one-year interval of prediagnostic periods (from >0 to >16 years).

To account for the potential role of caffeine in the effects of coffee on PD, stratified analyses were conducted for caffeinated and decaffeinated coffee consumption. Subjects for whom the sum of both coffee subtypes was equal to the total coffee intake were included in stratified analyses. Caffeinated and decaffeinated coffee consumers were divided into country-specific tertiles due to the smaller sample size, and models for caffeinated and decaffeinated coffee were mutually adjusted for one another. Coffee consumers were additionally categorized according to coffee types they consumed (only caffeinated, only decaffeinated, and both types of coffee).

In the nested case-control study, missing values of the detected caffeine metabolites (missing percentage ranging between 0% and 64.1%, Figure S1) that were below limits of detection were imputed using a quantile regression approach for left-censored missing data based on distributions of available values of metabolites, as implemented in *imputeLCMD* R package (28). Correlations among the metabolites and correlations between coffee consumption volume and metabolites were examined by Spearman correlation (*rho*). Ion intensities of metabolites were log2 transformed to reduce influence of extreme values and scaled (divided by standard deviation, SD) to make results of analysis comparable. Conditional logistic regression for the matched case-control sets was applied to estimate odds ratios (ORs) and 95% CIs for associations between caffeine metabolites and PD, adjusting for smoking status. The nested case-control study adopted the same stratified and sensitivity analyses as in the analysis of coffee consumption and PD in the cohort to evaluate the robustness of results.

Results

Study population

Following the application of exclusion criteria, our analysis included a total of 184,024 subjects from the EPIC4PD cohort, with a median follow-up of 13.1 years. Within this cohort, 308 and 285 incident PD cases were recorded among men and women, respectively (Table 1). The age-adjusted incidence rates for individuals aged 65 and older were 134 and 77 per 100,000 person-years for men and women, respectively. The median period between recruitment and PD diagnosis was 8.3 years. The median age at recruitment for subjects with PD was higher than those without PD (61.2 vs. 52.6 years). The prevalence of coffee consumption in the entire EPIC4PD population was 93%. The daily coffee consumption volume was highest in the Netherlands (median 500 ml/d) and lowest in Italy and Spain (median 100 ml/d for both countries) (Table S2). Participants in the highest quartile of coffee intake were more likely to be men, current

smokers, younger and reported higher alcohol consumption (Table S3). In our nested case-control study, which included 351 incident PD cases and 351 matched controls, the demographics, lifestyle factors and coffee consumption were comparable with those observed in the EPIC4PD cohort (Table 1).

Coffee consumption and Parkinson's disease

An inverse exposure-response relationship between coffee consumption and PD was observed (*p*-trend 0.003) with HR of 0.63 (95% CI 0.46-0.88) for the highest quartile of consumers vs. nonconsumers (Table 2). HRs based on overall coffee intake quartiles were similar to those using country-specific quartiles (Table S4). The point estimates of HR for coffee consumers compared to nonconsumers varied between 0.37 and 0.95 across countries, with a minimal heterogeneity noted (I^2=3.3%) (Figure S2). The PPF, with the HR of coffee consumers vs. nonconsumers at 0.72 (95% 0.56-0.94), was 26% (95% CI 6.6%-41%) for coffee consumption in the EPIC4PD population. In sub-analyses, the inverse association was limited to caffeinated coffee consumers (HR for highest tertile vs. nonconsumers 0.57, 95% CI 0.35-0.94; *p*-trend 0.007), and no association was observed for decaffeinated coffee consumption (Table S5).

No obvious difference was noted for associations between men and women (*p* for interaction 0.974), although a statistically significant trend for coffee intake and PD was only found in women (*p*-trend 0.025) (Table 2). Further adjustment for menopause status and hormone use did not materially change the associations among women (Table S6). A slightly stronger association for the highest quartile was observed among hormone never users (HR 0.50, 95% CI 0.26-0.96).

A stronger association between coffee consumption and PD was observed in never smokers (HR for highest quartile vs. nonconsumers 0.59, 95% CI 0.38-0.93), than in former and current smokers (Table 2). Interaction between smoking and coffee intake was not significant (*p* for interaction 0.185). Furthermore, compared with individual effect of smoking and coffee, a more pronounced inverse association was observed for subjects who were both cigarette smokers and coffee drinkers at baseline (HR vs. nonconsumers for both cigarettes and coffee 0.41, 95% CI 0.29-0.59) (Table S7).

In the analyses limited to 281 PD cases diagnosed after eight years of follow-up, the associations between coffee intake and PD were strengthened across all quartiles (HR for highest quartile 0.54, 95% CI 0.35-0.84) (Table 2). Slightly stronger inverse associations with increasing prediagnostic time lags were also reflected when we progressively excluded cases diagnosed within a certain time frame (Figure S3). Estimates based on the analysis limiting to 314 definite and very likely cases were similar to those in the main analysis (Table 2).

Table 1. Baseline characteristics among participants in the EPIC4PD cohort and nested case-control study

Characteristic	EPIC4PD cohort, n=184,024		Nested case-control study, n=702[1]	
	PD cases	Non-cases	PD cases	Controls
	n=593	n=183,431	n=351	n=351
Age at recruitment (years), median (IQR)	61.2 (55.2-65.8)	52.6 (46.7-59.9)	60.7 (54.8-65.6)	60.4 (55.0-65.2)
Age at diagnosis (years), median (IQR)	69.8 (63.6-74.4)	--	68.7 (62.8-74.0)	--
Years between recruitment and diagnosis, median (IQR)	8.3 (4.9-11.5)	--	7.8 (4.6-11.0)	--
Definite and very likely cases, n (%)	314 (53%)	--	188 (54%)	--
Sex, n (%)				
Male	308 (52%)	67,442 (37%)	195 (56%)	195 (56%)
Female	285 (48%)	115,989 (63%)	156 (44%)	156 (44%)
Country[2]				
Italy	64 (11%)	40,111 (22%)	54 (15%)	54 (15%)
Spain	101 (17%)	24,852 (13%)	97 (28%)	97 (28%)
UK	170 (29%)	23,227 (13%)	141 (40%)	141 (40%)
Netherlands	13 (2%)	16,813 (9%)	13 (4%)	13 (4%)
Germany	50 (8%)	25,349 (14%)	46 (13%)	46 (13%)
Sweden	195 (33%)	53,079(29%)	--	--
Coffee consumption at recruitment (ml/d)				
Nonconsumer, n (%)[3]	67 (11%)	12,826 (7%)	45 (12.9%)	36 (10.4%)
Total coffee, median (IQR)[3]	261 (104-475)	286 (113-500)	190 (79-475)	190 (73-475)
Caffeinated coffee, median (IQR)[4]	261 (100-475)	261 (90-475)	190 (60-475)	190 (82-475)
Decaffeinated coffee, median (IQR)[5]	190 (47-332)	62 (25-190)	190 (48-300)	86 (20-273)
Smoking status at recruitment, n (%)[6]				
Never smoker	321 (54%)	85,717 (47%)	183 (52%)	174 (49%)
Former smoker	198 (33%))	53,557 (29%)	115 (33%)	109 (31%)

Current smoker	74 (13%)	44,157 (24%)	43 (12%)	55 (16%)
BMI at recruitment (kg/m^2), median (IQR)	25.8 (23.8-28.5)	25.4 (23.1-28.2)	26.5 (24.2-29.2)	26.0 (23.8-29.1)
Alcohol consumption at recruitment (g/d)[7]				
Nonconsumer, n (%)	118 (20%)	30,994 (17%)	77 (22%)	62 (18%)
Total alcohol, median (IQR)	7.4 (2.7-18.7)	7.4 (2.1-18.8)	9.1 (2.8-24.2)	10.6 (2.6-26.7)
Higher education, n (%)[8]	84 (14%)	32,625 (18%)	41 (12%)	46 (13%)
Physically active, n (%)[9]	24 (4%)	15,674 (9%)	18 (5%)	26 (7%)
Postmenopausal, n (%)[10]	220 (77%)	57,280 (49%)	--	--
Ever use of menopausal hormone therapy, n (%)[10, 11]	59 (21%)	24,388 (21%)	--	--

IQR, interquartile range; BMI, body mass index.

[1] PD cases and controls were matched on age at recruitment, sex, country and fasting status in the nested case-control study

[2] No subjects from Sweden were included in the nested case-control study.

[3] Information on total coffee was missing for 3 PD cases and 4 controls in the nested case-control study.

[4] Information on caffeinated coffee was missing for 68 PD cases and 30,359 participants without PD in EPIC4PD cohort, and for 6 PD cases and 7 controls in the nested case-control study.

[5] Information on decaffeinated coffee was missing for 309 PD cases and 82,974 participants without PD in EPIC4PD cohort, and for 103 PD cases and 104 controls in the nested case-control study.

[6] Information on smoking status was missing for 10 PD cases and 13 controls in the nested case-control study.

[7] Information on alcohol consumption was missing for 3 PD cases and 4 controls in the nested case-control study.

[8] Information on education level was missing for 20 PD cases and 1,729 participants without PD in EPIC4PD cohort, and for 24 PD cases and 17 controls in the nested case-control study.

[9] Information on physical activity was missing for 58 PD cases and 25,590 participants without PD in EPIC4PD cohort, and for 13 PD cases and 10 controls in the nested case-control study.

[10] Only among women.

[11] Information on ever use of menopausal hormone therapy was missing for 55 PD cases and 16,667 participants without PD in EPIC4PD cohort

Table 2. Associations of total coffee consumption and risk of Parkinson's disease in the EPIC4PD cohort

Analysis	Coffee consumption[1]					*p* for trend	Per 100 ml/d
	Nonconsumers	Quartile 1	Quartile 2	Quartile 3	Quartile 4		
All participants n=184,024							
PD cases, n	67	203	145	90	88		
HR (95% CI)[2]	Reference	0.80 (0.61-1.06)	0.71 (0.53-0.96)	0.66 (0.48-0.91)	0.63 (0.46-0.88)	0.003	0.97 (0.94-1.00)
Men n=67,750							
PD cases, n	36	105	74	45	48		
HR (95% CI)[2]	Reference	0.81 (0.56-1.19)	0.68 (0.46-1.02)	0.70 (0.45-1.11)	0.69 (0.44-1.07)	0.090	0.97 (0.93-1.01)
Women n=116,274							
PD cases, n	31	98	71	45	40		
HR (95% CI)[2]	Reference	0.78 (0.52-1.17)	0.75 (0.49-1.15)	0.63 (0.39-1.00)	0.60 (0.37-0.96)	0.025	0.98 (0.93-1.03)
Never smokers n=86,038							
PD cases, n	41	103	83	55	39		
HR (95% CI)[2]	Reference	0.68 (0.47-0.97)	0.72 (0.49-1.04)	0.69 (0.46-1.05)	0.59 (0.38-0.93)	0.107	1.00 (0.95-1.05)
Former smokers n=53,755							
PD cases, n	99		41	24	34		
HR (95% CI)[2]	Reference[3]		0.64 (0.45-0.92)	0.63 (0.39-0.99)	0.78 (0.52-1.16)	0.085	0.96 (0.91-1.01)
Current smokers n=44,231							
PD cases, n	27		21	11	15		
HR (95% CI)[2]	Reference[3]		0.97 (0.55-1.72)	0.62 (0.30-1.30)	0.67 (0.35-1.28)	0.117	0.95 (0.87-1.04)
Late-diagnosed cases[4] n=183,743							
PD cases, n	38	110	69	47	48		
HR (95% CI)[2]	Reference	0.77 (0.53-1.11)	0.60 (0.41-0.90)	0.50 (0.32-0.77)	0.54 (0.35-0.84)	0.001	0.96 (0.91-1.01)
Definite&very likely cases n=183,745							
PD cases, n	33	95	90	52	44		
HR (95% CI)[2]	Reference	0.80 (0.54-1.19)	0.90 (0.60-1.34)	0.72 (0.46-1.13)	0.62 (0.39-0.98)	0.050	0.97 (0.93-1.02)

HR, hazard ratio; CI, confidence interval.

[1] Based on country-specific quartiles for coffee consumers. Quartile cutoffs were 62, 100, 145 ml/day in Italy, 47, 100, 184 ml/day in Spain, 190, 475, 557 ml/day in UK, 375, 500, 750 ml/day in the Netherlands, 261, 392, 573 ml/day in Germany, and 300, 400, 601 ml/day for Sweden.

[2] Cox regression adjusted for age at recruitment, sex (when appropriated), country and smoking status (when appropriated).

[3] Reference category merged with quartile 1 due to low case numbers among nonconsumers.

[4] PD cases diagnosed within 8 years of follow-up were excluded.

Caffeine metabolites and Parkinson's disease

Prediagnostic levels of most caffeine metabolites were positively associated with self-reported coffee volume, as indicated by correlation coefficients ranging from 0.10 to 0.41 (Table 3). Several metabolites including caffeine, theophylline, paraxanthine, AAMU (C18neg) were moderately correlated with each other (correlation coefficient, *rho*>0.4) (Figure S4).

Caffeine and three other metabolites (paraxanthine, theophylline, 1-methyluric acid), were negatively associated with PD risk (OR per SD increase (95% CI) 0.80 (0.67-0.95), 0.82 (0.69-0.96), 0.78 (0.65-0.93), 0.84 (0.72-0.98), respectively) (Table 3). Subtle associations with PD risk, although not statistically significant, were observed for 1,3,7-trimethyldihydrouric acid (OR 0.85, 95% CI 0.72-1.01), 5-acetylamino-6-formylamino-3-methyluracil (AFMU) (OR 1.19, 95% CI 0.99-1.43), and 1,3,7-trimethyl-5-hydroxyisourate (OR 1.19, 95% CI 0.99-1.41). Analyses for caffeine and theophylline were more pronounced among men than women (*p* for interaction by sex 0.020 and 0.011, respectively) (Figure 1). There was no evidence for effect modification by smoking status for the associations between caffeine metabolites and PD (Figure S5), although the inverse association for caffeine was stronger among current smokers (OR 0.55, 95% CI 0.31-0.92) than in noncurrent smokers. Furthermore, sensitivity analyses of limiting to cases diagnosed after eight years since recruitment and cases with high validity did not reveal substantial changes (Figure S6).

Analyses by increasing prediagnostic time periods for caffeine metabolites showed that for caffeine and the major metabolites paraxanthine and theophylline, associations became slightly stronger with increasing prediagnostic time periods (Figure S7). Interestingly, for several downstream metabolites (1-methyluric acid, 1,3,7-trimethyldihydrouric acid, AFMU and 1,3,7-trimethyl-5-hydroxyisourate), the associations became weaker for prediagnostic time periods longer than 10 years.

Table 3. Associations of caffeine metabolites and Parkinson's disease risk in the nested case-control study (n=702)

LC-MS mode	Retention time (s)	*m/z*	Metabolite name	Annotation level[1]	Correlation with coffee[2]	OR (95% CI)[3] Per SD increase of log2 ion intensity
HILIC-positive	195.0877	30.5	Caffeine	1	0.41	0.80 (0.67-0.95)
HILIC-positive	181.0773	31.8	Paraxanthine	2	0.21	0.82 (0.69-0.96)
HILIC-positive	181.0720	32.9	Theophylline	1	0.41	0.78 (0.65-0.93)
HILIC-positive	211.0770	57.2	1,3,7-Trimethyluric acid	2	-0.10	0.92 (0.77-1.10)
HILIC-positive	227.0789	39.5	AFMU	1	-0.16	1.19 (0.99-1.43)
HILIC-positive	199.0821	74.8	AAMU (HILpos)	2	0.12	1.05 (0.89-1.23)
C18-negative	197.0681	31.8	AAMU (C18neg)	1	0.42	0.94 (0.79-1.11)
HILIC-positive	167.0562	31.4	1-Methylxanthine (HILpos)	1	0.41	0.93 (0.78-1.09)
C18-negative	165.0414	33.0	1-Methylxanthine (C18neg)	1	0.44	0.87 (0.73-1.04)
HILIC-positive	153.0408	42.3	Xanthine	1	0.10	1.01 (0.86-1.20)
HILIC-positive	183.0505	69.9	1-Methyluric acid	2	0.13	0.84 (0.72-0.98)
HILIC-positive	201.0892	44.4	3,6,8-Trimethylallantoin (HILpos)	2	0.14	1.00 (0.83-1.21)
C18-negative	199.0782	32.9	3,6,8-Trimethylallantoin (C18neg)	2	-0.08	1.06 (0.90-1.26)
C18-negative	225.0627	33.8	1,3,7-Trimethyl-5-hydroxyisourate	2	-0.16	1.19 (0.99-1.41)
C18-negative	211.0837	33.2	1,3,7-Trimethyldihydrouric acid	2	0.41	0.85 (0.72-1.01)

LC-MS, liquid chromatography coupled with high-resolution mass spectrometry; *m/z*, mass/charge; OR, odds ratio; CI, confidence interval; AFMU, 5-acetylamino-6-formylamino-3-methyluracil; AAMU, 5-acetylamino-6-amino-3-methyluracil.

[1] Features matched with the in-house library were assigned with Level 1 annotation, features matched with predicted chemical retention time with Level 2 annotation.

[2] Spearman correlation coefficient between continuous coffee volume (ml/day) and ion intensity of caffeine metabolites.

[3] Conditional logistic regression for the matched case-control sets, adjusted for smoking status

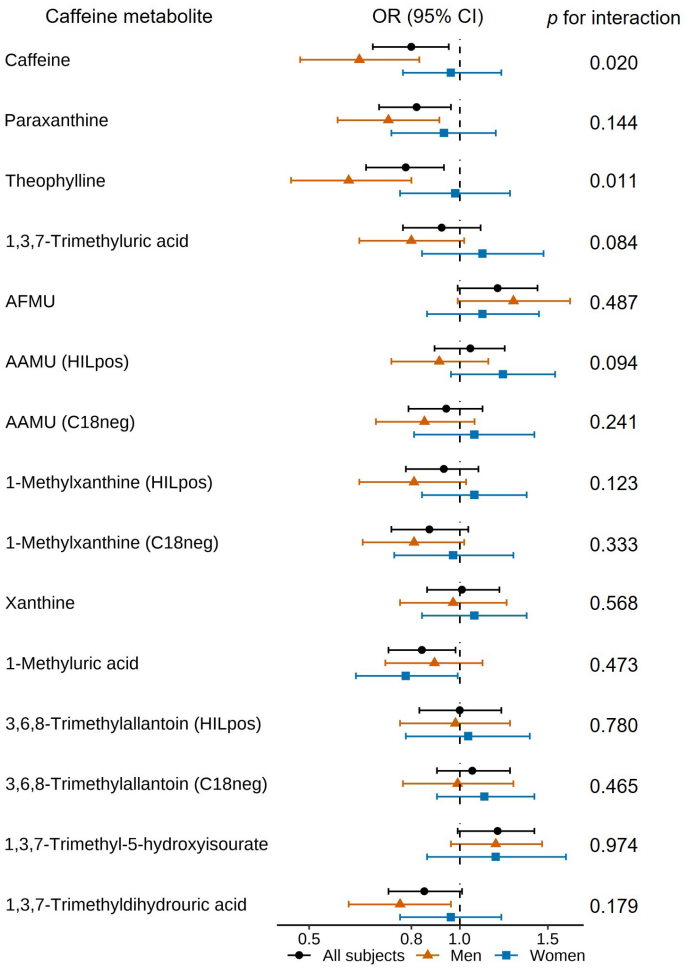


Figure 1. Associations between caffeine metabolites and Parkinson's disease risk among men and women in the nested case-control study (n=702)

OR, odds ratio; CI, confidence interval; AFMU, 5-acetylamino-6-formylamino-3-methyluracil; AAMU, 5-acetylamino-6-amino-3-methyluracil.

ORs and CIs (per SD increase of log2 ion intensity) were calculated by conditional logistic regression for the matched case-control sets, adjusted for smoking status, for men and women separately. P values for interaction of sex and metabolite were estimated by likelihood ratio test based on models with and without the interaction terms.

Discussion

This study demonstrated an inverse association of caffeinated coffee consumption with the risk of PD in one of the largest longitudinal cohorts worldwide with more than 20 years of follow-up. The neuroprotective effects of coffee were exposure-dependent, and individuals in the highest coffee consumption group had nearly 40% lower risk of PD compared with nonconsumers. This observation was strengthened with a comprehensive evaluation of prospectively measured plasma caffeine and its metabolites. These analyses showed strong inverse associations for caffeine and its major metabolites with the risk of PD.

The EPIC4PD population is not a strictly random sample for the entire European population, with women outnumbering men in the cohort. There was still a slight predominance of men among the PD patients. Notably, individuals who developed PD were generally older at the time of recruitment compared to those who did not develop PD. These findings further underscore the important role of aging and male gender for PD risk. Moreover, incidence of PD in the EPIC4PD cohort was comparable to those reported in the North America (162-277 per 100,000 person-years among males, 66-161 per 100,000 person-years among females) (29). Several large US prospective cohorts have reported comparable effect sizes for the highest coffee intake group (with adjusted HR ranging from 0.43 to 0.81 for men, from 0.61 to 0.90 for women), which aligns with our findings (5, 30, 31).

The strength of our study was utilizing objective blood markers for caffeine metabolism, which largely mitigates regional variations on coffee consumption. More notably, our study largely minimizes the possibility of reverse causation by collecting blood samples before PD diagnosis. In contrast, previous studies analyzed caffeine biochemical markers in biosamples from individuals who had been living with PD, with an average disease duration ranging from 6 to 8 years (6-8). This approach introduced potential bias, as these patients might change their coffee-consumption habits due to smell and taste dysfunction. In addition, caffeine is primarily metabolized by CYP1A2, an isoform of the hepatic cytochrome P450 enzyme family (32). Antiparkinsonian drugs such as levodopa has been shown to upregulate CYP1A2 activities, resulting in an increased metabolism of caffeine (33).

Coffee consumption has long been suggested to reduce or delay the development of PD, with caffeine identified as the most likely causal factor (1, 2, 5). Caffeine administration attenuated motor impairment, neuronal death and dopamine depletion in various animal models of PD (34-36). It is believed that caffeine's neuroprotective effects are mainly attributed to the blocking of adenosine 2A receptor (A2AR) (37).

Moreover, two major caffeine metabolites, paraxanthine and theophylline, have demonstrated the ability to mitigate symptoms in PD animal models (38) (39). These neuroprotective effects align with our findings, which revealed an inverse association between caffeine, paraxanthine, theophylline and the incidence of PD. More important, these associations tended to be marginally stronger with increasing prediagnostic period, indicating minimal effect due to reverse causality.

On the other hand, other caffeine-derived metabolites, specifically AFMU, 1,3,7-trimethyldihydrouric acid and 1,3,7-trimethyl-5-hydroxyisourate, exhibited altered associations with PD, although these association did not reach statistical significance. Notably, these associations were absent when the time between biomarker assessment and PD diagnosis was longer than 10 years. The molecular underpinnings of these metabolites remain elusive, partially due to lack of relevant publications, warranting investigations to decipher changes in other downstream caffeine metabolites.

Some studies have reported effect modification of estrogen and tobacco on the beneficial effects of caffeine on PD (5, 30, 31). Estrogen and caffeine are known to competitively metabolized by CYP1A2, resulting in an inhibitory effect on caffeine metabolism (40). Conversely, tobacco has been shown to strongly induce the CYP1A2 enzyme, thereby increasing the metabolism of caffeine in smokers (41). Because metabolites of caffeine, such as paraxanthine and theophylline, are also A2AR antagonists and neuroprotectants, the net effect of perturbed caffeine metabolism is difficult to predict. In this study, we did not observed statistically significant effect modification by hormonal use and smoking. However, more-marked trends toward a reduced risk of PD were observed among women who never used postmenopausal hormones with increasing coffee intake and among individuals who were both smokers and coffee drinkers. Our results might be hampered by limited power, and results from literature on the estrogen- and smoking-specific modification of caffeine and PD are inconsistent. The interaction mechanism needs to be elucidated further in experimental studies.

As a widely consumed beverage, coffee has a discernible public health impact on PD prevention, even though its effect size is relatively modest. However, the therapeutic effectiveness of caffeine and its metabolites for alleviating parkinsonian symptoms has been limited in current clinical trials (9) (10). Together with our findings, these results suggest that caffeine may exert a neuroprotective effect during the prodromal phase, rather than after the onset of classical motor PD symptoms. Therefore, the administration of caffeine to individuals at high risk for PD, such as those with rapid eye movement sleep behavior disorder, which is the strongest indicator of prodromal PD (42), could be a promising approach to stop or delay the disease deterioration. In

parallel, it remains to be determined if advocating for public intervention with promoting increased coffee consumption or caffeine supplements is appropriate due to the potential side effects of caffeine.

We acknowledged several limitations in our study. First, diet information and blood samples were collected on average eight years prior to PD diagnosis. Participants' coffee-consumption behavior can change due to an altered sense of taste and smell, which may occur up to 10 years before the onset of motor symptoms (43). Therefore, we cannot fully exclude that coffee-consumption habits were secondary to PD-related symptoms. However, as consistent inverse associations between coffee/caffeine and PD risk were found as early as 12 years prior to disease diagnosis, residual reverse causality does not seem to be substantial. Second, the observed associations could be confounded by unmeasured factors. Genes *CYP1A2* and *ADORA2A,* which encode caffeine metabolism enzyme CYP1A2 and action target A2AR respectively, might modify the protective effect of caffeine on PD (44). Environmental toxicants such as pesticides have been suggested to be associated with an increased risk of PD (45). However, to explain a 40% decrease in PD risk, any confounder must confer a risk ratio greater than 2.7 with both coffee consumption and PD risk (46). None of the known or suspected risk factors for PD demonstrate such strong associations. Thus, the chance that residual confounding fully explains our presented results is very limited. Third, coffee consumption assessed by dietary questionnaires might be subject to measurement error, in particular because of different coffee types and caffeine concentration in different countries. However, the dietary questionnaire has been validated in our cohort (11), and coffee intake was overall positively correlated with caffeine contents, although the extent varied across countries (Figure S8). In addition, non-differential measurement error would likely bias results towards the null, so the true coffee-PD associations could be stronger than the observed ones. Lastly, subjects with missing data on coffee consumption and smoking status exhibited notable differences from the entire EPIC4PD population regarding demographics and recruitment countries (Table S8). However, given the relatively small percentage of missing group (approximately 4%), the subjects included in our study remained a representative sample of the EPIC4PD cohort.

Conclusion

In summary, our study validated the protective effect of caffeine on PD risk in a large prospective cohort, and further confirmed the etiological role of caffeine utilizing biosamples before PD diagnosis in an untargeted exposomic framework. Our findings on the protective action of caffeine and its main metabolites provide insights into the etiology and prevention of PD.

References

1. Ross GW, Abbott RD, Petrovitch H, et al. Association of coffee and caffeine intake with the risk of Parkinson disease. Jama. 2000;283(20):2674-9.

2. Ascherio A, Zhang SM, Hernán MA, et al. Prospective study of caffeine consumption and risk of Parkinson's disease in men and women. Ann Neurol. 2001;50(1):56-63.

3. Hu G, Bidel S, Jousilahti P, Antikainen R, Tuomilehto J. Coffee and tea consumption and the risk of Parkinson's disease. Mov Disord. 2007;22(15):2242-8.

4. Tan LC, Koh WP, Yuan JM, et al. Differential effects of black versus green tea on risk of Parkinson's disease in the Singapore Chinese Health Study. Am J Epidemiol. 2008;167(5):553-60.

5. Palacios N, Gao X, McCullough ML, et al. Caffeine and risk of Parkinson's disease in a large cohort of men and women. Mov Disord. 2012;27(10):1276-82.

6. Takeshige-Amano H, Saiki S, Fujimaki M, et al. Shared Metabolic Profile of Caffeine in Parkinsonian Disorders. Mov Disord. 2020;35(8):1438-47.

7. Fujimaki M, Saiki S, Li Y, et al. Serum caffeine and metabolites are reliable biomarkers of early Parkinson disease. Neurology. 2018;90(5):e404-e11.

8. Hatano T, Saiki S, Okuzumi A, Mohney RP, Hattori N. Identification of novel biomarkers for Parkinson's disease by metabolomic technologies. J Neurol Neurosurg Psychiatry. 2016;87(3):295-301.

9. Postuma RB, Anang J, Pelletier A, et al. Caffeine as symptomatic treatment for Parkinson disease (Café-PD): A randomized trial. Neurology. 2017;89(17):1795-803.

10. Kulisevsky J, Barbanoj M, Gironell A, Antonijoan R, Casas M, Pascual-Sedano B. A double-blind crossover, placebo-controlled study of the adenosine A2A antagonist theophylline in Parkinson's disease. Clin Neuropharmacol. 2002;25(1):25-31.

11. Riboli E, Hunt KJ, Slimani N, et al. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. Public health nutrition. 2002;5(6b):1113-24.

12. Gallo V, Brayne C, Forsgren L, et al. Parkinson's Disease Case Ascertainment in the EPIC Cohort: The NeuroEPIC4PD Study. Neurodegener Dis. 2015;15(6):331-8.

13. Slimani N, Kaaks R, Ferrari P, et al. European Prospective Investigation into Cancer and Nutrition (EPIC) calibration study: rationale, design and population characteristics. Public health nutrition. 2002;5(6B):1125-45.

14. Gunter MJ, Murphy N, Cross AJ, et al. Coffee Drinking and Mortality in 10 European Countries: A Multinational Cohort Study. Ann Intern Med. 2017;167(4):236-47.

15. Soltow QA, Strobel FH, Mansfield KG, Wachtman L, Park Y, Jones DP. High-performance metabolic profiling with dual chromatography-Fourier-transform mass spectrometry (DC-FTMS) for study of the exposome. Metabolomics. 2013;9(1 Suppl):S132-s43.

16. Liu KH, Walker DI, Uppal K, et al. High-Resolution Metabolomics Assessment of Military Personnel: Evaluating Analytical Strategies for Chemical Detection. J Occup Environ Med. 2016;58(8 Suppl 1):S53-61.

17. Yu T, Park Y, Johnson JM, Jones DP. apLCMS--adaptive processing of high-resolution LC/MS data. Bioinformatics. 2009;25(15):1930-6.

18. Uppal K, Soltow QA, Strobel FH, et al. xMSanalyzer: automated pipeline for improved feature detection and downstream analysis of large-scale, non-targeted metabolomics data. BMC Bioinformatics. 2013;14:15.

19. Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M. KEGG for taxonomy-based analysis of pathways and genomes. Nucleic Acids Res. 2023;51(D1):D587-d92.

20. Liu KH, Lee CM, Singer G, et al. Large scale enzyme based xenobiotic identification for exposomics. Nat Commun. 2021;12(1):5418.

21. Tsugawa H, Kind T, Nakabayashi R, et al. Hydrogen Rearrangement Rules: Computational MS/MS Fragmentation and Structure Elucidation Using MS-FINDER Software. Anal Chem. 2016;88(16):7946-58.

22. Kind T, Fiehn O. Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. BMC Bioinformatics. 2007;8:105.

23. Bonini P, Kind T, Tsugawa H, Barupal DK, Fiehn O. Retip: Retention Time Prediction for Compound Annotation in Untargeted Metabolomics. Anal Chem. 2020;92(11):7515-22.

24. Schrimpe-Rutledge AC, Codreanu SG, Sherrod SD, McLean JA. Untargeted Metabolomics Strategies-Challenges and Emerging Directions. J Am Soc Mass Spectrom. 2016;27(12):1897-905.

25. Ascherio A, Schwarzschild MA. The epidemiology of Parkinson's disease: risk factors and prevention. The Lancet Neurology. 2016;15(12):1257-72.

26. Gallo V, Vineis P, Cancellieri M, et al. Exploring causality of the association between smoking and Parkinson's disease. Int J Epidemiol. 2019;48(3):912-25.

27. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Statistics in medicine. 2002;21(11):1539-58.

28. Lazer C. imputeLCMD: A Collection of Methods for Left-Censored Missing Data Imputation. R package version 2.0. https://CRAN.R-project.org/package=imputeLCMD. 2015.

29. Willis AW, Roberts E, Beck JC, et al. Incidence of Parkinson disease in North America. NPJ Parkinsons Dis. 2022;8(1):170.

30. Liu R, Guo X, Park Y, et al. Caffeine intake, smoking, and risk of Parkinson disease in men and women. Am J Epidemiol. 2012;175(11):1200-7.

31. Kim IY, O'Reilly É J, Hughes KC, Gao X, Schwarzschild MA, Ascherio A. Differences in Parkinson's Disease Risk with Caffeine Intake and Postmenopausal Hormone Use. J Parkinsons Dis. 2017;7(4):677-84.

32. Arnaud MJ. Pharmacokinetics and metabolism of natural methylxanthines in animal and man. Handb Exp Pharmacol. 2011(200):33-91.

33. Nehlig A. Interindividual Differences in Caffeine Metabolism and Factors Driving Caffeine Consumption. Pharmacol Rev. 2018;70(2):384-411.

34. Kachroo A, Irizarry MC, Schwarzschild MA. Caffeine protects against combined paraquat and maneb-induced dopaminergic neuron degeneration. Exp Neurol. 2010;223(2):657-61.

35. Sonsalla PK, Wong LY, Harris SL, et al. Delayed caffeine treatment prevents nigral dopamine neuron loss in a progressive rat model of Parkinson's disease. Exp Neurol. 2012;234(2):482-7.

36. Khadrawy YA, Salem AM, El-Shamy KA, Ahmed EK, Fadl NN, Hosny EN. Neuroprotective and Therapeutic Effect of Caffeine on the Rat Model of Parkinson's Disease Induced by Rotenone. J Diet Suppl. 2017;14(5):553-72.

37. Chen JF, Xu K, Petzer JP, et al. Neuroprotection by caffeine and A(2A) adenosine receptor inactivation in a model of Parkinson's disease. J Neurosci. 2001;21(10):Rc143.

38. Xu K, Xu YH, Chen JF, Schwarzschild MA. Neuroprotection by caffeine: time course and role of its metabolites in the MPTP model of Parkinson's disease. Neuroscience. 2010;167(2):475-81.

39. Pardo M, Paul NE, Collins-Praino LE, Salamone JD, Correa M. The non-selective adenosine antagonist theophylline reverses the effects of dopamine antagonism on tremor, motor activity and effort-based decision-making. Pharmacol Biochem Behav. 2020;198:173035.

40. Pollock BG, Wylie M, Stack JA, et al. Inhibition of caffeine metabolism by estrogen replacement therapy in postmenopausal women. J Clin Pharmacol. 1999;39(9):936-40.

41. Benowitz NL, Peng M, Jacob P, 3rd. Effects of cigarette smoking and carbon monoxide on chlorzoxazone and caffeine metabolism. Clin Pharmacol Ther. 2003;74(5):468-74.

42. Postuma RB, Pelletier A, Berg D, Gagnon JF, Escudier F, Montplaisir J. Screening for prodromal Parkinson's disease in the general community: a sleep-based approach. Sleep Med. 2016;21:101-5.

43. Oppo V, Melis M, Melis M, Tomassini Barbarossa I, Cossu G. "Smelling and Tasting" Parkinson's Disease: Using Senses to Improve the Knowledge of the Disease. Front Aging Neurosci. 2020;12:43.

44. Popat RA, Van Den Eeden SK, Tanner CM, et al. Coffee, ADORA2A, and CYP1A2: the caffeine connection in Parkinson's disease. Eur J Neurol. 2011;18(5):756-65.

45. Chambers-Richards T, Su Y, Chireh B, D'Arcy C. Exposure to toxic occupations and their association with Parkinson's disease: a systematic review with meta-analysis. Rev Environ Health. 2023;38(1):65-83.

46. VanderWeele TJ, Ding P. Sensitivity Analysis in Observational Research: Introducing the E-Value. Ann Intern Med. 2017;167(4):268-74.

Chapter 5 Supplemental materials

Text S1. Metabolomics method for caffeine metabolite measurement

To profile circulating caffeine metabolites in EPIC4PD, we performed global metabolomics analysis of all plasma samples using a liquid chromatography-high resolution mass spectrometry (LC-HRMS)-based platform that was established and optimized in our prior work (1, 2). Further *de novo* structural annotation was conducted for untargeted, semi-quantitative profiling of caffeine pathway metabolites through an integrated cheminformatic approach.

Sample preparation

Blood plasma samples were thawed on ice; each 50 µL was aliquoted and extracted into 100 µL cold acetonitrile (ACN) spiked with internal standard mix (ACN:plasma, 2:1, *v/v*). The extracts were cold incubated under -20 ºC before centrifugation for protein and particulate removal. Ten µL of the supernatant was injected on LC-MS, resulting in ~3.33 µL raw plasma loaded on column. Sample extracts were injected in triplicate alongside intermittent injection of quality control (QC) extracts of NIST1953 and BioIVT plasma, obtained respectively from National Institute of Standards and Technology (NIST) (Gaithersburg, MD, USA) and Bioreclamation Inc. (New Cassel, NY, USA).

Instrumental analysis

The injected extracts were chromatographically separated, ionized, and analyzed on a Thermo Fisher Scientific Vanquish dual chromatograph coupled to a high-resolution accurate-mass (HRAM) Orbitrap Exploris™ 240 mass spectrometer interfaced with a heated electrospray ionization (ESI) source (Waltham, MA, USA). To maximize detection of polar and nonpolar metabolites, two complementary analyses were performed, namely hydrophilic interaction liquid chromatography (HILIC)-ESI(+) and reverse-phase chromatography (RPC)-ESI(-), which were further referred to as "HILpos" and "C18neg," both operated in full scan mode at 120,000 mass resolution (FWHM) with a mass-to-charge (*m/z*) range of 85-1,275. Stringent quality assurance and quality control (QA/QC) procedures were applied, including sample randomization, timely mass calibration, and routine check of instrumental sensitivity. For this study, a total of 4,872 *.RAW data were acquired (18 batches in total), covering triplicate sample injections and intermittent QC injections.

Data processing

The acquired *.RAW data were converted to *.mzXML format using ProteoWizard msConvert and subject to feature extraction and sample alignment using apLCMS package of R (3). Crude alignment feature tables were yielded separately for HILpos and C18neg, both consisting of uniquely detected features (*m/z*, retention time) and their

ion abundance (peak area) across samples. The feature tables were processed through xMSanalyzer (4) where ion abundances were median summarized, cleaned up by QC, and corrected for batch effects (*ComBat*), yielding 9,435 ion features for HILpos and 8,439 for C18neg for statistical and informatic analyses.

Compound identification

Structural annotation of compounds of caffeine metabolism was implemented through an integrated cheminformatic strategy. We first built internal RT-*m/z* libraries respectively for HILpos and C18neg modes using hundreds of authentic chemical standards purchased from IROA Technologies (Chapel Hill, NC, USA) and Sigma Aldrich (St. Louis, MO, USA), from which caffeine and its major metabolites were included for a quick and reliable identification. Meanwhile, to expand the coverage, we leveraged *in silico* cheminformatic analyses for annotating all plausible metabolites involved in caffeine metabolism. We first retrieved a complete set of 22 structures from caffeine metabolism (map00232, version 11-30-18) referencing the Kyoto Encyclopedia of Genes and Genomes (KEGG) (5), alongside 1,3,7-trimethyldihydrourate, a novel caffeine metabolite recently discovered through our *in vitro* exposomic platform (6). Using accurate *m/z*, isotopic ratios, and chromatographic retention time (RT), we annotated detected peaks based on formula prediction (Seven Golden Rules) (7, 8) and chromatographic retention time estimation (Retip, XGBoost algorithm) (9), and manually curated based on extensive bioanalytical inferences and expert consultation. Level of annotation confidence was assigned based on the Schymanski scale for LC-HRMS metabolomics practice (10, 11).

References

1. Soltow QA, Strobel FH, Mansfield KG, Wachtman L, Park Y, Jones DP. High-performance metabolic profiling with dual chromatography-Fourier-transform mass spectrometry (DC-FTMS) for study of the exposome. Metabolomics. 2013;9(1 Suppl):S132-s43.

2. Liu KH, Walker DI, Uppal K, Tran V, Rohrbeck P, Mallon TM, et al. High-Resolution Metabolomics Assessment of Military Personnel: Evaluating Analytical Strategies for Chemical Detection. J Occup Environ Med. 2016;58(8 Suppl 1):S53-61.

3. Yu T, Park Y, Johnson JM, Jones DP. apLCMS--adaptive processing of high-resolution LC/MS data. Bioinformatics. 2009;25(15):1930-6.

4. Uppal K, Soltow QA, Strobel FH, Pittard WS, Gernert KM, Yu T, et al. xMSanalyzer: automated pipeline for improved feature detection and downstream analysis of large-scale, non-targeted metabolomics data. BMC Bioinformatics. 2013;14:15.

5. Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M. KEGG for taxonomy-based analysis of pathways and genomes. Nucleic Acids Res. 2023;51(D1):D587-d92.

6. Liu KH, Lee CM, Singer G, Bais P, Castellanos F, Woodworth MH, et al. Large scale enzyme based xenobiotic identification for exposomics. Nat Commun. 2021;12(1):5418.

7. Kind T, Fiehn O. Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. BMC Bioinformatics. 2007;8:105.

8. Tsugawa H, Kind T, Nakabayashi R, Yukihira D, Tanaka W, Cajka T, et al. Hydrogen Rearrangement Rules: Computational MS/MS Fragmentation and Structure Elucidation Using MS-FINDER Software. Anal Chem. 2016;88(16):7946-58.

9. Bonini P, Kind T, Tsugawa H, Barupal DK, Fiehn O. Retip: Retention Time Prediction for Compound Annotation in Untargeted Metabolomics. Anal Chem. 2020;92(11):7515-22.

10. Schymanski EL, Jeon J, Gulde R, Fenner K, Ruff M, Singer HP, et al. Identifying small molecules via high resolution mass spectrometry: communicating confidence. Environ Sci Technol. 2014;48(4):2097-8.

11. Schrimpe-Rutledge AC, Codreanu SG, Sherrod SD, McLean JA. Untargeted Metabolomics Strategies-Challenges and Emerging Directions. J Am Soc Mass Spectrom. 2016;27(12):1897-905.

Text S2. Formula of population preventable fraction (PPF)

$$PPF = P_e(1 - HR) \text{ (1)}$$

P_e , the prevalence of exposure. The prevalence of coffee consumption (coffee consumers) among the included EPIC4PD participants was 93.0%.

HR, hazard ratio of PD risk for coffee consumers compared with nonconsumers, 0.72 (95% confidence interval 0.56-0.94).

Reference:

1. Strain T, Brage S, Sharp SJ, Richards J, Tainio M, Ding D, et al. Use of the prevented fraction for the population to determine deaths averted by existing prevalence of physical activity: a descriptive study. Lancet Glob Health. 2020;8(7):e920-e30.

Table S1. Cox regression models additionally adjusting for other covariates

Adjusted covariate[1]	Coffee consumption HR (95% CI)[2]			
	Quartile 1	Quartile 2	Quartile 3	Quartile 4
BMI	0.80 (0.61-1.06)	0.72 (0.53-0.96)	0.66 (0.48-0.92)	0.64 (0.46-0.88)
Alcohol consumption	0.80 (0.60-1.06)	0.71 (0.53-0.96)	0.66 (0.47-0.91)	0.63 (0.46-0.88)
Physical activity	0.80 (0.61-1.06)	0.70 (0.53-0.94)	0.66 (0.48-0.92)	0.64 (0.46-0.89)
Education level	0.80 (0.61-1.06)	0.71 (0.53-0.96)	0.66 (0.48-0.92)	0.64 (0.46-0.88)

HRs and 95% CIs from the main analysis for quartile 1-4 were 0.80 (0.61-1.06), 0.71 (0.53-0.96), 0.66 (0.48-0.91), 0.63 (0.46-0.88), respectively.

[1] Besides age at recruitment, sex, country and smoking status, the main analysis was additionally adjusted for the covariate one by one.

[2] Based on country-specific quartiles for coffee consumers. Quartile cutoffs were 62, 100, 145 ml/day in Italy, 47, 100, 184 ml/day in Spain, 190, 475, 557 ml/day in UK, 375, 500, 750 ml/day in the Netherlands, 261, 392, 573 ml/day in Germany, and 300, 400, 601 ml/day for Sweden.

Table S2. Self-reported coffee consumption for each country

Country	Participants, n	PD cases, n	Coffee nonconsumers, n (%)	Coffee volume[1], median (IQR)
Italy	40,175	64	3,618 (9)	100 (62-145)
Spain	24,953	101	2,635 (10)	100 (47-184)
UK	23,397	170	2,781 (11)	475 (190-557)
Netherlands	16,826	13	466 (3)	500 (375-750)
Germany	25,399	50	1,280 (5)	392 (261-573)
Sweden	53,274	195	2,113 (4)	400 (300-601)

IQR, interquartile range.

[1] ml/day, for coffee consumers

Table S3. Selected characteristics in coffee consumption groups

	Coffee consumption[1]				
	Nonconsumers	Quartile 1	Quartile 2	Quartile 3	Quartile 4
Participants, n	12,893	48,593	44,900	38,566	39,072
Age at recruitment (years), median (IQR)	52.9 (46.4-60.2)	54.8 (48.2-61.6)	53.3 (47.4-60.1)	51.5 (45.7-58.8)	50.7 (45.5-57.3)
Male, %	34	36	37	36	39
Current smokers at recruitment, %	17	17	20	28	36
Alcohol consumption at recruitment (g/d), median (IQR)	6.5 (1.6-18.8)	7.4 (2.2-18.6)	6.5 (1.9-17.7)	7.6 (2.3-19.3)	8.1 (2.4-19.3)
BMI (kg/m^2), median (IQR)	25.7 (23.1-28.7)	25.3 (22.9-28.1)	25.4 (23.1-28.1)	25.4 (23.1-28.1)	25.6 (23.2-28.4)
Physically active, %	8	8	8	9	10

IQR, interquartile range.

[1] Based on country-specific quartiles for coffee consumers. Quartile cutoffs were 62, 100, 145 ml/day in Italy, 47, 100, 184 ml/day in Spain, 190, 475, 557 ml/day in UK, 375, 500, 750 ml/day in the Netherlands, 261, 392, 573 ml/day in Germany, and 300, 400, 601 ml/day for Sweden.

Table S4. Comparison of results based on country-specific and overall coffee intake quartiles

Analysis	Coffee consumption[2]				
	Nonconsumers	Quartile 1	Quartile 2	Quartile 3	Quartile 4
Country-specific quartiles[1]					
PD cases, n	67	203	145	90	88
HR (95% CI)[2]	Reference	0.80 (0.61-1.06)	0.71 (0.53-0.96)	0.66 (0.48-0.91)	0.63 (0.46-0.88)
Overall quartiles[3]					
PD cases, n	67	135	136	154	101
HR (95% CI)[2]	Reference	0.78 (0.58-1.06)	0.74 (0.55-0.99)	0.68 (0.50-0.92)	0.65 (0.47-0.91)

HR, hazard ratio; CI, confidence interval.

[1] Based on country-specific quartiles for coffee consumers. Quartile cutoffs were 62, 100, 145 ml/day in Italy, 47, 100, 184 ml/day in Spain, 190, 475, 557 ml/day in UK, 375, 500, 750 ml/day in Netherlands, 261, 392, 573 ml/day in Germany, and 300, 400, 601 ml/day for Sweden.

[2] Cox regression adjusted for age at recruitment, sex, country and smoking status.

[3] Based on overall quartiles (based on data from all countries combined) for coffee consumers. Quartile cutoffs were 113, 286, 500 ml/day.

Table S5. Associations of caffeinated and decaffeinated coffee consumption and Parkinson's disease risk

Coffee type	PD cases, n	HR (95% CI)[1]
Caffeinated coffee[2]		
Nonconsumers	55	Reference
Tertile 1	111	1.09 (0.78-1.52)
Tertile 2	62	0.77 (0.53-1.13)
Tertile 3	26	0.57 (0.35-0.94)
p for trend[3]		0.007
Decaffeinated coffee[4]		
Nonconsumers	180	Reference
Tertile 1	33	0.90 (0.61-1.32)
Tertile 2	22	1.04 (0.66-1.66)
Tertile 3	19	1.34 (0.79-2.26)
p for trend[3]		0.586
Coffee combination		
Nonconsumers for all coffee	31	Reference
Only caffeinated coffee	149	0.81 (0.55-1.20)
Both caffeinated and decaffeinated coffee	50	0.92 (0.57-1.49)
Only decaffeinated coffee	24	0.92 (0.54-1.57)

From the EPIC4PD cohort, subjects for whom the sum of both coffee subtypes was equal to the total coffee intake were included (79,689 participants and 254 incident PD cases) in the analyses.

HR, hazard ratio; CI, confidence interval.

[1] Cox regression adjusted for age at recruitment, sex, country, smoking status, caffeinated and decaffeinated coffee for one another.

[2] Based on country-specific tertiles for caffeinated coffee consumers. Quartile cutoffs were 60, 116 ml/day in Italy, 190, 475 ml/day in UK, 312, 450 ml/day in the Netherlands, 261, 458 ml/day in Germany.

[3] Coffee intake tertiles were entered into model as continuous term.

[4] Based on country-specific tertiles for decaffeinated coffee consumers. Quartile cutoffs were 8, 27 ml/day in Italy, 190, 475 ml/day in UK, 50, 100 ml/day in Netherlands, 19, 203 ml/day in Germany.

Table S6. Associations of coffee intake and Parkinson's disease according to hormone therapy use

Analysis	Coffee consumption[1]					*p* for trend[2]
	Nonconsumers	Quartile 1	Quartile 2	Quartile 3	Quartile 4	
All women, n=116,274						
PD cases, n	31	98	71	45	40	
HR (95% CI)[3]	Reference	0.77 (0.51-1.16)	0.74 (0.48-1.13)	0.62 (0.39-0.99)	0.59 (0.37-0.96)	0.025
Postmenopausal hormone never users, n=33,935						
PD cases, n	65		39	14	11	
HR (95% CI)[4, 5]	Reference		0.95 (0.64-1.43)	0.61 (0.34-1.10)	0.50 (0.26-0.96)	0.007
Postmenopausal hormone users, n=16,090						
PD cases, n	26		12	5	8	
HR (95% CI)[4, 5]	Reference		0.78 (0.39-1.55)	0.45 (0.17-1.19)	0.77 (0.35-1.73)	0.291

[1] Based on country-specific quartiles for coffee consumers. Quartile cutoffs were 62, 100, 145 ml/day in Italy, 47, 100, 184 ml/day in Spain, 190, 475, 557 ml/day in UK, 375, 500, 750 ml/day in Netherlands, 261, 392, 573 ml/day in Germany, and 300, 400, 601 ml/day for Sweden.

[2] Coffee intake quartiles were entered into model as continuous term.

[3] Cox regression adjusted for age at recruitment, country, smoking status, menopausal status and ever use of hormone therapy.

[4] Cox regression adjusted for age at recruitment, country and smoking status.

[5] Reference category merged with quartile 1 due to low case numbers among nonconsumers.

Table S7. Joint associations of coffee drinking and smoking in relation to Parkinson's disease risk

Exposure group	Participants, n	PD cases, n	HR (95% CI)[1]
Nonconsumers of both cigarettes and coffee	10,652	59	Reference
Consumers only of coffee	129,141	460	0.75 (0.57, 0.99)
Consumers only of cigarettes	2,241	8	0.86 (0.41, 1.80)
Consumers of both cigarettes and coffee	41,990	66	0.41 (0.29, 0.59)

HR, hazard ratio; CI, confidence interval.

[1] Cox regression adjusted for age at recruitment, sex and country.

Table S8. Characteristics of EPIC4PD cohort, subjects included and subjects with missing data

Characteristics	Whole EPIC4PD n=192,980	Subjects included n=184,024	Subjects with missing data on coffee and smoking n=8,484
Age at recruitment (years), median (IQR)	52.8 (46.9-60.0)	52.6 (46.7-60.0)	56.8 (49.2-65.7)
Male, n (%)	71,893 (37%)	67,750 (37%)	3,887 (46%)
Country, n (%)			
Italy	41,140 (21%)	40,175 (22%)	904 (11%)
Spain	25,017 (13%)	24,953 (13%)	16 (0.2%)
UK	30,440 (16%)	23,397 (13%)	6,931 (82%)
Netherlands	17,031 (9%)	16,826 (9%)	200 (2%)
Germany	25,538 (13%)	25,399 (14%)	75 (0.8%)
Sweden	53,814 (28%)	53,274 (29%)	358 (4%)

IQR, interquartile range.

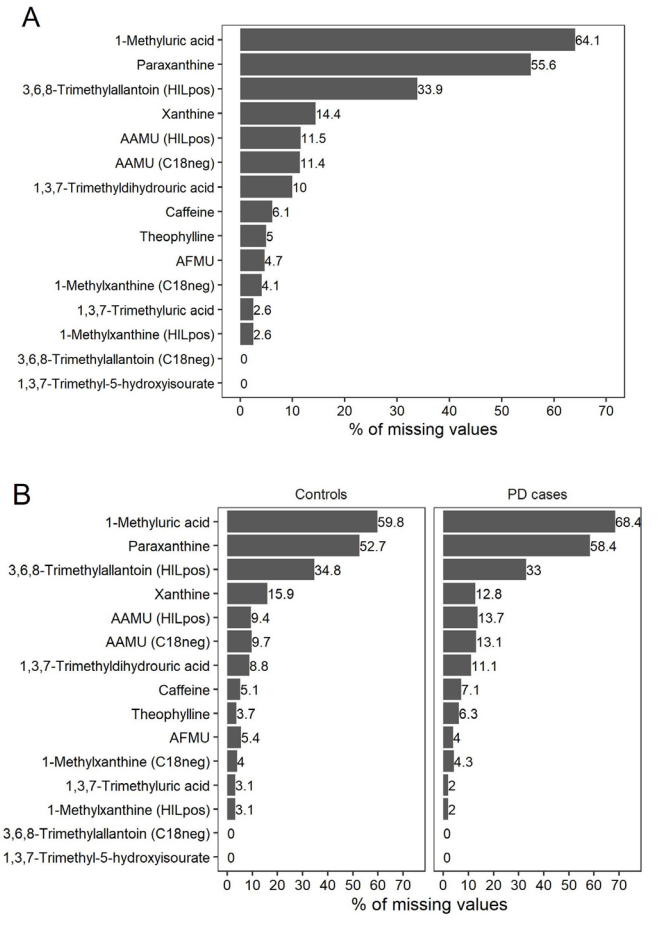


Figure S1. Percentage of missing values for each caffeine metabolite

Missing percentage was calculated among all subjects (A) and among PD cases and controls separately (B).

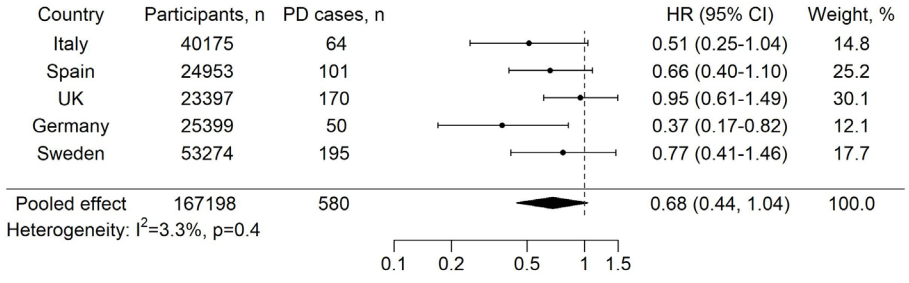


Figure S2. Associations of coffee consumption and Parkinson's disease for each country

HR, hazard ratio; CI, confidence interval. HRs and CIs were calculated for coffee consumers vs. nonconsumers, from Cox regressions adjusted for age at recruitment, sex, country, smoking status. Netherlands was not included because there was no nonconsumer in the country.

p-value from Cochran's Q test=0.4.

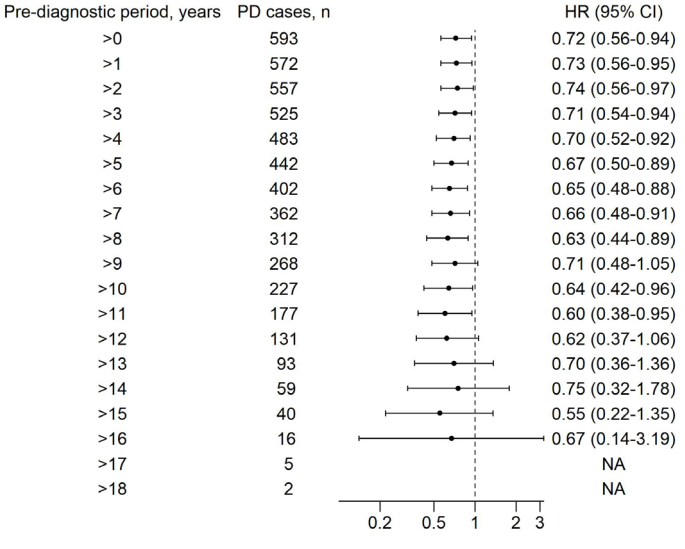


Figure S3. Associations of coffee consumption and Parkinson's disease in different pre-diagnostic periods

HR, hazard ratio; CI, confidence interval. For PD cases, periods between recruitment and diagnosis (pre-diagnostic periods) ranged from <1 to 18 years. HRs and CIs were calculated for coffee consumers vs. nonconsumers, from Cox regressions adjusted for age at recruitment, sex, country, smoking status.

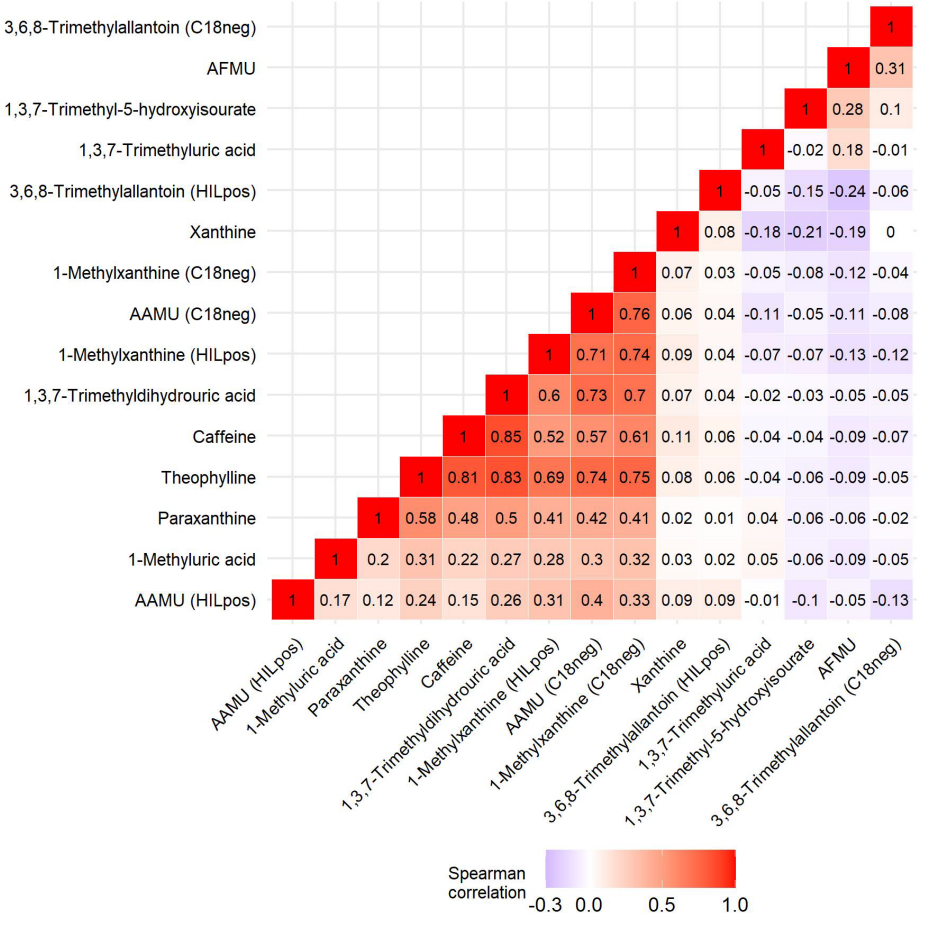


Figure S4. Correlation matrix of caffeine metabolites

Values in tiles were Spearman correlation coefficients for corresponding two metabolites

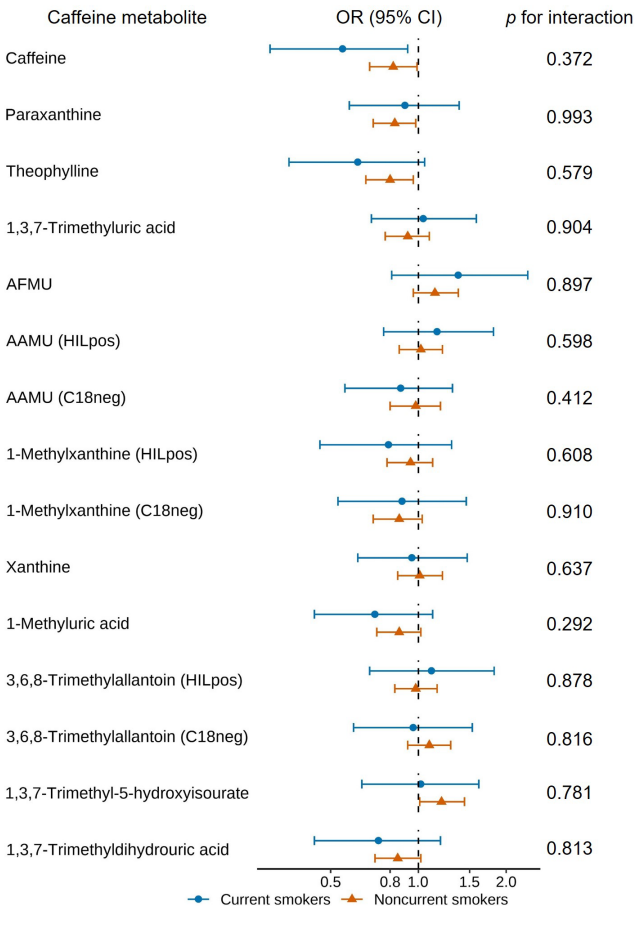


Figure S5. Associations between caffeine metabolites and Parkinson's disease among current and noncurrent smokers

OR, odds ratio; CI, confidence interval; AFMU, 5-Acetylamino-6-formylamino-3-methyluracil; AAMU, 5-acetylamino-6-amino-3-methyluracil.

ORs and CIs (per SD increase of log2 ion intensity) were calculated by unconditional logistic regression adjusted for age at recruitment, sex and study center for current and noncurrent smokers separately. P values for interaction of smoking status and metabolite were estimated by likelihood ratio test based on models with and without the interaction terms.

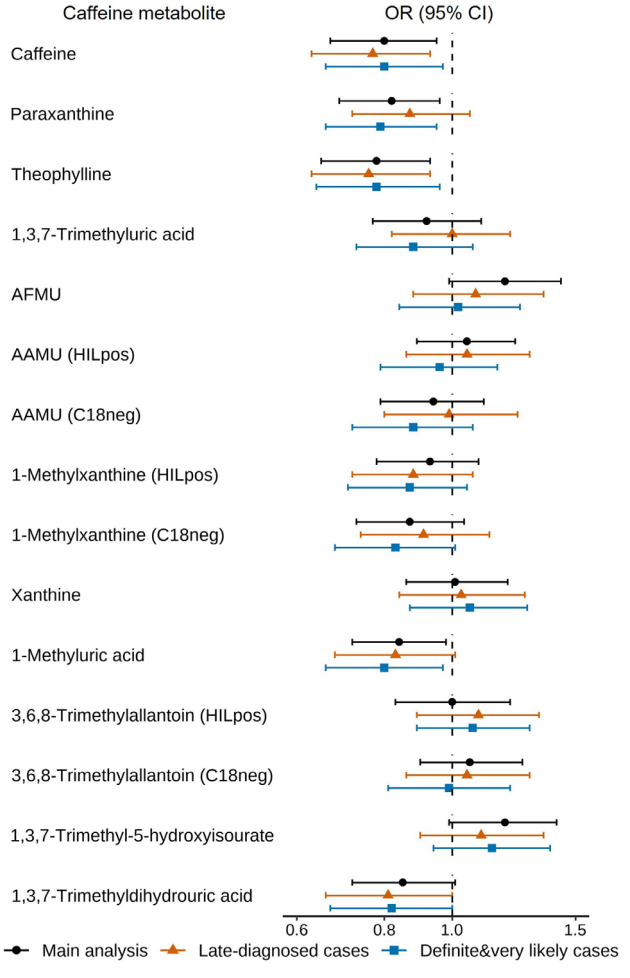


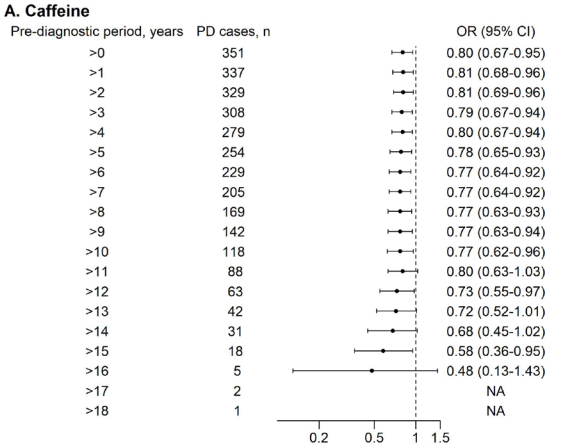
Figure S6. Sensitivity analyses of associations between caffeine metabolites and Parkinson's disease

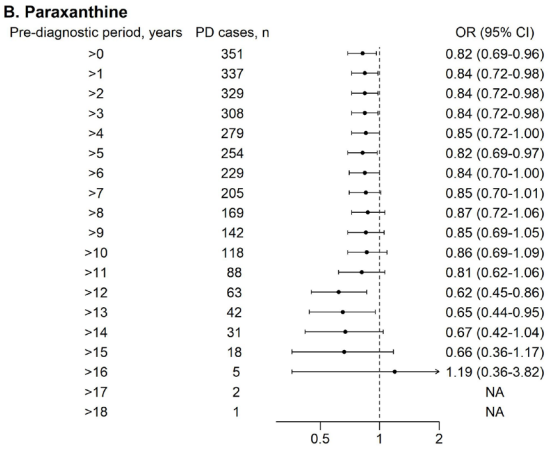
OR, odds ratio; CI, confidence interval; AFMU, 5-Acetylamino-6-formylamino-3-methyluracil; AAMU, 5-acetylamino-6-amino-3-methyluracil.

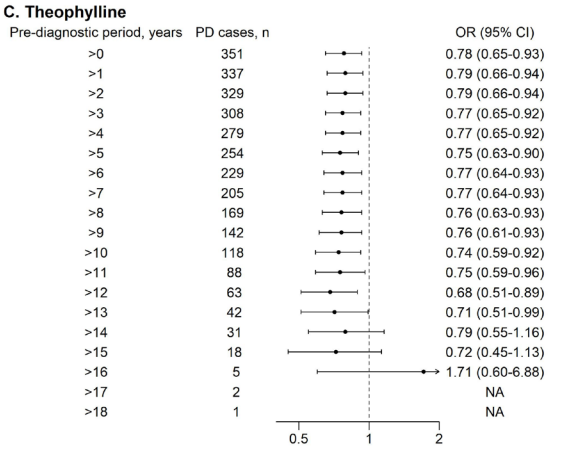
ORs and CIs (per SD increase of log2 ion intensity) for main analyses were calculated by conditional logistic regression for the matched case-control sets, adjusted for smoking status. ORs and CIs for sensitivity analyses were calculated by unconditional logistic regression adjusted for age at recruitment, sex, study center and smoking status.

Figure S7. Associations of selected caffeine metabolites and Parkinson's disease in different pre-diagnostic periods

For PD cases, periods between recruitment and diagnosis (pre-diagnostic periods) ranged from 0.2 to 18.3 years. ORs and CIs (per SD increase of log2 ion intensity) were calculated by unconditional logistic regression adjusted for age at recruitment, sex, study center and smoking status.

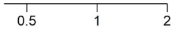






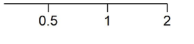
D. 1-Methyluric acid

Pre-diagnostic period, years	PD cases, n	OR (95% CI)
>0	351	0.84 (0.72-0.98)
>1	337	0.84 (0.72-0.98)
>2	329	0.84 (0.71-0.98)
>3	308	0.84 (0.72-0.99)
>4	279	0.85 (0.72-1.01)
>5	254	0.85 (0.72-1.01)
>6	229	0.86 (0.73-1.02)
>7	205	0.86 (0.72-1.03)
>8	169	0.83 (0.68-1.01)
>9	142	0.82 (0.66-1.00)
>10	118	0.85 (0.68-1.07)
>11	88	0.83 (0.64-1.07)
>12	63	0.97 (0.72-1.31)
>13	42	0.95 (0.67-1.36)
>14	31	1.06 (0.70-1.60)
>15	18	0.93 (0.56-1.55)
>16	5	1.40 (0.49-4.38)
>17	2	NA
>18	1	NA



E. 1,3,7-Trimethyldihydrouric acid

Pre-diagnostic period, years	PD cases, n	OR (95% CI)
>0	351	0.85 (0.72-1.01)
>1	337	0.86 (0.73-1.02)
>2	329	0.87 (0.73-1.03)
>3	308	0.85 (0.71-1.01)
>4	279	0.84 (0.71-1.00)
>5	254	0.81 (0.68-0.97)
>6	229	0.80 (0.66-0.96)
>7	205	0.81 (0.67-0.98)
>8	169	0.81 (0.66-1.00)
>9	142	0.82 (0.66-1.02)
>10	118	0.82 (0.65-1.04)
>11	88	0.90 (0.69-1.17)
>12	63	0.88 (0.65-1.21)
>13	42	0.87 (0.61-1.25)
>14	31	0.92 (0.61-1.40)
>15	18	0.87 (0.53-1.43)
>16	5	0.88 (0.36-2.32)
>17	2	NA
>18	1	NA



F. AFMU

Pre-diagnostic period, years	PD cases, n	OR (95% CI)
>0	351	1.19 (0.99-1.43)
>1	337	1.13 (0.96-1.34)
>2	329	1.15 (0.97-1.36)
>3	308	1.13 (0.95-1.34)
>4	279	1.15 (0.97-1.38)
>5	254	1.18 (0.98-1.42)
>6	229	1.14 (0.95-1.39)
>7	205	1.11 (0.91-1.35)
>8	169	1.08 (0.88-1.35)
>9	142	1.10 (0.88-1.38)
>10	118	1.04 (0.82-1.34)
>11	88	1.08 (0.82-1.44)
>12	63	1.25 (0.90-1.77)
>13	42	1.06 (0.73-1.58)
>14	31	1.07 (0.70-1.70)
>15	18	0.95 (0.59-1.63)
>16	5	1.05 (0.49-3.50)
>17	2	NA
>18	1	NA

G. 1,3,7-Trimethyl-5-hydroxyisourate

Pre-diagnostic period, years	PD cases, n	OR (95% CI)
>0	351	1.19 (0.99-1.41)
>1	337	1.18 (1.00-1.40)
>2	329	1.17 (0.99-1.38)
>3	308	1.17 (0.99-1.38)
>4	279	1.18 (1.00-1.41)
>5	254	1.18 (0.99-1.42)
>6	229	1.15 (0.96-1.38)
>7	205	1.10 (0.91-1.33)
>8	169	1.10 (0.90-1.35)
>9	142	1.00 (0.80-1.24)
>10	118	0.99 (0.78-1.27)
>11	88	0.96 (0.72-1.28)
>12	63	0.91 (0.66-1.27)
>13	42	0.81 (0.54-1.22)
>14	31	0.86 (0.53-1.39)
>15	18	0.87 (0.48-1.57)
>16	5	0.75 (0.15-2.99)
>17	2	NA
>18	1	NA

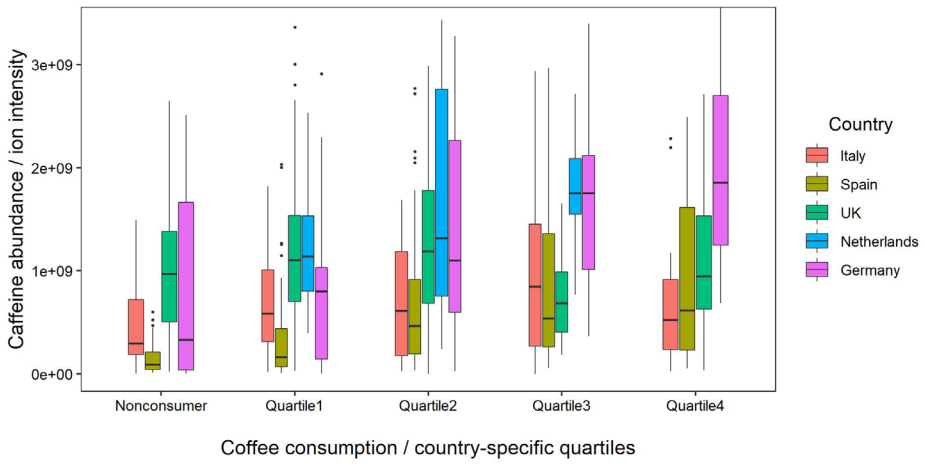


Figure S8. Caffeine levels in coffee intake quartiles for each country

Chapter 6

Lipopolysaccharide-binding Protein and Future Parkinson's Disease Risk: A European Prospective Cohort

Yujia Zhao, Douglas I Walker, Christina M Lill, Bastiaan R Bloem, Sirwan KL Darweesh, Brismar Pinto-Pacheco, Brooklyn McNeil, Gary W Miller, Alicia K Heath, Myrthe Frissen, Dafina Petrova, Maria-Jose Sánchez, María-Dolores Chirlaque, Marcela Guevara, Maurizio Zibetti, Salvatore Panico, Lefkos Middleton, Verena Katzke, Rudolf Kaaks, Elio Riboli, Giovanna Masala, Sabina Sieri, Raul Zamora-Ros, Pilar Amiano, Mazda Jenab, Susan Peters, Roel Vermeulen

Corresponding authors: Roel Vermeulen, Susan Peters

Published in *Journal of Neuroinflammation*, July 2023

https://pubmed.ncbi.nlm.nih.gov/37480114/

Abstract

Introduction: Lipopolysaccharide (LPS) is the outer membrane component of gram-negative bacteria. LPS-binding protein (LBP) is an acute-phase reactant that mediates immune responses triggered by LPS and has been used as a blood marker for LPS. LBP has recently been indicated to be associated with Parkinson's disease (PD) in small-scale retrospective case-control studies. We aimed to investigate the association between LBP blood levels with PD risk in a nested case-control study within a large European prospective cohort.

Methods: A total of 352 incident PD cases (55% males) were identified and one control per case was selected, matched by age at recruitment, sex and study center. LBP levels in plasma collected at recruitment, which was on average 7.8 years before diagnosis of the cases, were analyzed by enzyme linked immunosorbent assay. Odds ratios (ORs) were estimated for one unit increase of the natural log of LBP levels and PD incidence by conditional logistic regression.

Results: Plasma LBP levels were higher in prospective PD cases compared to controls (median (interquartile range) 26.9 (18.1-41.0) vs. 24.7 (16.6-38.4) μg/ml). The OR for PD incidence per one unit increase of log LBP was elevated (1.46, 95%CI 0.98-2.19). This association was more pronounced among women (OR 2.68, 95%CI 1.40-5.13) and overweight/obese subjects (OR 1.54, 95%CI 1.09-2.18).

Conclusion: The findings suggest that higher plasma LBP levels may be associated with an increased risk of PD and may thus pinpoint to a potential role of endotoxemia in the pathogenesis of PD, particularly in women and overweight/obese individuals.

Keywords: Parkinson's disease, lipopolysaccharide-binding protein, pre-diagnostic, systemic inflammation, endotoxemia

Introduction

Parkinson's disease (PD) is the second most common neurodegenerative disease affecting more than 1% of the population aged 60 years and older (1). The pathogenesis is complex and multifaceted, with increasing evidence suggesting that neuroinflammation likely plays a fundamental role (2). Mounting evidence suggests a crosstalk between brain inflammation and peripheral inflammation (2). Lipopolysaccharide (LPS, also known as endotoxin), the component of the outer membrane of gram-negative bacteria, is a potent activator of innate immune responses. LPS normally presents at much higher concentrations in the human gut as compared to blood (3). It has been speculated that already during the early stages of PD, LPS increasingly enters the blood due to a disrupted intestinal barrier, possibly induced by altered gut microbiota (4). The resultant systemic inflammation, induced by elevated circulating LPS, might in turn exacerbate ongoing neurodegeneration in PD by reinforcing microglial activation in the brain (5). Thus, endotoxemia has been hypothesized as a potential pathological mechanism of PD (6).

Although accumulating animal studies indicate the involvement of the systemic innate immune response in PD (5), the associated components and underlying mechanism in humans remain unclear. The LPS-binding protein (LBP) is a secretory acute-phase protein synthesized mainly in the liver. LBP has the dual role of promoting innate immune responses to LPS (7), as well as enhancing the neutralization and clearance of LPS by high density lipoprotein (8). Given technical limitations of measuring LPS in biofluids with sufficient accuracy (9), elevated LBP concentrations in serum or plasma have been suggested as a useful marker indicating endotoxemia and systemic inflammation in chronic diseases (10-12). Elevated LBP blood concentration was also recently proposed as a biomarker for intestinal permeability (13).

Several studies have reported lower LBP blood levels in prevalent PD cases as compared to healthy controls (14-18) (Table S1). However, these studies were mostly based on small groups of PD patients from hospitals. Moreover, LBP levels were assessed after the diagnosis, which might be subject to reverse causation. To the best of our knowledge, no prospective study has yet been conducted on the possible relation between LBP and the risk of PD. Therefore, we aimed to investigate the association between plasma LBP levels and the future PD onset measuring LBP levels in pre-diagnostic plasma samples in a large prospective European cohort.

Methods

Study design

The European Prospective Investigation into Cancer and Nutrition (EPIC) is a large prospective cohort study that was initiated in 1992 and recruited more than half a million people in ten European countries. At the time of enrollment, information on diet and lifestyle was collected through validated questionnaires, anthropometric measurements were performed, and blood samples were collected (19). The EPIC study was approved by the ethical committee of the International Agency for Research on Cancer (IARC) and by the ethical review boards of each study center. All participants signed a written informed consent.

The EPIC4PD study was conducted within EPIC and aimed to prospectively assess the role of risk factors in PD. This sub-study was based on a source population of 220,494 subjects from Sweden, the United Kingdom (UK), the Netherlands, Germany, Spain, Italy and Greece (20). Potential PD cases were identified through medical record linkage and further validated by experts in movement disorders through clinical records, according to the diagnostic criteria of the UK Brain Bank. A total of 881 PD cases were ascertained. Cases who received a diagnosis after the date of recruitment were defined as incident cases (n=734) (20).

Reliability of diagnoses was determined by the quality of clinical data (rated as 'poor', 'good' or 'excellent'), as well as the confidence degree of the expert neurologist's final judgement (rated as 'low', 'medium' or 'high') (20). Diagnoses were defined as 'definite' only when the confidence degree of the neurologist was high and the data quality was excellent; 'very likely' when the confidence degree was high, but data quality was either good or poor; 'probable' when the confidence degree was medium and data quality was either excellent or good; and diagnoses were defined as 'possible' in all remaining cases.

Here, we conducted a nested case-control study within the EPIC4PD cohort. A total of 352 incident PD cases for whom a plasma sample was available were included for current analyses. No subjects from Sweden and Greece were included as no blood samples were available in the central EPIC biobank. Among the included cases, 45 and 144 were categorized as 'definite' and 'very likely' cases, respectively. One control per case was selected by incidence density sampling matched by age at recruitment, sex and study center.

Lipopolysaccharide-binding protein measurement

A blood sample was obtained from each participant at recruitment and stored at a central biobank in liquid nitrogen (at -196 °C). LBP in plasma was measured by enzyme

linked immunosorbent assay (ELISA) (HK315-02, Hycult Biotech) in duplicates. The arithmetic mean of the duplicate measurements for each participant was calculated and used in the statistical analyses. Relative standard deviation (RSD) was calculated to check the variability of the two measurements. The ELISA assays were run on eleven plates, with samples from 64 subjects in each. All pairs of PD case and matched control were measured on the same plates to avoid batch effects in the case-control comparison.

Statistical analysis

Conditional logistic regression for the matched case-control sets was applied to investigate the association between plasma LBP levels and PD. LBP levels were naturally log-transformed to reduce influence of extreme values. Smoking is a well recognized inverse risk factor for PD (21), and increased body mass index (BMI) is indicated associated with higher PD incidence (22). Further considering being overweight and a smoker are related to higher LBP levels (23), smoking status and BMI were deemed as potential confounders and adjusted for in our models. Subjects with missing information on smoking status (n=10 for cases, n=13 for controls) were coded as 'unknown'.

To assess possible effect modification by sex, we performed stratified analyses on men and women. We also stratified analyses by smoking status at recruitment (current, former, never smokers) and BMI (BMI<25 kg/m^2, BMI≥25 kg/m^2), for which mixed effects logistic regression was used, adjusted for the matching variables (age at recruitment, sex, study center) and BMI/smoking status as fixed effects. ELISA plate was treated as random effect since batch effects might exist given different RSDs in analytical plates (Figure S1).

To investigate LBP levels in different stages prior to diagnosis, we ran mixed linear regression analyses on the pre-diagnostic periods (defined as time between recruitment and PD diagnosis) and natural log LBP levels among PD cases.

Furthermore, we performed sensitivity analyses: i) excluding PD cases diagnosed within 8 years (median) after recruitment to rule out possible reverse causality; ii) including only definite and very likely PD diagnosis; and iii) excluding case-control pairs in the two plates of ELISA where the RSDs for duplicate LBP measurements were relatively high (Figure S1).

Results

A total of 352 PD cases (55% males) and 352 matched controls were included in this study (Table 1). The median age at recruitment was 60 years, and the median period

between recruitment and PD diagnosis was 7.8 years. Smoking status and BMI at recruitment were not significantly different between PD cases and controls.

Plasma LBP levels were right-skewed distributed (Figure S2). LBP levels were slightly higher in PD cases than in controls (median (interquartile range), 26.9 (18.1-41.0) vs. 24.7 (16.6-38.4) µg/ml). LBP concentrations in overweight/obese subjects (BMI≥25 kg/m^2) were significantly higher than in normal-weight subjects (BMI<25 kg/m^2) (median (interquartile range), 27.2 (18.1-42.7) vs. 23.3 (16.5-37.1) µg/ml; *p*-value 0.004)

A positive association was observed between LBP levels at baseline and future PD diagnosis (Table 2). The adjusted odds ratio (OR) was 1.46 (95% confidence interval (CI) 0.98-2.19) per one natural log unit increase in LBP levels. Stratified by sex, elevated LBP levels were associated with PD incidence among women (OR 2.68, 95%CI 1.40-5.13), while for men the OR was 0.93 (0.55-1.59) (*p*-value for interaction 0.016). The association among never smokers was slightly stronger (OR 1.30, 95%CI 0.87-1.95) than in current and former smokers but there was no evidence for effect modification (*p*-value for interaction >0.05). In contrast, BMI significantly modified the association between LBP and PD (*p*-value for interaction 0.018). LBP was positively associated with incident PD in overweight/obese subjects (OR 1.54, 95%CI 1.09-2.18), but not in normal-weight counterparts (OR 0.78, 95%CI 0.48-1.28). Sensitivity analyses did not materially change the observed associations (Table 2).

There was no obvious trend for plasma LBP levels in different pre-diagnostic periods among PD cases (Figure 1). The coefficient estimate between logarithmic LBP and year to diagnosis from the linear regression was -0.004 (95%CI -0.015-0.007).

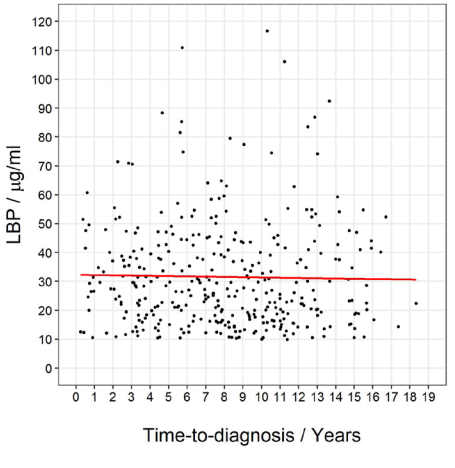


Figure 1. Scatter plot of pre-diagnostic plasma LBP levels in PD cases only (n=352) in relation to years to diagnosis.

The line represents the linear regression fitted on log-transformed LBP levels (β -0.004).

Table 1. Characteristics of study participants

Characteristic	PD cases n=352	Controls n=352
Age at recruitment, years[a]	60.8 (54.8-65.7)	60.4 (55.0-65.2)
Age at PD diagnosis, years	68.7 (62.8-74.0)	--
Years between recruitment and PD diagnosis	7.8 (4.6-11.0)	--
Sex, n (%)[a]		
Male	195 (55%)	195 (55%)
Female	157 (45%)	157 (45%)
Country, n (%)[a]		
Italy	54 (15%)	54 (15%)
Spain	97 (28%)	97 (28%)
UK	142 (40%)	142 (40%)
Netherlands	13 (4%)	13 (4%)
Germany	46 (13%)	46 (13%)
Smoking status at recruitment, n (%)		
Never smokers	183 (52%)	174 (49%)
Former smokers	116 (33%)	110 (31%)
Current smokers	43 (12%)	55 (16%)
Unknown	10 (3%)	13 (4%)
BMI categories, n (%)		
BMI<25 kg/m^2	119 (34%)	137 (39%)
BMI≥25	233 (66%)	215 (61%)

[a] Matching variables

Continuous data were expressed as median (interquartile range)

Categorical data were expressed as number (percentage)

BMI, body mass index, calculated as the weight in kilograms divided by the square of the height in meters.

Table 2. Association between plasma LBP levels and risk of PD

	PD cases plasma LBP (µg/ml)[a]/n	Controls plasma LBP (µg/ml)[a]/n	Odds ratio (95% CI)[b]	*p*-value
Main analysis[c]	26.9 (18.1-41.0) / 352	24.7 (16.6-38.4) / 352	1.46 (0.98-2.19)	0.061
Subgroup analysis by sex[c]				
Men	24.6 (18.1-38.9) / 195	24.2 (16.9-39.3) / 195	0.93 (0.55-1.59)	0.797
Women	29.6 (18.2-44.0) / 157	25.4 (16.1-37.9) / 157	2.68 (1.40-5.13)	0.003
Subgroup analysis by smoking status[d]				
Current smokers	29.7 (19.1-42.6) / 43	23.0 (16.3-43.5) / 55	1.12 (0.49-2.51)	0.793
Former smokers	24.2 (17.2-38.1) / 116	25.4 (16.6-37.7) / 110	1.09 (0.65-1.84)	0.736
Never smokers	27.8 (18.2-40.9) / 183	24.4 (17.1-37.0) / 174	1.30 (0.87-1.95)	0.198
Subgroup analysis by BMI[e]				
BMI<25 kg/m^2	22.7 (15.8-33.3) / 119	24.5 (16.7-38.0) / 137	0.78 (0.48-1.28)	0.331
BMI≥25 kg/m^2	29.6 (19.4-43.9) / 233	24.8 (16.6-38.4) / 215	1.54 (1.09-2.18)	0.016
Exclude cases diagnosed within eight years[c]	23.7 (16.3-40.1) / 169	23.9 (16.1-37.3) / 169	1.45 (0.79-2.65)	0.234
Include definite and very likely cases only[c]	24.8 (17.1-40.8) / 189	24.7 (16.6-39.6) / 189	1.19 (0.69-2.07)	0.532
Exclude ELISA plates with high variability between duplicates[c]	26.4 (18.2-39.7) / 288	24.2 (16.6-36.8) / 288	1.55 (0.99-2.41)	0.053

[a] Median (interquartile range)/the number of subjects

[b] Odds ratio for one unit increase of naturally logarithmic LBP

[c] Conditional logistic regression for the matched case-control sets, adjusted for BMI and smoking status

[d] Mixed effects logistic regression, adjusted for age at recruitment, sex, study center and BMI, ELISA plate as random effect

[e] Mixed effects logistic regression, adjusted for age at recruitment, sex, study center and smoking status, ELISA plate as random effect

Discussion

This is the first study to investigate the association between pre-diagnostic LBP levels and PD risk, utilizing a large European-wide prospective cohort. LBP levels at recruitment were higher in those who subsequently developed PD, which was most pronounced among women and overweight/obese subjects. No differences in LBP levels were observed in different pre-diagnostic periods.

LBP is an acute-phase protein synthesized mainly by hepatocytes. It binds and transfers LPS in blood to the cellular receptor complex consisting of CD14, MD2 and Toll-like receptor 4, resulting in the activation of immune cells and production of inflammatory factors (7). The levels of LBP rapidly increase when LPS enters the circulation, even at a subclinical level (24). LBP is more stable than LPS in blood (half-life, 12-24 hours (25) vs. 2-4 minutes (26)) and can be easily measured with immunoassay, making it a good indicator of exposure to LPS.

Elevated LBP levels in blood have been shown to be associated with a higher risk of diverse chronic diseases, including cardiovascular disease (10), allergy (27), arthritis (12), Crohn's disease (11), and metabolic syndrome (28). Similar to our findings (median level 26.9 μg/ml), LBP concentrations in patients from these studies were around 20 μg/ml, which are much lower than the levels in acute-phase response (>100 μg/ml) (29). These low but increased LBP levels possibly reflect low-grade endotoxemia and chronic systemic inflammation. Chronic systemic inflammation, which, unlike acute inflammation, fails to resolve and leads to a persistent and chronic state (7), has been well acknowledged to be involved in PD pathogenesis (5). Our findings, in particular those among women and overweight/obese individuals, provide further support for the role of low-grade chronic inflammation in the development of PD.

Excess LPS in the circulation has been hypothesized to result from leaky gut in PD due to intestinal barrier dysfunction. A few preliminary studies observed reduced tight junction protein expression in colonic samples of PD patients (18, 30, 31), and some studies reported increased intestinal permeability in patients (14, 18, 32, 33). Changes of gut microbiota composition may cause alterations in the gut barrier function and permeability. Recent studies have reported depletion of bacteria in the *Lachnospiraceae* family and the *Faecalibacterium* genus (34), which produce butyrate that is a fundamental energy source for intestinal epithelial cells and plays a role in the maintenance of colonic homeostasis (35, 36). Bacterial products such as LPS in gut lumen gain access to lamina propria and the bloodstream by gut hyperpermeability. In our study, higher LBP levels in PD cases, which are suggested as an indicator for

increased intestinal permeability (13), further supports the involvement of LPS from the intestine in PD pathogenesis.

Aside from the inflammation pathway induced by LPS in the bloodstream, a putative mechanism by which compromised intestinal barrier may influence the brain in PD is the vagal pathway. It has been speculated that alpha-synuclein misfolding and aggregation initially start at the intersection of the gut lumen and the enteric nervous system, which are probably triggered by microbes and their products (37). Alpha-synuclein may then be propagated to neurons in the central nervous system through the vagus nerve via retrograde transport, causing abnormal alpha-synuclein deposits in the brain, which is a hallmark of PD pathology (37).

In our study, circulating LBP levels were higher in overweight/obese than in normal-weight individuals, which is in line with previous findings from general populations in China (28, 38) and Spain (23). Effect modification of BMI of the association between LBP and PD risk was noted, and the association was strongest among overweight/obese individuals. This phenomenon is reasonable due to the low-grade chronic inflammation in the development of obesity (39), and gut microbiota is considered as one of the factors in the process. Many experimental studies demonstrated altered microbiota composition, enhanced intestinal permeability, low-grade endotoxemia in animal models of obesity (40-42). However, the exact biological mechanisms of obesity in the crosslink between gut and PD need further exploration.

Several epidemiological studies have reported an association between LBP levels in serum or plasma and the presence of PD (Table S1) (14-18). However, contrary to our results, they all found lower LBP levels in clinically manifest PD compared to controls. Therefore, the difference in LBP is possibly driven by the consequences of the disease (e.g., chronic constipation that is typical of PD, changes in diet, medications) and does not necessarily contribute to the occurrence of the disease (the average disease duration in these studies ranged from 2 to 9.5 years (14-18)). One study indicated that LBP could be transported from the systemic circulation to the intestinal epithelial basolateral and finally into the gut lumen during endotoxemia (43). This transintestinal efflux possibly explains lower LBP levels in prevalent PD cases, especially when the disease progresses, and the intestinal barrier function deteriorates further. The mechanisms underlying discordant blood LBP levels before and after PD onset remain to be elucidated.

A limitation of our study is that we did not simultaneously analyze proinflammatory cytokines, systemic endotoxin, and intestinal barrier function, thus we could not verify the mechanistic associations between intestinal microbiota composition, increased intestinal permeability, and LPS invasion. Second, our study was performed in patients

whose samples were taken during the prodromal phase of PD, on average eight years prior to the diagnosis. Constipation occurs as early as 20 years before the onset of motor symptoms (44), and we have no information about the presence of bowel dysfunction in our participants at the time of sample collection. Therefore, we cannot fully exclude that the higher LBP levels in our study were secondary to PD-related constipation. However the fact that we see no association between time-to-diagnosis and LBP levels while symptoms worsen closer to diagnosis might speak against this. Furthermore, the observed sex-related discrepancy of LBP levels and PD risk, as the association was significant in women only, suggests the potential presence of distinct sex-dependent pathological mechanisms, which need further confirmation in an independent cohort and warrant further exploration.

Conclusion

Overall, our results showed that elevated LBP levels prior to diagnosis were associated with higher PD risk in women and overweight/obese individuals. This is the first study to evaluate pre-diagnostic blood LBP levels, shedding some light on LPS-mediated inflammation in the gut-brain axis hypothesis of PD. Future prospective studies are needed to elucidate the association between markers of endotoxemia and the risk of PD.

References

1. Ascherio A, Schwarzschild MA. The epidemiology of Parkinson's disease: risk factors and prevention. The Lancet Neurology. 2016;15(12):1257-72.

2. Tansey MG, Wallings RL, Houser MC, Herrick MK, Keating CE, Joers V. Inflammation and immune dysfunction in Parkinson disease. Nat Rev Immunol. 2022.

3. Sender R, Fuchs S, Milo R. Revised Estimates for the Number of Human and Bacteria Cells in the Body. PLoS biology. 2016;14(8):e1002533.

4. Ghosh SS, Wang J, Yannie PJ, Ghosh S. Intestinal Barrier Dysfunction, LPS Translocation, and Disease Development. J Endocr Soc. 2020;4(2):bvz039.

5. Ferrari CC, Tarelli R. Parkinson's disease and systemic inflammation. Parkinsons Dis. 2011;2011:436813.

6. Brown GC. The endotoxin hypothesis of neurodegeneration. Journal of neuroinflammation. 2019;16(1):180.

7. Morris MC, Gilliam EA, Li L. Innate immune programing by endotoxin and its pathological consequences. Front Immunol. 2014;5:680.

8. Lamping N, Dettmer R, Schroder NW, et al. LPS-binding protein protects mice from septic shock caused by LPS or gram-negative bacteria. J Clin Invest. 1998;101(10):2065-71.

9. Cohen J. The detection and interpretation of endotoxaemia. Intensive care medicine. 2000;26 Suppl 1:S51-6.

10. Asada M, Oishi E, Sakata S, et al. Serum Lipopolysaccharide-Binding Protein Levels and the Incidence of Cardiovascular Disease in a General Japanese Population: The Hisayama Study. Journal of the American Heart Association. 2019;8(21):e013628.

11. Lakatos PL, Kiss LS, Palatka K, et al. Serum lipopolysaccharide-binding protein and soluble CD14 are markers of disease activity in patients with Crohn's disease. Inflammatory bowel diseases. 2011;17(3):767-77.

12. Fotis L, Shaikh N, Baszis KW, et al. Serologic Evidence of Gut-driven Systemic Inflammation in Juvenile Idiopathic Arthritis. J Rheumatol. 2017;44(11):1624-31.

13. Seethaler B, Basrai M, Neyrinck AM, et al. Biomarkers for assessment of intestinal permeability in clinical practice. Am J Physiol Gastrointest Liver Physiol. 2021;321(1):G11-G7.

14. Forsyth CB, Shannon KM, Kordower JH, et al. Increased intestinal permeability correlates with sigmoid mucosa alpha-synuclein staining and endotoxin exposure markers in early Parkinson's disease. PloS one. 2011;6(12):e28032.

15. Hasegawa S, Goto S, Tsuji H, et al. Intestinal Dysbiosis and Lowered Serum Lipopolysaccharide-Binding Protein in Parkinson's Disease. PloS one. 2015;10(11):e0142164.

16. Pal GD, Shaikh M, Forsyth CB, Ouyang B, Keshavarzian A, Shannon KM. Abnormal lipopolysaccharide binding protein as marker of gastrointestinal inflammation in Parkinson disease. Frontiers in neuroscience. 2015;9:306.

17. Chen SJ, Chi YC, Ho CH, Yang WS, Lin CH. Plasma Lipopolysaccharide-Binding Protein Reflects Risk and Progression of Parkinson's Disease. Journal of Parkinson's disease. 2021;11(3):1129-39.

18. Perez-Pardo P, Dodiya HB, Engen PA, et al. Role of TLR4 in the gut-brain axis in Parkinson's disease: a translational study from men to mice. Gut. 2019;68(5):829-43.

19. Riboli E, Hunt KJ, Slimani N, et al. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. Public health nutrition. 2002;5(6b):1113-24.

20. Gallo V, Brayne C, Forsgren L, et al. Parkinson's Disease Case Ascertainment in the EPIC Cohort: The NeuroEPIC4PD Study. Neuro-degenerative diseases. 2015;15(6):331-8.

21. Gallo V, Vineis P, Cancellieri M, et al. Exploring causality of the association between smoking and Parkinson's disease. Int J Epidemiol. 2019;48(3):912-25.

22. Chen J, Guan Z, Wang L, Song G, Ma B, Wang Y. Meta-analysis: overweight, obesity, and Parkinson's disease. International journal of endocrinology. 2014;2014:203930.

23. Gonzalez-Quintela A, Alonso M, Campos J, Vizcaino L, Loidi L, Gude F. Determinants of serum concentrations of lipopolysaccharide-binding protein (LBP) in the adult population: the role of obesity. PloS one. 2013;8(1):e54600.

24. Hudgins LC, Parker TS, Levine DM, et al. A single intravenous dose of endotoxin rapidly alters serum lipoproteins and lipid transfer proteins in normal volunteers. Journal of lipid research. 2003;44(8):1489-98.

25. Schumann RR. Old and new findings on lipopolysaccharide-binding protein: a soluble pattern-recognition molecule. Biochemical Society transactions. 2011;39(4):989-93.

26. Yao Z, Mates JM, Cheplowitz AM, et al. Blood-Borne Lipopolysaccharide Is Rapidly Eliminated by Liver Sinusoidal Endothelial Cells via High-Density Lipoprotein. J Immunol. 2016;197(6):2390-9.

27. Ha EK, Kim JH, Yon DK, et al. Association of serum lipopolysaccharide-binding protein level with sensitization to food allergens in children. Scientific reports. 2021;11(1):2143.

28. Liu X, Lu L, Yao P, et al. Lipopolysaccharide binding protein, obesity status and incidence of metabolic syndrome: a prospective study among middle-aged and older Chinese. Diabetologia. 2014;57(9):1834-41.

29. Opal SM, Palardy JE, Marra MN, Fisher CJ, Jr., McKelligon BM, Scott RW. Relative concentrations of endotoxin-binding proteins in body fluids during infection. Lancet (London, England). 1994;344(8920):429-31.

30. Clairembault T, Leclair-Visonneau L, Coron E, et al. Structural alterations of the intestinal epithelial barrier in Parkinson's disease. Acta neuropathologica communications. 2015;3:12.

31. Kuan WL, Bennett N, He X, et al. α-Synuclein pre-formed fibrils impair tight junction protein expression without affecting cerebral endothelial cell function. Experimental neurology. 2016;285(Pt A):72-81.

32. Schwiertz A, Spiegel J, Dillmann U, et al. Fecal markers of intestinal inflammation and intestinal permeability are elevated in Parkinson's disease. Parkinsonism & related disorders. 2018;50:104-7.

33. Dumitrescu L, Marta D, Dănău A, et al. Serum and Fecal Markers of Intestinal Inflammation and Intestinal Barrier Permeability Are Elevated in Parkinson's Disease. Frontiers in neuroscience. 2021;15:689723.

34. Romano S, Savva GM, Bedarf JR, Charles IG, Hildebrand F, Narbad A. Meta-analysis of the Parkinson's disease gut microbiome suggests alterations linked to intestinal inflammation. NPJ Parkinsons Dis. 2021;7(1):27.

35. Hamer HM, Jonkers D, Venema K, Vanhoutvin S, Troost FJ, Brummer RJ. Review article: the role of butyrate on colonic function. Alimentary pharmacology & therapeutics. 2008;27(2):104-19.

36. Ferreira-Halder CV, Faria AVS, Andrade SS. Action and function of Faecalibacterium prausnitzii in health and disease. Best practice & research Clinical gastroenterology. 2017;31(6):643-8.

37. Breen DP, Halliday GM, Lang AE. Gut-brain axis and the spread of alpha-synuclein pathology: Vagal highway or dead end? Mov Disord. 2019;34(3):307-16.

38. Sun L, Yu Z, Ye X, et al. A marker of endotoxemia is associated with obesity and related metabolic disorders in apparently healthy Chinese. Diabetes Care. 2010;33(9):1925-32.

39. Artemniak-Wojtowicz D, Kucharska AM, Pyrżak B. Obesity and chronic inflammation crosslinking. Cent Eur J Immunol. 2020;45(4):461-8.

40. Frazier TH, DiBaise JK, McClain CJ. Gut microbiota, intestinal permeability, obesity-induced inflammation, and liver injury. JPEN J Parenter Enteral Nutr. 2011;35(5 Suppl):14S-20S.

41. Aoun A, Darwish F, Hamod N. The Influence of the Gut Microbiome on Obesity in Adults and the Role of Probiotics, Prebiotics, and Synbiotics for Weight Loss. Prev Nutr Food Sci. 2020;25(2):113-23.

42. Brun P, Castagliuolo I, Di Leo V, et al. Increased intestinal permeability in obese mice: new evidence in the pathogenesis of nonalcoholic steatohepatitis. Am J Physiol Gastrointest Liver Physiol. 2007;292(2):G518-25.

43. Vreugdenhil AC, Snoek AM, Greve JW, Buurman WA. Lipopolysaccharide-binding protein is vectorially secreted and transported by cultured intestinal epithelial cells and is present in the intestinal mucus of mice. J Immunol. 2000;165(8):4561-6.

44. Savica R, Carlin JM, Grossardt BR, et al. Medical records documentation of constipation preceding Parkinson disease: A case-control study. Neurology. 2009;73(21):1752-8.

Chapter 6 Supplemental materials

Table S1. LBP levels in Parkinson's disease cases and controls in previous studies

Study	Study design	No. of PD cases	No. of controls	LBP in PD[a]	LBP in controls[a]	Test and *p*-value PD cases vs. controls
Forsyth 2011 (1)	Case-control	9	10	22.9±5.5	84.3±31.4	Mann-Whitney U test *p*=0.016
Hasegawa 2015 (2)	Case-control	51	36	7.8±2.4	10.1±5.1	Student's t-test *p*<0.01
Pal 2015 (3)	Case-control	94	99	9.3±6.7	11.3±7.4	Analysis of covariance, adjusting age and gender *p*=0.03
Perez-Pardo 2019 (4)	Case-control	5	5	15.73±3.75	33.47±6.20	Student's t-test *p*=0.04
Chen 2021 (5)	Case-control	248	149	9.08±2.91	10.10±3.00	Mann-Whitney U test *p*=0.003

[a] Mean ± standard deviation, µg/ml

References:

1. Forsyth CB, Shannon KM, Kordower JH, et al. Increased intestinal permeability correlates with sigmoid mucosa alpha-synuclein staining and endotoxin exposure markers in early Parkinson's disease. PloS one. 2011;6(12):e28032.

2. Hasegawa S, Goto S, Tsuji H, et al. Intestinal Dysbiosis and Lowered Serum Lipopolysaccharide-Binding Protein in Parkinson's Disease. PloS one. 2015;10(11):e0142164.

3. Pal GD, Shaikh M, Forsyth CB, Ouyang B, Keshavarzian A, Shannon KM. Abnormal lipopolysaccharide binding protein as marker of gastrointestinal inflammation in Parkinson disease. Frontiers in neuroscience. 2015;9:306.

4. Perez-Pardo P, Dodiya HB, Engen PA, et al. Role of TLR4 in the gut-brain axis in Parkinson's disease: a translational study from men to mice. Gut. 2019;68(5):829-43.

5. Chen SJ, Chi YC, Ho CH, Yang WS, Lin CH. Plasma Lipopolysaccharide-Binding Protein Reflects Risk and Progression of Parkinson's Disease. Journal of Parkinson's disease. 2021;11(3):1129-39.

Figure S1. Boxplot of relative standard deviations of two LBP measurements from the same subjects, grouped on eleven ELISA plates.

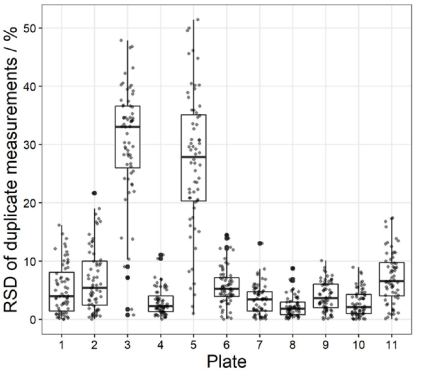
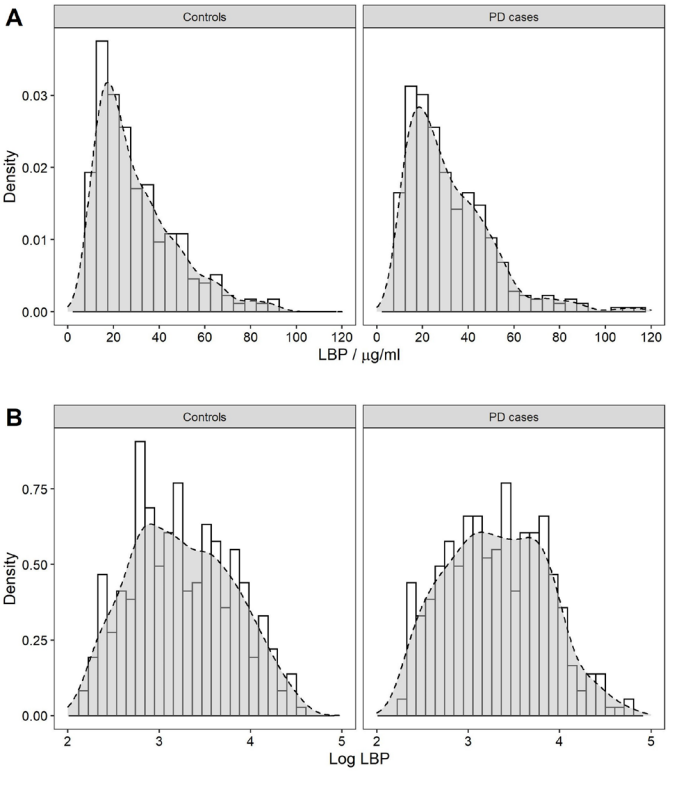


Figure S2. Histogram with density plot of LBP concentrations on original scale (A) or on natural log scale (B), faceting on controls and Parkinson's disease cases.



Chapter 7

Gut Microbial Metabolites and Future Risk of Parkinson's Disease: A Metabolome-wide Association Study

Yujia Zhao[#], Yunjia Lai[#], Sirwan KL Darweesh, Bastiaan R Bloem, Lars Forsgren, Johnni Hansen, Verena A Katzke, Giovanna Masala, Sabina Sieri, Carlotta Sacerdote, Salvatore Panico, Raul Zamora-Ros, Maria-Jose Sánchez, José María Huerta, Marcela Guevara, Ana Vinagre-Aragon, Paolo Vineis, Christina M. Lill, Gary Miller, Susan Peters, Roel Vermeulen

[#] Joint first authors

Corresponding author: Roel Vermeulen

Submitted

Abstract

Background: Gut microbiota is altered in individuals with Parkinson's disease (PD). Metabolites derived from microbiota activity have been utilized as functional indicators to investigate the gut-brain interaction in PD by exploratory case-control studies. However, these studies were typically small-scale and adopted post-diagnosis measurements, which are susceptible to reverse causality.

Objectives: To prospectively investigate the association between plasma microbial metabolites and PD risk within a metabolomics framework.

Methods: A nested case-control study was conducted within the EPIC prospective population cohort, comprising 351 cases with incident PD and 351 controls who were matched on age, sex and study center. Plasma samples were obtained at the cohort recruitment, on average 8 years (ranged 0.2-18 years) before PD diagnosis for the cases. The untargeted metabolome was profiled in plasma, with 167 microbial metabolites annotated. We performed a metabolome-wide association study (MWAS) to identify metabolites associated with future PD risk. Microbiota-relevant pathways related to PD risk were explored across different groups based on sex, smoking status, body mass index, and lipopolysaccharide-binding protein (LBP) levels as a biomarker of intestinal permeability.

Results: Our MWAS identified 13 microbial metabolites which were nominally associated with PD risk (p-value<0.05), including amino acids, bile acid, indoles, and hydroxy acid. However, none of these results remained significant after false discovery control (FDR=20%). Pathway analyses across the whole metabolome implicated three pathways in PD risk: (i) valine, leucine and isoleucine degradation, (ii) butanoate metabolism, and (iii) propanoate metabolism. In stratified analyses, more pronounced PD-associated microbial pathways were found in men, smokers, overweight/obese individuals, and those with higher LBP levels.

Conclusion: This study suggests that changes in microbial metabolites may be a pre-diagnostic feature of PD. We observed biologically plausible associations between microbial pathways and PD, and our results indicate that pathway enrichment may depend on individual characteristics.

Keywords: Parkinson's disease, gut-brain axis, microbial metabolites, pre-diagnostic biosamples, untargeted metabolomics

Introduction

The hypothesis that the gut plays an etiologic role in the development of Parkinson's disease (PD) has garnered interest, ever since the discovery of α-synuclein aggregation, a PD hallmark, originating in the gastrointestinal tract in many PD cases (1). Subsequent studies have shown alterations in the composition of microbiota in prevalent PD cases compared to healthy controls (2). The causes for these microbiota changes in prevalent cases and their functional consequences currently remain largely unclear. Importantly, as these measurements were performed in already diagnosed patients, it remains unclear whether these changes preceded the onset of PD or—alternatively—were a consequence of it, raising concerns about reversed causality.

To overcome that issue, there is a need for prospective investigations which examine the association between measures of the microbiota and the future risk of PD. Ideally, such studies should not only focus on microbiota composition, but also on microbiota functionalities, dysbiosis, and host-microbe interactions. To this end, metabolites originating from microbiota activity and released in the blood have emerged as promising biomarkers (3). Microbial metabolites can be broadly categorized into three types: i) metabolites directly produced by gut microbiota from diets, such as short-chain fatty acids and indole derivatives; ii) metabolites initially produced by the host and subsequently modified by gut microbiota, such as secondary bile acids; iii) metabolites synthesized de novo by microbiota, such as polysaccharide A (4). These metabolites play crucial roles in host-microbiota crosstalk by participating in a range of physiological and pathological functions. They contribute to the modulation of energy metabolism, facilitate nutrition absorption, regulate the composition and function of gut microbiota, and influence intestinal barrier integrity and permeability (5).

Previous studies have observed alterations in the levels of short-chain fatty acids and secondary bile acids in prevalent PD cases compared to controls (6, 7). Similar to studies on microbiota composition profiles, reverse causality cannot be ruled out since biosamples were collected post-diagnosis, while the disease process may have started up to 20 years earlier (8). Importantly, PD comorbidities such as constipation, as well as anti-PD medications like levodopa and catechol-O-methyltransferase inhibitors, could all influence gut microbiota and associated metabolites (7, 9).

In this study, we utilized a population-based prospective cohort study, EPIC4PD, where plasma samples were collected at recruitment, and participants were subsequently followed until PD onset (10). Being one of the largest biorepositories for PD, EPIC4PD offers significant opportunities for biomarker discovery in PD research. An earlier study within this cohort showed that plasma lipopolysaccharide-binding protein (LBP), a

biomarker of bacterial invasion and intestinal permeability, is associated with an elevated risk of PD, indicating potential involvement of gut microbiota in the disease development (11). Here, we aim to conduct a metabolome-wide association study (MWAS) within the EPIC4PD cohort, focusing on plasma microbial metabolites, to further illuminate the gut-brain interplay in PD.

Methods

Study design

This MWAS was performed as a nested case-control study within the EPIC4PD cohort, a multinational prospective cohort with up to two decades of follow-up. Cohort details and PD case ascertainment have been reported previously (10, 11). Out of the 734 incident PD cases confirmed in the EPIC4PD cohort, 351 cases were eligible for inclusion in the current study based on the availability of their pre-diagnostic plasma samples. One control per case was selected with incidence density sampling matching by age at recruitment, sex, and study center. Detailed information on diet and lifestyle factors was collected at enrollment (12).

Untargeted metabolomics

Microbial metabolites in plasma were characterized using untargeted metabolomics. Plasma samples were analyzed by liquid chromatography-high resolution mass spectrometry (LC-HRMS) in two complementary modes, namely hydrophilic interaction liquid chromatography-electrospray ionization [ESI(+)] and reverse-phase chromatography-ESI(-), referred to as 'HILIC-ESI$^+$' and 'C18-ESI$^-$' modes, respectively (13). Individual ion features were picked and aligned in all samples with unique retention time (RT), mass-to-charge (*m/z*), and ion abundance (peak area). In total, 9,435 features for HILIC-ESI$^+$ and 8,439 for C18-ESI$^-$ were aligned across samples and kept for analysis after quality control procedures and batch effect corrections. Details of analytical methods and data processing are provided in Text S1.

Before statistical analysis, ion features with >50% missing data were excluded. Missing values for the remaining features were considered below the limit of detection and imputed using quantile regression for left-censored missing data via the *imputeLCMD* R package (14). After imputation, 7,487 HILIC-ESI$^+$ and 6,968 C18-ESI$^-$ features were retained for subsequent analysis.

Annotation of microbial metabolites

Structural annotation of microbial metabolites within the metabolome was conducted based on two sources, including a large internal spectral library (established from >800 chemical standards) and a reference compound list of curated microbial metabolites from the Exposome-Explorer database that contains 457 unique structures (4). The

Exposome-Explorer database defines metabolites of microbial origin based on three types of evidence: a) *in vitro* experiments with fecal samples showing the metabolite production by gut bacteria; b) *in vivo* manipulation of gut microbiota with antibiotics, resulting in decreased metabolite concentrations; and c) comparison of metabolite concentrations in germ-free and conventional animals (4). In the Exposome-Explorer database, metabolites supported by both *in vitro* (a) and *in vivo* (b or c) evidence are considered largely or exclusively produced by human gut microbiota. Metabolites supported by only one type of evidence are probably to be influences or modified by microbiota. Details for compound annotation are provided in Text S1.

Of the whole metabolome, 167 features (78 HILIC-ESI$^+$ and 89 C18-ESI$^-$ features) were annotated as microbial metabolites, primarily falling into the categories of amino acids and peptides (n=42), fatty acyls (n=20), carbohydrates and conjugates (n=10), indoles and derivatives (n=10), phenylpropanoids and polyketides (n=10) (Table S1).

Statistical analysis

The relationship between microbial metabolism perturbation and PD was investigated by an MWAS, including the 167 microbial metabolites. Ion intensities of metabolites were log2-transformed and scaled by standard deviation (SD) for subsequent analyses. Conditional logistic regression was applied for the matched case-control sets. Odds ratios (ORs) and 95% confidence intervals (CIs) were estimated, with each model including a single metabolite as the independent variable. Smoking status and body mass index (BMI) were considered as covariates. Subjects with missing smoking status (n=23) were coded as 'unknown' and retained as a distinct group within the variable in the analyses. Multiple statistical tests in the MWAS raises concerns regarding an inflated rate of false positives if the significance cutoff remains at the conventionally stated *p*-value of 0.05 for a single test. Therefore, the Benjamini-Hochberg procedure, a commonly used multiple testing correction method, was applied and features were considered significant at a false discovery rate (FDR) threshold of 20%.

Stratified analyses were conducted by sex, smoking status (smokers or non-smokers at baseline), and BMI (normal weight, <25 kg/m^2; overweight/obese, ≥25 kg/m^2) to assess potential effect modification of the associations between metabolome and PD. Stratified analyses were also performed for plasma LBP concentrations, grouping subjects into those with lower (<median, 25.6 µg/ml) and higher LBP levels (≥median). We conducted the following, separate sensitivity analyses by repeating the main model after: i) limiting PD cases to those diagnosed more than eight years since recruitment to further rule out possible reverse causality (n=170); ii) limiting PD cases to those with higher validity of diagnosis (n=188) (labeled as 'definite' and 'very likely' in the EPIC4PD cohort (10)); or iii) additional adjustment for individual food category (fruit, vegetables,

meat, cereal, dairy, fish/seafood intake) for which consumption was reported at baseline to account for the impact of diet on gut microbiota.

We further examined the relationship between time-to-diagnosis (defined as years between recruitment and PD diagnosis) and the intensity of metabolites among PD cases, aiming to explore potential metabolite changes across different PD prodromal phases. Given the potential for a non-linear relationship, we utilized natural spline regression with 3 degrees of freedom.

To enable a pathway enrichment analysis, we performed an MWAS including the whole metabolome (7,487 HILIC-ESI$^+$ and 6,968 C18-ESI$^-$ features). The analytic framework and stratified analyses were same as those applied in the aforementioned MWAS focusing the microbial metabolites. Metabolites with a raw p-value below 0.05 were chosen for the pathway enrichment analysis using MetaboAnalystR 5.0 (15). The mass tolerance, analytical mode, and permutation number were set as ±5 ppm, mixed ('positive' for HILIC-ESI$^+$ and 'negative' for C18-ESI$^-$), and 1000, respectively. Enriched metabolic pathways were identified using a Gamma threshold (*p*-value derived from the permutation tests for the pathway) of 0.05 and required the presence of at least three metabolites associated with PD risk. This study presents enriched pathways related to gut microbiota. The selection criteria for determining pathways associated with the microbiota were knowledge-driven, based on extensive literature searches, particularly using the Exposome-Explorer database. For instance, the butanoate metabolism pathway is chosen due to documented evidence of butyric acid in both *in vitro* and *in vivo* studies within the database.

Results

Our study included participants with a median age of 60 years (ranged 38-76 years) at recruitment (Table S2). PD patients had a median period of 8 years (ranged 0.2-18 years) between recruitment and PD diagnosis. Smoking status and BMI at baseline were not significantly different between the case and control groups.

In the MWAS of annotated microbial metabolites, 13 metabolites showed associations with PD risk (raw *p*-value<0.05); however, none met the FDR 20% threshold (Table 1). Most of these metabolites exhibited higher abundance in pre-diagnostic plasma samples of PD cases compared to controls, including five amino acids and derivatives (arginylglutamate, creatinine, leucylproline, phenylalanine, delta-aminovaleric acid), xylose, indole-3-propionic acid, purine, gallic acid, L-2-hydroxyglutarate and 4-hydroxyphenylpyruvic acid. Conversely, dihydroresveratrol and taurocholic acid had lower plasma abundance among PD cases than their corresponding controls.

Short-chain fatty acids and bile acids are among the extensively studied microbial metabolites, and several metabolites in the two categories were identified in our study. Butyric acid, a common short-chain fatty acid, exhibited lower, albeit non-significant, abundance in PD cases compared to controls (OR 0.91, 95%CI 0.76-1.08). Primary bile acids, chenodeoxycholic and taurocholic acid, exhibited altered abundance within PD cases, although with opposing directions, with ORs of 1.13 (95%CI 0.97-1.31) and 0.81 (95% 0.67-0.96), respectively. However, the secondary bile acids, including lithocholic acid, glycodeoxycholic acid, and 3-oxodeoxycholic acid, did not show obvious changes among cases and controls (Table S1).

In the MWAS involving the whole metabolome, we identified 1000 metabolic features, comprising 547 HILIC-ESI$^+$ and 453 C18-ESI$^-$ features, exhibiting nominal associations with PD risk (raw *p*-value<0.05), but none met the FDR 20% threshold. Through pathway enrichment analysis, three pathways relating to microbiota activity showed enrichment, namely valine, leucine and isoleucine degradation, butanoate metabolism, and propanoate metabolism (Figure 1).

Stratified analyses showed limited overlap across subgroups of sex, smoking status, and BMI for PD-associated features (raw *p*-value<0.05) from the MWAS (Figure S1). Formal tests of interaction between sex, smoking status, BMI, and the 13 associated microbial metabolites (raw *p*-value<0.05) were non-significant for most features. Yet interestingly, microbiota-relevant pathways enriched for PD-associated features were heterogeneous across subgroups (Figure 1). In general, more PD-associated pathways were seen for individuals who were male, overweight/obese, and smoked. Only amino acid pathways were found associated with PD among individuals with lower LBP levels. In contrast, pathways of short-chain fatty acids (butanoate and propanoate) were further enriched among those with higher LBP levels.

In the sensitivity analyses, we limited PD cases to those with higher certainty levels of diagnosis, as well as those diagnosed more than eight years since recruitment. The effect estimates for metabolic features were consistent with those from the analyses including all cases (Figure S2). Additional adjustments for food intake yielded similar results compared to the main analyses (Figure S3).

Table 1. Microbial metabolites associated with Parkinson's disease risk (raw *p*-value<0.05)

Metabolite	Evidence on microbial origin[1]	Analytical method	*m/z*	Retention time (s)	OR (95% CI)[2]	*p*-value
Arginylglutamate	In vivo	C18-ESI$^-$	302.1553	32.2	1.22 (1.03-1.45)	0.020
Creatinine	In vivo	HILIC-ESI$^+$	114.0662	33.6	1.22 (1.01-1.48)	0.035
Leucylproline	In vivo	HILIC-ESI$^+$	229.1548	50.0	1.27 (1.03-1.55)	0.022
Phenylalanine	Both	C18-ESI$^-$	164.0716	34.3	1.19 (1.01-1.41)	0.037
delta-Aminovaleric acid	Both	C18-ESI$^-$	116.0718	30.4	1.18 (1.00-1.40)	0.049
Xylose	In vivo	C18-ESI$^-$	149.0456	37.3	1.25 (1.05-1.49)	0.011
Indolepropionic acid	Both	HILIC-ESI$^+$	190.0864	34.2	1.22 (1.05-1.43)	0.011
Dihydroresveratrol	In vitro	C18-ESI$^-$	229.0808	73.6	0.84 (0.70-1.00)	0.047
Purine	In vitro	C18-ESI$^-$	119.0351	34.7	1.24 (1.03-1.48)	0.020
Gallic acid	Both	C18-ESI$^-$	169.0134	38.8	1.22 (1.03-1.46)	0.024
Taurocholic acid	In vivo	HILIC-ESI$^+$	516.3055	35.5	0.81 (0.67-0.96)	0.016
L-2-hydroxyglutarate	In vitro	C18-ESI$^-$	147.0299	29.2	1.40 (1.05-1.85)	0.020
4-Hydroxyphenylpyruvic acid	In vivo	HILIC-ESI$^+$	181.0530	41.0	1.30 (1.11-1.52)	0.001

m/z, mass-to-charge ratio; OR, odds ratio; CI, confidence interval.

[1] 'Both' indicated that the specific metabolite simultaneously met in vitro (production by gut bacteria) and in vivo (reduction of concentrations by antibiotics or in germ-free animals) evidence supporting a microbial origin. 'In vivo/In vitro' represented that evidence was limited to either in vivo or in vitro studies.

[2] Conditional logistic regression for the matched case-control sets, adjusted by smoking status and BMI. OR was calculated per standard deviation (SD) increase of log2 ion intensity.

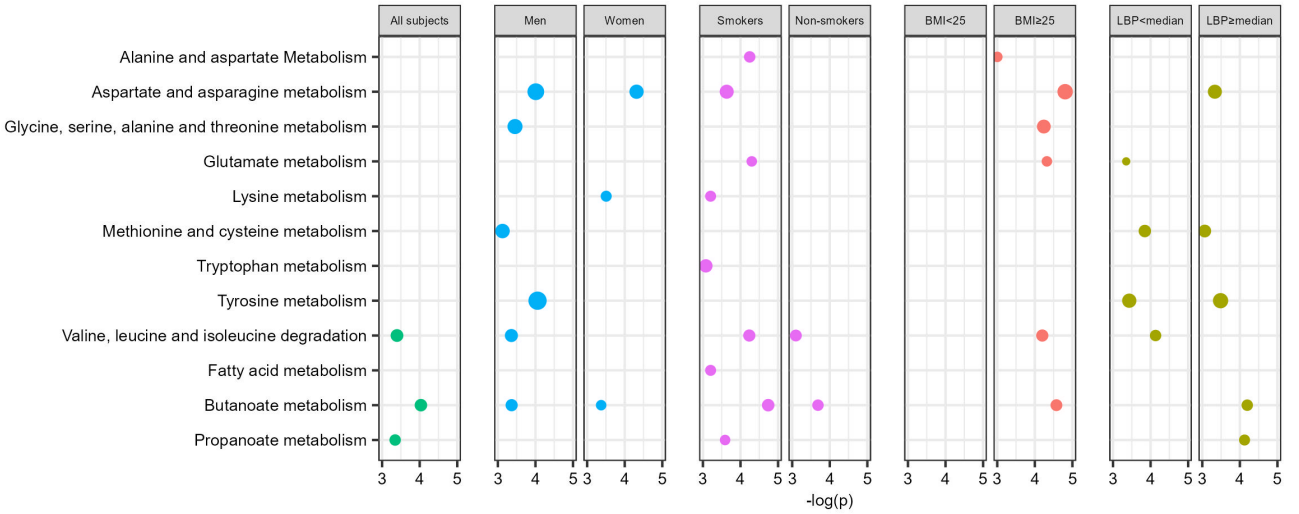


Figure 1. Microbiota-relevant pathways associated with Parkinson's disease in the main analyses and stratified analyses

Pathways enriched with PD-associated metabolic features were identified using MetaboAnalyst 5.0. The position of the dot on the *x*-axis corresponds to the -log10 *p*-value calculated for the pathway, and the size of the dot corresponds to the number of features meeting *p*-value<0.05 in the pathway.

Discussion

For the first time, we explored the associations between microbial metabolites and the future risk of PD within a large prospective cohort. We detected 13 metabolites which were nominally associated with PD risk. Furthermore, through the analysis of whole metabolome, we uncovered three microbiota-relevant pathways associated with PD risk, namely valine, leucine and isoleucine degradation, butanoate metabolism, and propanoate metabolism. Notably, individual characteristics might contribute to the variability in the involvement of gut microbiota in PD.

Over the past decade, numerous preclinical and clinical studies have shed light on the highly complex relationship between the gut and the brain in PD. The prevailing is that there are two distinct PD phenotypes: the 'gut-first' and 'brain-first', each with different symptomatic trajectories (17). Pre-diagnostic changes of gut microbiota may primarily be associated with the 'gut-first' phenotype but not the 'brain-first' one. However, the mechanisms behind the varied impacts of gut microbiota on PD remain unexplored.

Previous studies have suggested increased blood levels of both short-chain fatty acids and secondary bile acids among individuals with PD compared with healthy controls (sTable3) (6, 18). However, our study observed minimal changes in the abundance of these metabolites. These discrepancies may arise from differences in the timing of metabolite assessment, after-diagnostic versus pre-diagnostic samples. Nonetheless, alterations in the pathway level of short-chain fatty acid metabolism (butanoate and propanoate) were found to be associated with PD in our study. Therefore, further exploration of the involvement and mechanisms of short-chain fatty acids and secondary bile acids in the gut-brain axis is warranted.

Our study presents novel evidence of possible involvement of certain microbial metabolites in PD development, and their observed associations aligns with their previously implicated roles in the neurological system. For example, phenylalanine and L-2-hydroxyglutarate, recognized for their neurotoxic effects (19, 20), were associated with increased PD risk, whereas dihydroresveratrol, known for its neuroprotective properties (21), showed an inverse association with PD. Notably, phenylalanine, being a precursor of dopamine, may be influenced by dopamine medication. However, this bias is mitigated by utilizing pre-diagnostic samples from PD cases in our study. Interestingly, despite its well-known anti-inflammatory and antioxidant properties (22), indole-3-propionic acid displayed a positive relationship with PD risk.

In addition to their direct effects on the brain, blood metabolites and pathways may serve as indicator of gut-related pathogenic processes in PD. α-Synuclein aggregation, a PD hallmark, potentially originates in the enteric nervous system in cases of the 'gut-

first' PD type (1). Arginylation, a post-transcriptional modifications of α-synuclein involving the transfer of arginine to aspartate or glutamate (23), has been revealed to reduce the aggregation of α-synuclein, subsequently mitigating neurotoxicity (24). Notably, in our study, one arginylation product, arginylglutamate, showed a positive association with PD risk, suggesting an increased arginylation process of α-synuclein preceding the onset of PD. In a healthy gut, microbiota reside as biofilms on the intestinal mucus (25). However, under environmental pressures, these biofilms get disassembled, leading to the release of pathogenic bacteria. These pathogens might cause increased intestinal inflammation and permeability—two critical steps implicated in the gut-brain crosstalk of PD (26). Microbiota-produced amino acids, like aspartate, leucine, methionine, tyrosine and tryptophan, can trigger biofilm disassembly (27, 28). Importantly, our study linked pathways involving these amino acids with PD among analyses of all subjects or subgroups. Arginylation of α-synuclein and biofilm disassembly need further confirmation in the gut-related pathogenesis of PD.

This study noted differences in metabolite profiles and pathway enrichment related to sex, smoking status, and BMI. Similarly, our previous study on another microbiota-related marker, LBP, also revealed sex- and BMI-specific effects (11). Among smokers and overweight/obese individuals, more microbial pathways were found to be associated with PD, and stronger association between LBP and PD risk was observed among overweight/obese individuals. The effect modification of smoking and obesity on the relationship of microbial dysbiosis and PD is plausible. Previous research has indicated that both factors can induce gut microbiota dysbiosis, increased intestinal permeability and contribute to gut inflammation in animal models (29, 30). However, sex appears to play distinct roles. The positive association between LBP and PD risk was observed exclusively among women (11), whereas a larger number of microbial pathways were identified for men in the current study. These sex-related differences may be attributed to sex hormones, which are known as a potent driver to affect the microbiota composition (31). Additionally, diets, antibiotics, and environmental factors impact gut microbiota in a sex-dependent manner. In summary, our research suggests that personal traits and lifestyles might influence the involvement of the gut-brain axis in PD, potentially contributing to differentiation of 'gut-first' and 'brain-first' subtypes. Further investigation is warranted to elucidate the mechanisms through which these factors affect microbial metabolites and other microbiota-related markers, such as LBP.

Besides personal characteristics, this study observed the impact of LBP on the relationship between metabolites and PD. Individuals with higher levels of LBP exhibited more microbiota-relevant pathways compared to those with lower LBP levels. Elevated LBP levels has been proposed as a potential biomarker for intestinal

permeability (32). Therefore, our findings imply the presence of intestinal leakage and increased microbial metabolites in the blood.

Our study's strengths include using blood samples collected up to 18 years before PD diagnosis, minimizing influences from prodromal symptoms and medication. Second, our sample size featuring 351 case-control pairs is among the largest studies of microbial metabolites in PD. This surpasses previous studies, which included a maximum of 96 PD patients (Table S3). Third, our untargeted metabolomics framework broadens the scope of microbial alterations examined, including both metabolite and pathway analyses. This expansion enables a more comprehensive analysis of the role of gut microbiota in PD development. Fourth, the cases and controls in this study were drawn from the general European population, and their demographics are similar to those reported in other community-based European PD research, albeit with slightly younger age at diagnosis age. Therefore, our findings can be fairly generalized to the wider population of individuals affected by PD.

However, some limitations are acknowledged. First, extended prodromal phases of PD could still introduce residual reverse causality. Specifically, early prodromal symptoms of PD such as chronic constipation and loss of taste and smell, which may occur up to 10 years before the onset of the cardinal motor symptoms (33), might influence the gut microbiota. Sensitivity analysis with a cut-off of 8 years before diagnosis was consistent with the main findings. Additionally, an extended analysis regressing time-to-diagnosis (varying from <1 to 18 years) to the 13 metabolites prospectively related to PD in our study showed no effect between time-to-diagnosis and abundance of these metabolites (Figure S4). We therefore infer that the influence of the prodromal stage was limited in our study. Second, many microbial metabolites are co-metabolized by the host and microbiota. It is difficult to distinguish the contributions from microbiota, particularly without direct measurement of the gut microbiota. Third, the microbiota undergoes a rapid turnover and exhibit high heterogenous between individuals. Even with the large sample size available, this study might still lack sufficient power to detect statistically significant changes.

In summary, among the reported microbiota-relevant metabolites in humans (4), our study singled out a number of metabolites nominally associated pre-diagnostically with PD incidence. These functional metabolites were linked to biological pathways including valine, leucine and isoleucine degradation, butanoate metabolism, and propanoate metabolism. To our knowledge, this is the first large-scale prospective study examining metabolites relevant to gut microbiota in pre-clinical PD. The results of this study warrant further investigation of the role of the gut microbiota in PD development.

References

1. Braak H, de Vos RA, Bohl J, Del Tredici K. Gastric alpha-synuclein immunoreactive inclusions in Meissner's and Auerbach's plexuses in cases staged for Parkinson's disease-related brain pathology. Neurosci Lett. 2006;396(1):67-72.

2. Romano S, Savva GM, Bedarf JR, Charles IG, Hildebrand F, Narbad A. Meta-analysis of the Parkinson's disease gut microbiome suggests alterations linked to intestinal inflammation. NPJ Parkinson's disease. 2021;7(1):27.

3. Zierer J, Jackson MA, Kastenmuller G, et al. The fecal metabolome as a functional readout of the gut microbiome. Nat Genet. 2018;50(6):790-5.

4. Neveu V, Nicolas G, Amara A, Salek RM, Scalbert A. The human microbial exposome: expanding the Exposome-Explorer database with gut microbial metabolites. Sci Rep. 2023;13(1):1946.

5. Liu J, Tan Y, Cheng H, Zhang D, Feng W, Peng C. Functions of Gut Microbiota Metabolites, Current Status and Future Perspectives. Aging Dis. 2022;13(4):1106-26.

6. Chen SJ, Chen CC, Liao HY, et al. Association of Fecal and Plasma Levels of Short-Chain Fatty Acids With Gut Microbiota and Clinical Severity in Patients With Parkinson Disease. Neurology. 2022;98(8):e848-e58.

7. Shao Y, Li T, Liu Z, et al. Comprehensive metabolic profiling of Parkinson's disease by liquid chromatography-mass spectrometry. Mol Neurodegener. 2021;16(1):4.

8. Kalia LV, Lang AE. Parkinson's disease. Lancet. 2015;386(9996):896-912.

9. Shin C, Lim Y, Lim H, Ahn TB. Plasma Short-Chain Fatty Acids in Patients With Parkinson's Disease. Mov Disord. 2020;35(6):1021-7.

10. Gallo V, Brayne C, Forsgren L, et al. Parkinson's Disease Case Ascertainment in the EPIC Cohort: The NeuroEPIC4PD Study. Neurodegener Dis. 2015;15(6):331-8.

11. Zhao Y, Walker DI, Lill CM, et al. Lipopolysaccharide-binding protein and future Parkinson's disease risk: a European prospective cohort. J Neuroinflammation. 2023;20(1):170.

12. Riboli E, Hunt KJ, Slimani N, et al. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. Public health nutrition. 2002;5(6b):1113-24.

13. Soltow QA, Strobel FH, Mansfield KG, Wachtman L, Park Y, Jones DP. High-performance metabolic profiling with dual chromatography-Fourier-transform mass spectrometry (DC-FTMS) for study of the exposome. Metabolomics. 2013;9(1 Suppl):S132-s43.

14. Lazer C. imputeLCMD: A Collection of Methods for Left-Censored Missing Data Imputation. R package version 2.0. https://CRAN.R-project.org/package=imputeLCMD. 2015.

15. Pang Z, Zhou G, Ewald J, et al. Using MetaboAnalyst 5.0 for LC-HRMS spectra processing, multi-omics integration and covariate adjustment of global metabolomics data. Nat Protoc. 2022;17(8):1735-61.

16. Tan AH, Lim SY, Lang AE. The microbiome-gut-brain axis in Parkinson disease - from basic research to the clinic. Nat Rev Neurol. 2022;18(8):476-95.

17. Borghammer P, Van Den Berge N. Brain-First versus Gut-First Parkinson's Disease: A Hypothesis. J Parkinsons Dis. 2019;9(s2):S281-S95.

18. Li P, Killinger BA, Ensink E, et al. Gut Microbiota Dysbiosis Is Associated with Elevated Bile Acids in Parkinson's Disease. Metabolites. 2021;11(1).

19. Kranendijk M, Struys EA, Salomons GS, Van der Knaap MS, Jakobs C. Progress in understanding 2-hydroxyglutaric acidurias. J Inherit Metab Dis. 2012;35(4):571-87.

20. Rovelli V, Longo N. Phenylketonuria and the brain. Mol Genet Metab. 2023;139(1):107583.

21. Chung JY, Jeong JH, Song J. Resveratrol Modulates the Gut-Brain Axis: Focus on Glucagon-Like Peptide-1, 5-HT, and Gut Microbiota. Front Aging Neurosci. 2020;12:588044.

22. Negatu DA, Gengenbacher M, Dartois V, Dick T. Indole Propionic Acid, an Unusual Antibiotic Produced by the Gut Microbiota, With Anti-inflammatory and Antioxidant Properties. Front Microbiol. 2020;11:575586.

23. Zhang J, Li X, Li JD. The Roles of Post-translational Modifications on alpha-Synuclein in the Pathogenesis of Parkinson's Diseases. Front Neurosci. 2019;13:381.

24. Wang J, Han X, Leu NA, et al. Protein arginylation targets alpha synuclein, facilitates normal brain health, and prevents neurodegeneration. Sci Rep. 2017;7(1):11323.

25. Buret AG, Allain T. Gut microbiota biofilms: From regulatory mechanisms to therapeutic targets. J Exp Med. 2023;220(3).

26. Klann EM, Dissanayake U, Gurrala A, et al. The Gut-Brain Axis and Its Relation to Parkinson's Disease: A Review. Front Aging Neurosci. 2021;13:782082.

27. Kolodkin-Gal I, Romero D, Cao S, Clardy J, Kolter R, Losick R. D-amino acids trigger biofilm disassembly. Science. 2010;328(5978):627-9.

28. Warraich AA, Mohammed AR, Perrie Y, Hussain M, Gibson H, Rahman A. Evaluation of anti-biofilm activity of acidic amino acids and synergy with ciprofloxacin on Staphylococcus aureus biofilms. Sci Rep. 2020;10(1):9021.

29. Frazier TH, DiBaise JK, McClain CJ. Gut microbiota, intestinal permeability, obesity-induced inflammation, and liver injury. JPEN J Parenter Enteral Nutr. 2011;35(5 Suppl):14S-20S.

30. Gui X, Yang Z, Li MD. Effect of Cigarette Smoke on Gut Microbiota: State of Knowledge. Front Physiol. 2021;12:673341.

31. Valeri F, Endres K. How biological sex of the host shapes its gut microbiota. Frontiers in Neuroendocrinology. 2021;61.

32. Seethaler B, Basrai M, Neyrinck AM, et al. Biomarkers for assessment of intestinal permeability in clinical practice. Am J Physiol Gastrointest Liver Physiol. 2021;321(1):G11-G7.

33. Pellicano C, Benincasa D, Pisani V, Buttarelli FR, Giovannelli M, Pontieri FE. Prodromal non-motor symptoms of Parkinson's disease. Neuropsychiatr Dis Treat. 2007;3(1):145-52.

Chapter 7 Supplemental materials

Text S1. Methods for untargeted metabolomics measurement and compound annotation

Table S1. Annotated microbial metabolites and results of metabolome-wide association study with Parkinson's disease

Table S2. Basic information/demographics of participants from the case-control study nested in the EPIC4PD cohort

Table S3. Summary of associations between short-chain fatty acids and secondary bile acids with Parkinson's disease from the literature

Figure S1. Numbers of features associated with Parkinson's disease from metabolome-wide association study

Figure S2. Comparison of metabolome-wide association study results between main analyses and sensitivity analyses

Figure S3. Comparison of metabolome-wide association study results between main analyses with analyses adjusting food intake

Figure S4. Nature spline regression for associations between time-to-diagnosis and metabolites among PD cases

Text S1. Methods for untargeted metabolomics measurement and compound annotation

To profile circulating microbial metabolites in EPIC4PD, we performed global metabolomics analysis of all plasma samples using a liquid chromatography-high resolution mass spectrometry (LC-HRMS)-based platform (1, 2). Further de novo structural annotation was conducted for untargeted, semi-quantitative profiling of microbial metabolites through an integrated cheminformatic approach.

Sample preparation

Blood plasma samples were thawed on ice, and a 50 µL sample was aliquoted and extracted into 100 µL cold acetonitrile (ACN) spiked with internal standard mix (ACN:plasma, 2:1, *v/v*). The extracts were cold incubated under -20 ºC before centrifugation for protein and particulate removal. Ten µL of the supernatant was injected on LC-MS, loading ~3.33 µL raw plasma on the column. Sample extracts were injected in triplicate alongside intermittent injection of quality control (QC) extracts of NIST1953 and BioIVT plasma, obtained respectively from the National Institute of Standards and Technology (NIST) (Gaithersburg, MD, USA) and Bioreclamation Inc. (New Cassel, NY, USA).

Instrumental analysis

The injected extracts were chromatographically separated, ionized, and analyzed on a Thermo Fisher Scientific Vanquish dual chromatograph coupling to a high-resolution accurate-mass (HRAM) Orbitrap Exploris™ 240 mass spectrometer interfaced with a heated electrospray ionization (ESI) source (Waltham, MA, USA). To maximize the detection of polar and nonpolar metabolites, two complementary analyses were performed, namely hydrophilic interaction liquid chromatography (HILIC)-ESI(+) and reverse-phase chromatography (RPC)-ESI(-), which were further referred to as "HILIC-ESI$^+$" and "C18-ESI$^-$," both operated in full scan mode at 120,000 mass resolution (FWHM) with a mass-to-charge (*m/z*) range of 85-1,275. Stringent quality assurance and quality control (QA/QC) procedures were applied, including sample randomization, timely mass calibration, and routine check of instrumental sensitivity. For this study, a total of 4,872 *.RAW data were acquired (18 batches), covering triplicate sample injections and intermittent QC injections.

Data processing

The acquired *.RAW data were converted to *.mzXML format using ProteoWizard msConvert and subject to feature extraction and sample alignment using the apLCMS package of R (3). Crude alignment feature tables yielded separately for HILIC-ESI$^+$ and

C_{18}-ESI$^-$, both consisting of uniquely detected features (*m/z*, retention time) and their ion abundance (peak area) across samples. The feature tables are processed through xMSanalyzer (4) where ion abundances were median summarized, cleaned up by QC, and corrected for batch effects (ComBat), yielding 9,435 ion features for HILIC-ESI$^+$ and 8,439 for C_{18}-ESI$^-$ for statistical and informatic analyses.

Compound identification

Structural annotation of compounds derived from microbiota was implemented through an integrated cheminformatic strategy. We first retrieved a set of 457 metabolites relevant to gut microbiota, by referring to the Exposome-Explorer microbial exposome database (5). We built internal RT-m/z libraries respectively for HILIC-ESI$^+$ and C_{18}-ESI$^-$ modes using hundreds of authentic chemical standards purchased from IROA Technologies (Chapel Hill, NC, USA) and Sigma Aldrich (St. Louis, MO, USA), from which some acknowledged microbial metabolites were included for quick and reliable identification. Meanwhile, to expand the coverage, we leveraged in silico cheminformatic analyses for annotating all plausible metabolites relevant to microbiota. Using accurate *m/z*, isotopic ratios, and chromatographic retention time (RT), we annotated detected peaks based on formula prediction (Seven Golden Rules) (6, 7) and chromatographic retention time estimation (Retip, XGBoost algorithm) (8), and manually curated based on extensive bioanalytical inferences and expert consultation. The level of annotation confidence was assigned based on the Schymanski scale for LC-HRMS metabolomics practice (9, 10). Annotation confidence was assigned as Level 1 for features matched with our in-house library, and Level 2 for features with predicted parameters but not validated by authentic chemical standards.

Table S1. Annotated microbial metabolites and results of metabolome-wide association study with Parkinson's disease

Metabolite name	Evidence on microbial origin[1]	LC-MS mode	Retention time (s)	m/z	Annotation level[2]	OR (95% CI)[3]
Amino acids and peptides, n=42						
Asparagine	Both	HILIC-ESI$^+$	133.0608	85.7	1	0.98 (0.81-1.18)
Aspartic acid	Both	C18-ESI$^-$	132.0302	23.5	1	1.02 (0.83-1.24)
Citrulline (C18neg)	Both	C18-ESI$^-$	174.0885	29.2	1	0.99 (0.85-1.15)
Citrulline (HILpos)	Both	HILIC-ESI$^+$	176.1031	83	1	1.15 (0.98-1.36)
Glutamic acid (C18neg)	Both	C18-ESI$^-$	146.0459	27.5	1	1.05 (0.88-1.25)
Glutamic acid (HILpos)	Both	HILIC-ESI$^+$	148.0606	68.9	1	1.03 (0.84-1.25)
Histidine	Both	C18-ESI$^-$	154.0623	35.3	1	1.09 (0.92-1.30)
Lysine	Both	C18-ESI$^-$	145.0982	54.1	2	0.87 (0.72-1.05)
Methionine (C18neg)	Both	C18-ESI$^-$	148.0438	32.5	2	0.99 (0.84-1.17)
Methionine (HILpos)	Both	HILIC-ESI$^+$	150.0578	47.9	1	1.02 (0.83-1.24)
Ornithine	Both	C18-ESI$^-$	131.0827	35.5	2	1.02 (0.86-1.21)
Phenylalanine (C18neg)	Both	C18-ESI$^-$	164.0716	34.3	2	1.19 (1.01-1.41)
Phenylalanine (HILpos)	Both	HILIC-ESI$^+$	166.0839	46.4	1	1.09 (0.92-1.28)
Proline betaine	Both	HILIC-ESI$^+$	144.102	57.5	1	0.95 (0.80-1.13)
Tyrosine (C18neg)	Both	C18-ESI$^-$	180.0667	33.7	1	1.04 (0.89-1.22)
Tyrosine (HILpos)	Both	HILIC-ESI$^+$	182.0809	53.1	1	1.04 (0.87-1.23)
delta-Aminovaleric acid	Both	C18-ESI$^-$	116.0718	30.4	1	1.18 (1.00-1.40)
3,4-Dihydroxy-L-phenylalanine	Only in vivo	C18-ESI$^-$	196.0593	33	1	0.98 (0.83-1.15)
4-Hydroxyproline	Only in vivo	HILIC-ESI$^+$	132.0654	63.3	1	0.97 (0.83-1.13)
Arginine	Only in vivo	HILIC-ESI$^+$	175.1191	92.7	2	0.98 (0.84-1.16)
Arginylglutamate (C18neg)	Only in vivo	C18-ESI$^-$	302.1553	32.2	2	1.22 (1.03-1.45)
Arginylglutamate (HILpos)	Only in vivo	HILIC-ESI$^+$	304.16	105.3	2	1.08 (0.92-1.28)
Arogenic acid	Only in vivo	HILIC-ESI$^+$	228.0812	55.7	2	1.02 (0.87-1.21)
Aspartylhistidine	Only in vivo	C18-ESI$^-$	269.0876	32	2	1.05 (0.87-1.26)
Creatine (C18neg)	Only in vivo	C18-ESI$^-$	130.0594	34.9	2	1.12 (0.92-1.36)
Creatine (HILpos)	Only in vivo	HILIC-ESI$^+$	132.0767	62.3	2	0.97 (0.82-1.16)
Creatinine	Only in vivo	HILIC-ESI$^+$	114.0662	33.6	1	1.22 (1.01-1.48)
D-Alanyl-D-alanine	Only in vivo	HILIC-ESI$^+$	161.0922	73.9	2	1.01 (0.86-1.18)
Diaminopimelic acid	Only in vivo	HILIC-ESI$^+$	191.1027	81.4	2	1.06 (0.83-1.37)
Dimethylarginine (ADMA)	Only in vivo	HILIC-ESI$^+$	203.1504	76.9	1	1.22 (1.00-1.51)
Glutamine (C18neg)	Only in vivo	C18-ESI$^-$	145.0619	29.7	1	1.03 (0.87-1.21)
Glutamine (HILpos)	Only in vivo	HILIC-ESI$^+$	147.0766	79.5	1	1.04 (0.88-1.24)
Isoleucine	Only in vivo	HILIC-ESI$^+$	132.102	40.3	1	1.07 (0.89-1.28)
Leucylproline	Only in vivo	HILIC-ESI$^+$	229.1548	50	2	1.27 (1.03-1.55)
N-Acetyl-L-glutamic acid	Only in vivo	HILIC-ESI$^+$	190.0762	40.6	2	1.06 (0.89-1.25)
Phenylacetylglutamine	Only in vivo	C18-ESI$^-$	263.1038	30.1	2	1.04 (0.89-1.21)
Phenylacetylglycine	Only in vivo	HILIC-ESI$^+$	194.0811	20	1	0.95 (0.79-1.14)
Pyroglutamic acid (C18neg)	Only in vivo	C18-ESI$^-$	128.0354	29.5	2	1.00 (0.80-1.26)
Pyroglutamic acid (HILpos)	Only in vivo	HILIC-ESI$^+$	130.0499	35.4	1	1.02 (0.86-1.21)
Threonine (C18neg)	Only in vivo	C18-ESI$^-$	118.0511	38.3	2	0.99 (0.84-1.16)
Threonine (HILpos)	Only in vivo	HILIC-ESI$^+$	120.0655	68.2	1	1.02 (0.85-1.23)
beta-Alanine	Only in vivo	HILIC-ESI$^+$	90.055	60.4	1	1.09 (0.90-1.32)

Fatty acyls, n=20						
4-Butyric acid betaine	Both	HILIC-ESI$^+$	146.1177	43	2	0.96 (0.79-1.16)
Butyric acid (4:0)	Both	HILIC-ESI$^+$	89.0597	28.2	2	0.91 (0.76-1.08)
Margaric acid (17:0)	Both	C18-ESI$^-$	269.2486	168.7	1	0.93 (0.79-1.10)
Palmitic acid (16:0)	Both	C18-ESI$^-$	255.233	161.3	2	0.93 (0.79-1.11)
Palmitoleic acid (cis-16:1n-7) (C18neg)	Both	C18-ESI$^-$	253.2173	148.9	1	0.98 (0.82-1.16)
Palmitoleic acid (cis-16:1n-7) (HILpos)	Both	HILIC-ESI$^+$	255.2321	24.9	2	0.96 (0.82-1.14)
Acetylcarnitine	Only in vivo	HILIC-ESI$^+$	204.123	38.9	1	1.10 (0.93-1.29)
Dihomo-gamma-linolenic acid (DGLA) (C18neg)	Only in vivo	C18-ESI$^-$	305.2482	154.7	1	1.00 (0.84-1.19)
Dihomo-gamma-linolenic acid (DGLA) (HILpos)	Only in vivo	HILIC-ESI$^+$	307.2627	18.4	2	1.04 (0.88-1.23)
Myristic acid (14:0)	Only in vivo	C18-ESI$^-$	227.2017	144.7	2	0.94 (0.79-1.11)
Stearic acid (18:0)	Only in vivo	C18-ESI$^-$	283.2644	181.7	2	0.96 (0.81-1.13)
5-Dodecenoic acid (C18neg)	Only in vitro	C18-ESI$^-$	197.1547	104.5	2	0.87 (0.70-1.08)
5-Dodecenoic acid (HILpos)	Only in vitro	HILIC-ESI$^+$	199.1694	25.2	2	1.06 (0.83-1.36)
Conjugated linoleic acids (CLA)	Only in vitro	HILIC-ESI$^+$	281.2477	24.6	2	1.04 (0.88-1.22)
Isocaproic acid (6:0)	Only in vitro	HILIC-ESI$^+$	117.091	22.4	1	0.98 (0.80-1.19)
Lauric acid (12:0)	Only in vitro	C18-ESI$^-$	199.1703	124.8	2	0.86 (0.73-1.03)
Myristoleic acid (cis-14:1n-5) (C18neg)	Only in vitro	C18-ESI$^-$	225.186	133.2	2	0.95 (0.80-1.14)
Myristoleic acid (cis-14:1n-5) (HILpos)	Only in vitro	HILIC-ESI$^+$	227.2007	24.4	2	1.00 (0.83-1.19)
Oleic acid (cis-18:1n-9) (C18neg)	Only in vitro	C18-ESI$^-$	281.2487	163.9	2	1.02 (0.86-1.20)
Oleic acid (cis-18:1n-9) (HILpos)	Only in vitro	HILIC-ESI$^+$	283.2635	24.9	2	1.01 (0.85-1.20)
Carbohydrates and conjugates, n=10						
2-Dehydro-3-deoxy-D-arabino-heptonate-7-phosphate	Only in vivo	C18-ESI$^-$	287.0126	52	2	1.02 (0.83-1.24)
D-Ribulose-5-phosphate	Only in vivo	C18-ESI$^-$	229.018	31.6	2	1.12 (0.96-1.31)
Glucose 6-phosphate	Only in vivo	C18-ESI$^-$	259.0283	32.6	2	1.01 (0.87-1.18)
Glyceric acid	Only in vivo	C18-ESI$^-$	105.0194	27.1	1	0.98 (0.80-1.19)
Indole-3-carboxylic acid glucuronide	Only in vivo	HILIC-ESI$^+$	338.081	68.9	2	1.03 (0.85-1.25)
N-Acetylglucosamine	Only in vivo	C18-ESI$^-$	220.0831	46.8	2	0.89 (0.73-1.09)
N-Acetylneuraminic acid	Only in vivo	HILIC-ESI$^+$	310.1132	86.7	2	0.89 (0.75-1.06)
Xylose (C18neg)	Only in vivo	C18-ESI$^-$	149.0456	37.3	2	1.25 (1.05-1.49)
Xylose (HILpos)	Only in vivo	HILIC-ESI$^+$	151.0618	47.4	2	1.04 (0.86-1.26)
p-Cresol glucuronide	Only in vivo	HILIC-ESI$^+$	285.0994	40.1	2	1.11 (0.94-1.32)
Indoles and derivatives, n=10						
Indole-3-aldehyde	Both	HILIC-ESI$^+$	146.0601	40.2	2	1.04 (0.87-1.24)
Indoleacetic acid	Both	C18-ESI$^-$	174.0561	43.7	2	0.96 (0.81-1.15)
Indolelactic acid	Both	C18-ESI$^-$	204.0666	31.8	2	0.99 (0.84-1.17)
Indolepropionic acid (C18neg)	Both	C18-ESI$^-$	188.0718	32.6	1	1.02 (0.87-1.20)
Indolepropionic acid (HILpos)	Both	HILIC-ESI$^+$	190.0864	34.2	1	1.22 (1.05-1.43)
Tryptophan (C18neg)	Both	C18-ESI$^-$	203.0825	33.7	2	1.03 (0.87-1.21)
Tryptophan (HILpos)	Both	HILIC-ESI$^+$	205.0973	41.1	1	1.06 (0.90-1.26)
Indoleacrylic acid	Only in vitro	HILIC-ESI$^+$	188.0707	41	2	1.04 (0.88-1.24)
Indolepyruvic acid	Only in vitro	C18-ESI$^-$	202.0492	32.3	2	0.98 (0.84-1.15)
Skatole	Only in vitro	C18-ESI$^-$	130.0663	65.3	2	0.85 (0.72-1.01)

Phenylpropanoids and polyketides, n=10						
3-Hydroxyphenylpropionic acid	Both	HILIC-ESI$^+$	167.0709	17.8	2	1.00 (0.85-1.17)
3-Phenylpropionic acid	Both	C18-ESI$^-$	149.0609	40.1	2	0.90 (0.78-1.05)
Dihydrodaidzein	Both	C18-ESI$^-$	255.0632	77.4	2	1.14 (0.94-1.38)
Dihydroferulic acid	Both	C18-ESI$^-$	195.0664	19.7	1	1.02 (0.87-1.18)
Equol	Both	C18-ESI$^-$	241.0918	58.5	2	1.00 (0.85-1.19)
Ferulic acid	Both	C18-ESI$^-$	193.0499	52.9	2	0.92 (0.79-1.08)
Gentiobiose	Only in vivo	HILIC-ESI$^+$	343.1233	91	2	1.09 (0.91-1.31)
Cinnamic acid (C18neg)	Only in vitro	C18-ESI$^-$	147.0452	32.9	2	1.10 (0.93-1.30)
Cinnamic acid (HILpos)	Only in vitro	HILIC-ESI$^+$	149.0597	41.1	2	1.11 (0.96-1.28)
Dihydroresveratrol	Only in vitro	C18-ESI$^-$	229.0808	73.6	2	0.84 (0.70-1.00)
Purines and derivatives, n=9						
Adenine (C18neg)	Only in vivo	C18-ESI$^-$	134.0473	50.7	2	0.93 (0.75-1.15)
Adenine (HILpos)	Only in vivo	HILIC-ESI$^+$	136.0619	42.5	2	1.10 (0.92-1.31)
Guanine	Only in vivo	HILIC-ESI$^+$	152.0535	68.3	2	0.95 (0.78-1.17)
Hypoxanthine (C18neg)	Only in vivo	C18-ESI$^-$	135.0304	28.6	1	1.01 (0.82-1.25)
Hypoxanthine (HILpos)	Only in vivo	HILIC-ESI$^+$	137.0452	40.7	1	1.11 (0.84-1.48)
Uric acid (C18neg)	Only in vivo	C18-ESI$^-$	167.021	30.4	2	0.98 (0.83-1.17)
Uric acid (HILpos)	Only in vivo	HILIC-ESI$^+$	169.0361	57.5	1	1.17 (0.96-1.43)
Xanthine	Only in vivo	HILIC-ESI$^+$	153.0408	42.3	1	1.03 (0.87-1.23)
Purine	Only in vitro	C18-ESI$^-$	119.0351	34.7	2	1.24 (1.03-1.48)
Keto acids and derivatives, n=7						
2-Oxobutyric acid (C18neg)	Only in vivo	C18-ESI$^-$	101.0245	30.7	2	1.15 (0.97-1.36)
2-Oxobutyric acid (HILpos)	Only in vivo	HILIC-ESI$^+$	103.039	53.1	2	0.94 (0.79-1.11)
2-Oxoglutaric acid	Only in vivo	C18-ESI$^-$	145.0142	27.3	1	0.89 (0.71-1.12)
3-Methyl-2-oxobutyric acid (C18neg)	Only in vivo	C18-ESI$^-$	115.0401	31.6	2	1.19 (0.95-1.49)
3-Methyl-2-oxobutyric acid (HILpos)	Only in vivo	HILIC-ESI$^+$	117.0567	38.7	2	1.08 (0.91-1.28)
4-Methyl-2-oxovaleric acid	Only in vivo	C18-ESI$^-$	129.0558	32.4	2	1.06 (0.86-1.30)
Oxaloacetic acid	Only in vitro	C18-ESI$^-$	130.9986	27.7	2	1.13 (0.93-1.36)
Phenols, n=7						
1,2-Dihydroxybenzene	Both	C18-ESI$^-$	109.0296	31.3	2	1.00 (0.86-1.16)
3,4-Dihydroxybenzeneacetic acid	Both	C18-ESI$^-$	167.0349	29.6	1	1.15 (0.95-1.38)
Homovanillic acid	Both	C18-ESI$^-$	181.0507	28.1	2	1.13 (0.96-1.33)
5-(3',4',5'-Trihydroxyphenyl)-gamma-valerolactone	Only in vitro	C18-ESI$^-$	223.0575	30.1	2	0.89 (0.75-1.06)
5-(3',4'-Dihydroxyphenyl)-gamma-valerolactone	Only in vitro	C18-ESI$^-$	207.0664	30.4	2	1.05 (0.89-1.24)
5-(3',4'-Dihydroxyphenyl)-valeric acid	Only in vitro	C18-ESI$^-$	209.0809	36.2	2	0.99 (0.84-1.18)
5-(3'-Hydroxyphenyl)-gamma-valerolactone	Only in vitro	C18-ESI$^-$	191.071	62	2	0.99 (0.84-1.17)
Organic acids and derivatives, n=6						
Guaiacol sulfate	Only in vivo	C18-ESI$^-$	203.0004	29.6	2	0.97 (0.83-1.13)
Indoxyl sulfate	Only in vivo	C18-ESI$^-$	212.0016	32	1	1.02 (0.88-1.18)
Taurine (C18neg)	Only in vivo	C18-ESI$^-$	124.0075	31.3	2	0.88 (0.74-1.05)
Taurine (HILpos)	Only in vivo	HILIC-ESI$^+$	126.0219	119.7	2	1.17 (0.98-1.38)
p-Cresol sulfate	Only in vivo	C18-ESI$^-$	187.0012	35	2	1.02 (0.87-1.19)
N1-Acetylspermidine	Only in vitro	HILIC-ESI$^+$	188.1757	83.9	2	1.00 (0.84-1.20)

Benzoic acids and derivatives, n=5						
Gallic acid	Both	C18-ESI$^-$	169.0134	38.8	2	1.22 (1.03-1.46)
Hippuric acid (C18neg)	Both	C18-ESI$^-$	178.051	30.2	1	1.02 (0.88-1.19)
Hippuric acid (HILpos)	Both	HILIC-ESI$^+$	180.0658	29.2	1	1.08 (0.93-1.26)
2,6-Dihydroxybenzoic acid	Only in vitro	C18-ESI$^-$	153.0186	32.5	1	1.03 (0.87-1.22)
Syringic acid	Only in vitro	C18-ESI$^-$	197.0436	32.2	2	1.01 (0.82-1.25)
Bile acids and derivatives, n=5						
Chenodeoxycholic acid	Both	C18-ESI$^-$	391.286	100.1	1	1.13 (0.97-1.31)
Lithocholic acid	Both	C18-ESI$^-$	375.293	153.3	2	0.92 (0.78-1.08)
Glycodeoxycholic acid	Only in vivo	HILIC-ESI$^+$	450.3206	27.7	1	0.95 (0.79-1.13)
Taurocholic acid	Only in vivo	HILIC-ESI$^+$	516.3055	35.5	1	0.81 (0.67-0.96)
3-oxodeoxycholic acid	Only in vitro	C18-ESI$^-$	389.27	101.1	2	1.03 (0.89-1.19)
Pyrimidines and derivatives, n=5						
Cytosine	Only in vivo	HILIC-ESI$^+$	112.0507	50	2	1.11 (0.95-1.30)
Orotic acid	Only in vivo	HILIC-ESI$^+$	HILIC-ESI$^+$	94.7	2	1.11 (0.93-1.32)
Thymine	Only in vivo	HILIC-ESI$^+$	127.0503	27.4	2	1.09 (0.93-1.28)
Uracil	Only in vivo	HILIC-ESI$^+$	113.0346	30.7	1	0.94 (0.80-1.10)
Thiamin	Only in vitro	HILIC-ESI$^+$	265.1168	41.4	2	1.16 (0.98-1.37)
Organicnitrogen compounds, n=4						
Alanine betaine	Both	C18-ESI$^-$	130.0874	35.8	2	1.07 (0.90-1.28)
N-Oleoylethanolamine	Only in vivo	HILIC-ESI$^+$	326.3048	25.2	2	1.16 (0.95-1.41)
Choline	Only in vitro	HILIC-ESI$^+$	104.107	38.9	1	0.89 (0.75-1.07)
Histamine	Only in vitro	HILIC-ESI$^+$	112.0898	64.6	2	1.04 (0.89-1.21)
Organoheterocyclic compounds, n=4						
Piperidine	Both	HILIC-ESI$^+$	86.0965	40.3	2	1.05 (0.88-1.26)
Ascorbic acid (C18neg)	Only in vivo	C18-ESI$^-$	175.0248	29.9	2	1.09 (0.90-1.32)
Ascorbic acid (HILpos)	Only in vivo	HILIC-ESI$^+$	177.0441	66.2	2	0.92 (0.78-1.07)
Biotin	Only in vitro	C18-ESI$^-$	243.0749	54.7	2	1.13 (0.89-1.43)
Pyridines and derivatives, n=4						
4-Pyridoxic acid	Only in vivo	C18-ESI$^-$	182.0447	37.4	1	1.09 (0.92-1.29)
Nicotinamide	Only in vivo	HILIC-ESI$^+$	123.0553	35.8	1	1.11 (0.93-1.34)
Pyridoxal (C18neg)	Only in vivo	C18-ESI$^-$	166.051	58.4	2	1.08 (0.92-1.26)
Pyridoxal (HILpos)	Only in vivo	HILIC-ESI$^+$	168.0641	45.6	2	0.89 (0.75-1.06)
Alcohols and polyols, n=3						
4-Hydroxycyclohexylacetic acid	Only in vivo	C18-ESI$^-$	157.0871	31.8	1	0.91 (0.77-1.06)
Pantothenic acid	Only in vivo	HILIC-ESI$^+$	220.118	33	2	0.94 (0.81-1.09)
Quinic acid	Only in vitro	HILIC-ESI$^+$	193.0687	63.8	2	1.08 (0.90-1.29)
Carboxylic acids and derivatives, n=3						
Methylmalonic acid (MMA)	Both	C18-ESI$^-$	117.0194	27.9	2	1.02 (0.86-1.20)
Citric acid	Only in vitro	C18-ESI$^-$	191.0198	27.8	1	1.15 (0.95-1.40)
Malonic acid	Only in vitro	C18-ESI$^-$	103.0037	29	2	0.92 (0.74-1.13)
Hydroxy acids and derivatives, n=3						
Malic acid (C18neg)	Only in vivo	C18-ESI$^-$	133.0143	27.9	1	1.05 (0.88-1.25)
Malic acid (HILpos)	Only in vivo	HILIC-ESI$^+$	135.0276	48.9	2	1.15 (0.96-1.39)
L-2-Hydroxyglutaric acid	Only in vitro	C18-ESI$^-$	147.0299	29.2	1	1.40 (1.05-1.85)
Steroids and derivatives, n=3						
16alpha-Hydroxyestrone	Only in vitro	C18-ESI$^-$	285.1543	70.4	2	1.03 (0.88-1.21)
Estrone (C18neg)	Only in vitro	C18-ESI$^-$	269.1576	127	2	0.94 (0.80-1.10)
Estrone (HILpos)	Only in vitro	HILIC-ESI$^+$	271.166	23.4	1	1.06 (0.90-1.25)
Benzene and substituted derivatives, n=2						
4-Hydroxyphenylpyruvic acid	Only in vivo	HILIC-ESI$^+$	181.053	41	2	1.30 (1.11-1.52)
Phenylpyruvic acid	Only in vitro	C18-ESI$^-$	163.0401	34	2	0.99 (0.84-1.16)

Lipids and lipid-like molecules, n=2						
alpha-Tocopherol	Only in vivo	HILIC-ESI$^+$	431.3837	24.6	2	0.97 (0.82-1.14)
Phylloquinone	Only in vivo	HILIC-ESI$^+$	451.3608	23.9	2	1.04 (0.89-1.22)
Carbonyl compound, n=1						
Kynurenine	Only in vitro	HILIC-ESI$^+$	209.0921	42.7	1	1.08 (0.92-1.28)
Nucleoside, n=1						
Uridine	Only in vivo	C18-ESI$^-$	243.0622	33.8	1	1.01 (0.85-1.19)
Penol ether, n=1						
5-(3'-Methoxyphenyl)valeric acid	Only in vitro	HILIC-ESI$^+$	209.1172	25.4	2	0.96 (0.81-1.16)

[1] 'Both' indicated that the specific metabolite simultaneously met in vitro (production by gut bacteria) and in vivo (reduction of concentrations by antibiotics or in germ-free animals) evidence supporting a microbial origin. 'Only in vivo/in vitro' indicates that evidence was limited to either in vivo or in vitro studies.

[2] Features matched with the in-house library were assigned with Level 1 annotation, and features matched with predicted chemical retention time with Level 2 annotation.

[3] Main analyses included all subjects. Conditional logistic regression for the matched case-control sets, adjusted by smoking status and BMI. OR was calculated per standard deviation (SD) increase of log2 ion intensity.

Table S2. Basic information/demographics of participants from the case-control study nested in the EPIC4PD cohort

Characteristic	PD cases n=351	Controls n=351
Age at recruitment[1], median (range)	60 (38-76)	60 (38-76)
Age at PD diagnosis, median (range)	68 (45-86)	--
Years between recruitment and PD diagnosis, median (range)	8 (0.2-18)	--
Diagnosis validity		
Definite	45 (13%)	--
Very likely	143 (40%)	--
Probable	59 (17%)	--
Possible	104 (30%)	--
Sex[1], n (%)[a]		
Male	195 (55%)	195 (55%)
Female	156 (44%)	156 (44%)
Country[1], n (%)[a]		
Italy	54 (15%)	54 (15%)
Spain	97 (28%)	97 (28%)
UK	141 (40%)	141 (40%)
Netherlands	13 (4%)	13 (4%)
Germany	46 (13%)	46 (13%)
Smoking status at recruitment, n (%)		
Smokers	43 (12%)	55 (16%)
Non-smokers	298 (85%)	283 (80%)
Unknown	10 (3%)	13 (4%)
BMI categories, n (%)		
BMI<25 kg/m^2	118 (34%)	136 (39%)
BMI≥25 kg/m^2	233 (66%)	215 (61%)
LBP categories, n (%)		
LBP<25.6 µg/ml	165 (47%)	183 (52%)
LBP≥25.6 µg/ml	186 (53%)	168 (48%)

PD, Parkinson's disease; BMI, body mass index; LBP, lipopolysaccharide-binding protein.

[1] Matching variables.

Table S3. Summary of associations between short-chain fatty acids and secondary bile acids with Parkinson's disease from the literature

Study	N. of PD patients	N. of controls	Biospecimen	Direction of change in PD patients compared to controls		
				Increased	Decreased	Non-changed
Short-chain fatty acids						
Unger 2016 (11)	34	34	Feces		Acetic acid Butyric acid Propionic acid	Valeric acid iso-Butyric acid iso-Valeric acid
Aho 2021 (12)	55	56	Feces		Butyric acid Propionic acid	Acetic acid iso-Butyric acid iso-Valeric acid Valeric acid
Chen 2022 (13)	96	85	Feces		Acetic acid Propionic acid Butyric acid	
Shin 2020 (14)	38	33	Plasma	Acetic acid		Propionic acid Butyric acid
Chen 2022 (13)	96	85	Plasma	Propionic acid Butyric acid Valeric acid		Acetic acid
Secondary bile acids						
Li 2021 (15)	15	12	Appendix	Deoxycholic acid Lithocholic acid		Ursodeoxycholic acid
Cirstea 2020 (16)	75	50	Serum			Sulfolithocholic acid Deoxycholic acid
Shao 2021 (17)	36	43	Plasma	Deoxycholic acid		Lithocholic acid Ursodeoxycholic acid

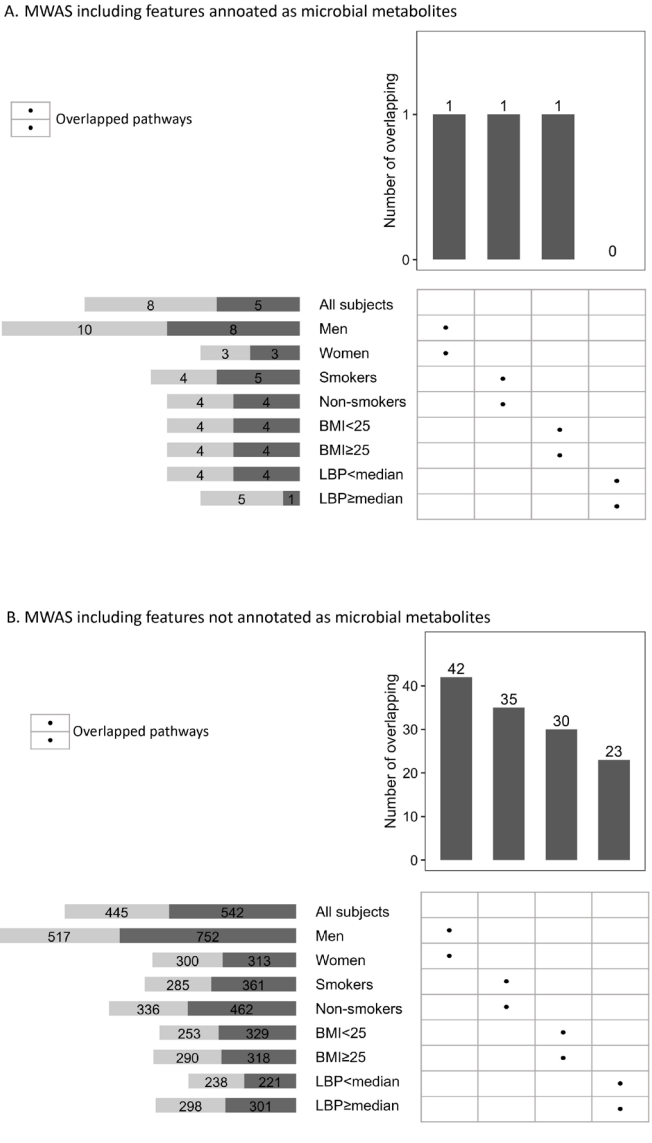


Figure S1. Numbers of features associated with Parkinson's disease from metabolome-wide association study

Panel A: metabolome-wide association study (MWAS) including features annotated as microbial metabolites (n=167). Panel B: MWAS including the rest of features (not annotated as microbial metabolites) (7,409 HILIC-ESI$^+$ and 6,879 C18-ESI$^-$ features).

In each panel, left part showed the number of features (dark grey for HILIC-ESI$^+$, light grey for C18-ESI$^-$) associated with PD risk (raw *p*-value<0.05) in the main analyses and stratified analyses. Upper part shows the number of overlapped features among different categories of stratified analyses.

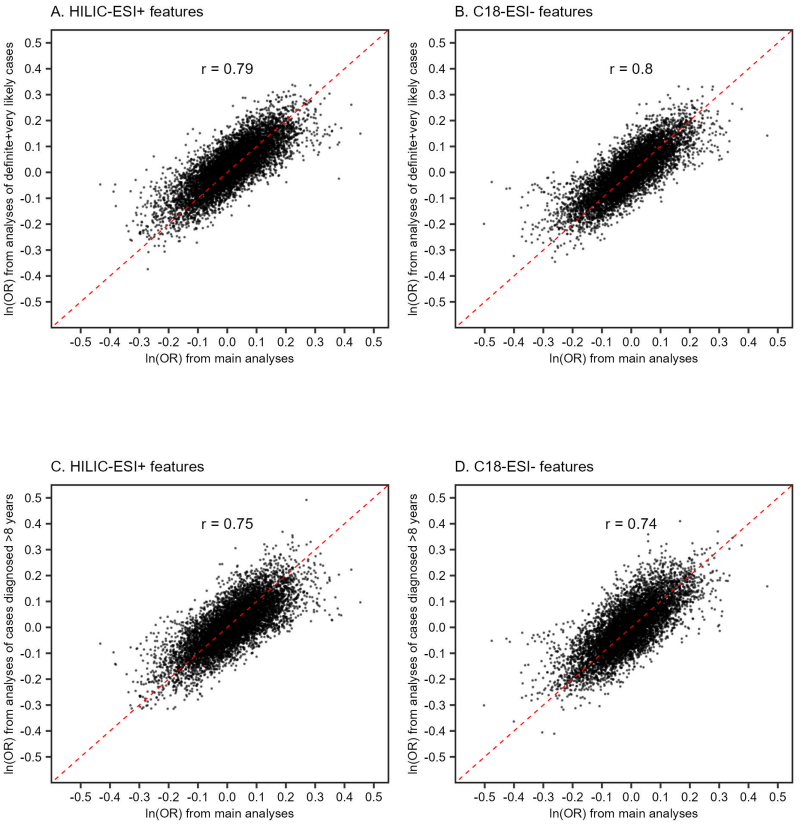


Figure S2. Comparison of metabolome-wide association study results between main analyses and sensitivity analyses

Panel A: effect estimate comparison between main analyses and analyses limiting to definite and very likely PD diagnoses for HILIC-ESI+ features. Panel B: effect estimate comparison between main analyses and analyses limiting to definite and very likely PD diagnoses for C18-ESI– features. Panel C: effect estimate comparison between main analyses and analyses limiting to cases diagnosed >8 years since recruitment for HILIC-ESI+ features. Panel D: effect estimate comparison between main analyses and analyses limiting to cases diagnosed >8 years since recruitment for C18-ESI– features.

In each panel, each dot represented one metabolic feature (in total 7,487 HILIC-ESI+ features and 6,968 C18-ESI– features). They were displayed by ln(OR) estimated from the main analyses (all subjects) against the corresponding values from sensitivity analyses. Pearson correlation coefficients were labelled on the plot. The dotted line is the line of identity.

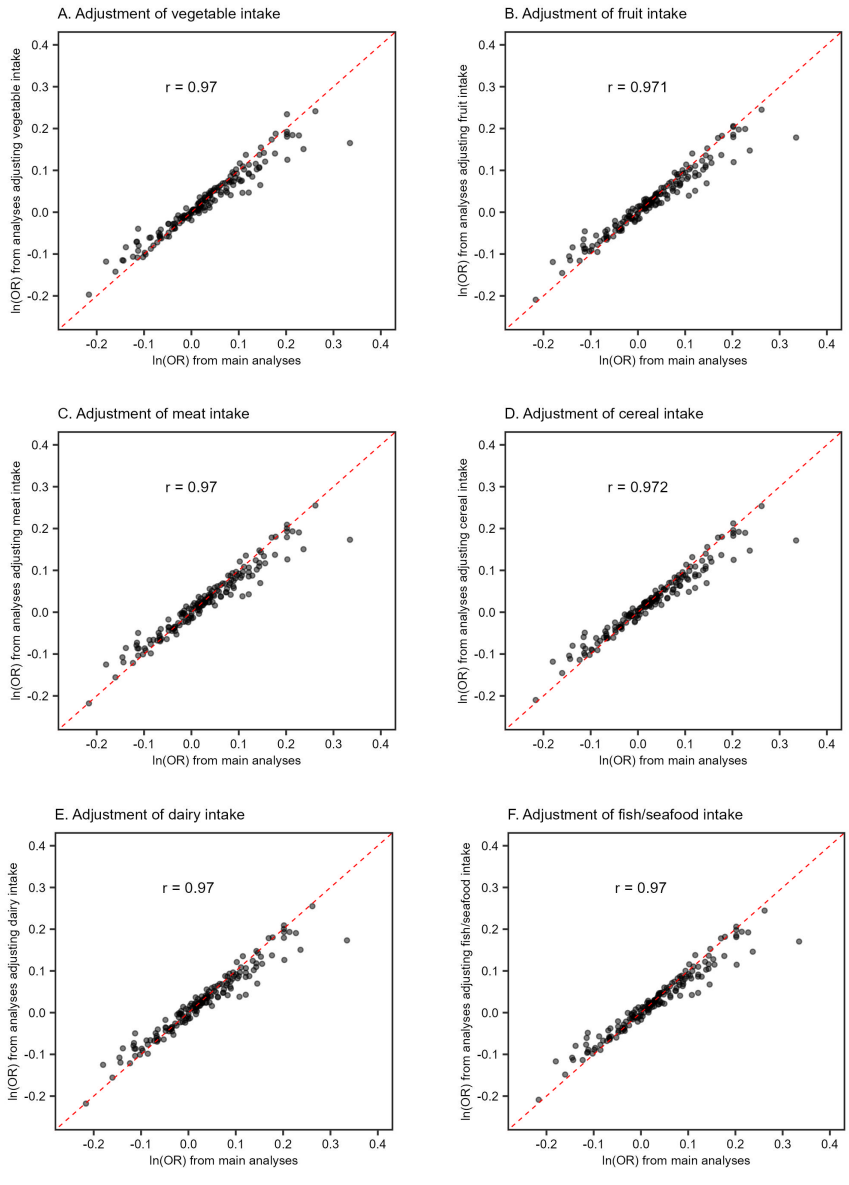
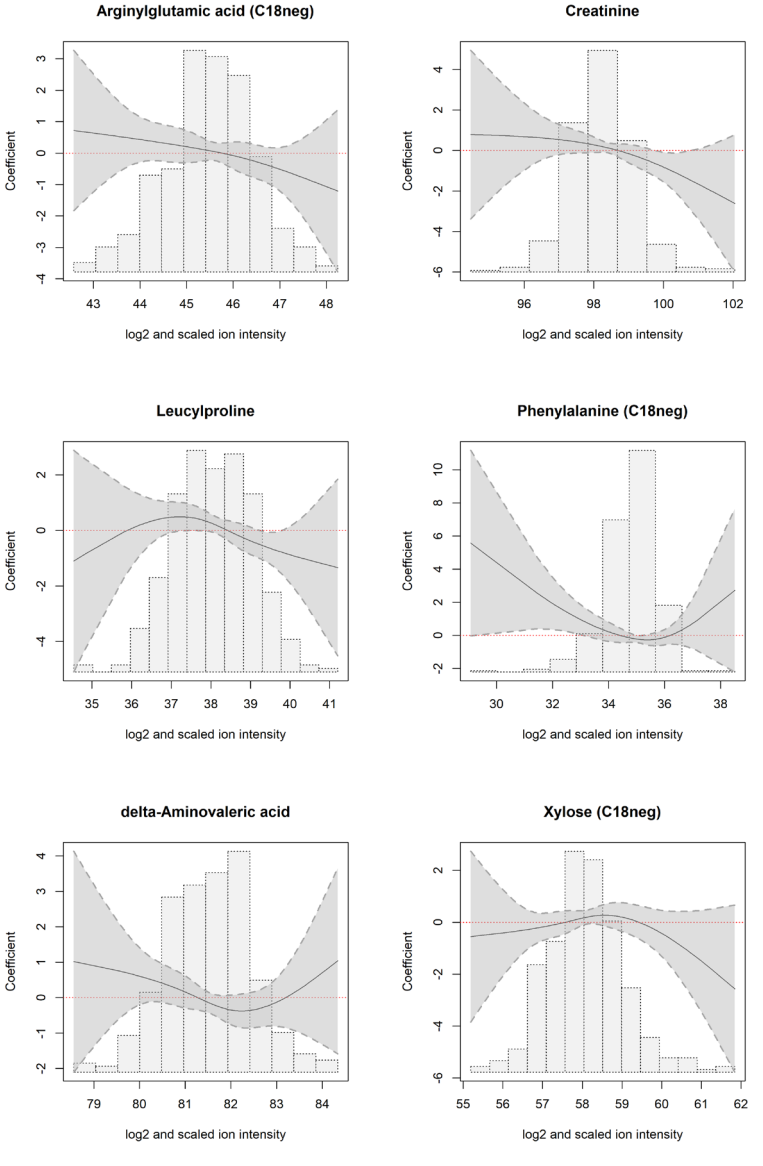


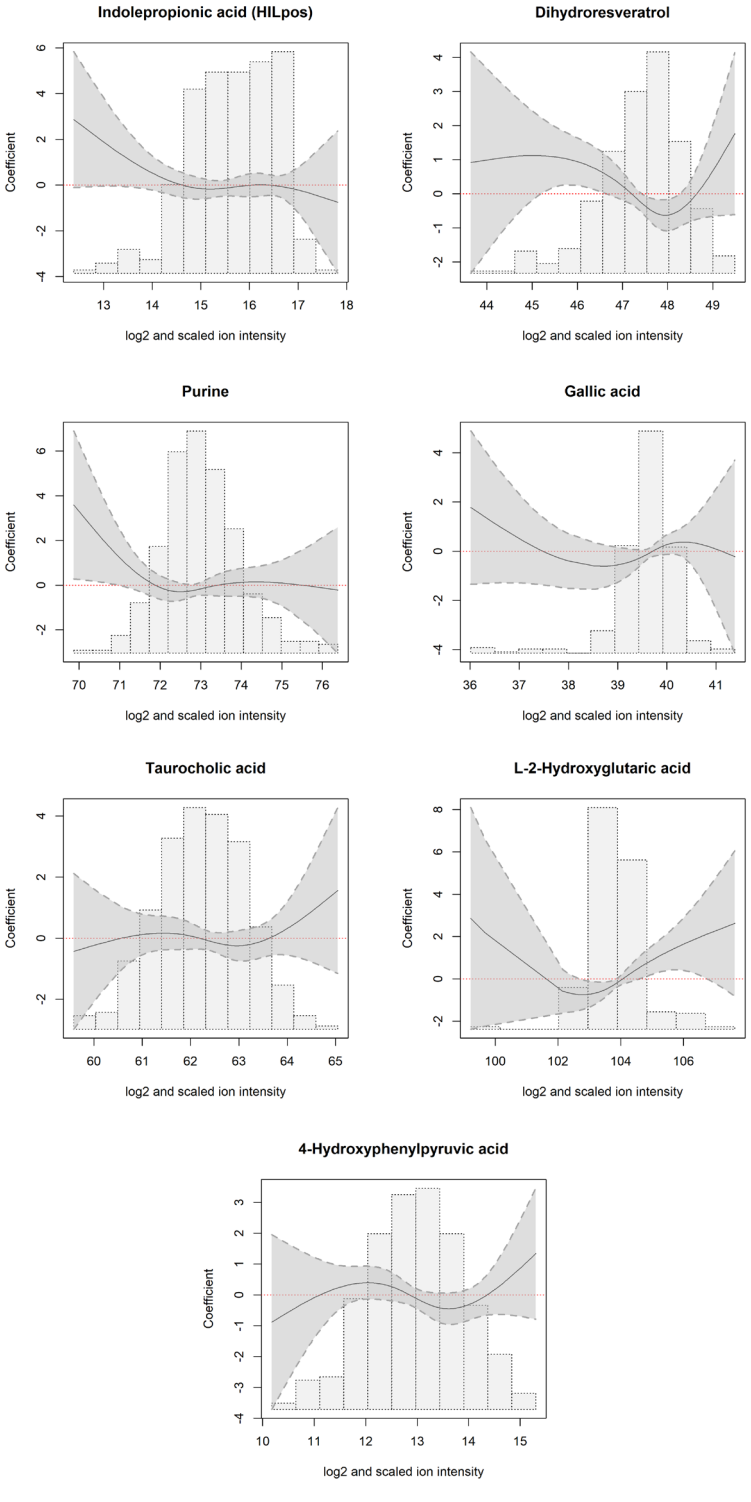
Figure S3. Comparison of metabolome-wide association study results between main analyses with analyses adjusting food intake

In each panel, each dot represented one metabolite (in total 167 microbial metabolites). They were displayed by ln(OR) estimated from the main analyses against the corresponding values from analyses further adjusting for food intake. Six main food groups were further adjusted, and they entered the model one-by-one. Pearson correlation coefficients were labelled on the plot. The dotted line is the line of identity.

Figure S4. Nature spline regression for associations between time-to-diagnosis and metabolites among PD cases

Histogram shows the distribution of ion intensities of metabolites; solid line represents coefficients estimated by analyses where time-to-diagnosis was regressed on metabolite intensities (natural spline with 3 degrees of freedom) among PD cases only. Covariates including age at recruitment, sex, study center, smoking status and BMI, were taken into consideration; gray areas are the 95% confidence intervals.




Indolepropionic acid (HILpos)
Coefficient
log2 and scaled ion intensity
Dihydroresveratrol
Coefficient
log2 and scaled ion intensity
Purine
Coefficient
log2 and scaled ion intensity
Gallic acid
Coefficient
log2 and scaled ion intensity
Taurocholic acid
Coefficient
log2 and scaled ion intensity
L-2-Hydroxyglutaric acid
Coefficient
log2 and scaled ion intensity
4-Hydroxyphenylpyruvic acid
Coefficient
log2 and scaled ion intensity

References

1. Liu KH, Walker DI, Uppal K, et al. High-Resolution Metabolomics Assessment of Military Personnel: Evaluating Analytical Strategies for Chemical Detection. J Occup Environ Med. 2016;58(8 Suppl 1):S53-61.

2. Soltow QA, Strobel FH, Mansfield KG, Wachtman L, Park Y, Jones DP. High-performance metabolic profiling with dual chromatography-Fourier-transform mass spectrometry (DC-FTMS) for study of the exposome. Metabolomics. 2013;9(1 Suppl):S132-s43.

3. Yu T, Park Y, Johnson JM, Jones DP. apLCMS--adaptive processing of high-resolution LC/MS data. Bioinformatics. 2009;25(15):1930-6.

4. Uppal K, Soltow QA, Strobel FH, et al. xMSanalyzer: automated pipeline for improved feature detection and downstream analysis of large-scale, non-targeted metabolomics data. BMC Bioinformatics. 2013;14:15.

5. Neveu V, Nicolas G, Amara A, Salek RM, Scalbert A. The human microbial exposome: expanding the Exposome-Explorer database with gut microbial metabolites. Sci Rep. 2023;13(1):1946.

6. Kind T, Fiehn O. Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. BMC Bioinformatics. 2007;8:105.

7. Tsugawa H, Kind T, Nakabayashi R, et al. Hydrogen Rearrangement Rules: Computational MS/MS Fragmentation and Structure Elucidation Using MS-FINDER Software. Anal Chem. 2016;88(16):7946-58.

8. Bonini P, Kind T, Tsugawa H, Barupal DK, Fiehn O. Retip: Retention Time Prediction for Compound Annotation in Untargeted Metabolomics. Anal Chem. 2020;92(11):7515-22.

9. Schrimpe-Rutledge AC, Codreanu SG, Sherrod SD, McLean JA. Untargeted Metabolomics Strategies-Challenges and Emerging Directions. J Am Soc Mass Spectrom. 2016;27(12):1897-905.

10. Schymanski EL, Jeon J, Gulde R, et al. Identifying small molecules via high resolution mass spectrometry: communicating confidence. Environ Sci Technol. 2014;48(4):2097-8.

11. Unger MM, Spiegel J, Dillmann KU, et al. Short chain fatty acids and gut microbiota differ between patients with Parkinson's disease and age-matched controls. Parkinsonism Relat Disord. 2016;32:66-72.

12. Aho VTE, Houser MC, Pereira PAB, et al. Relationships of gut microbiota, short-chain fatty acids, inflammation, and the gut barrier in Parkinson's disease. Mol Neurodegener. 2021;16(1):6.

13. Chen SJ, Chen CC, Liao HY, et al. Association of Fecal and Plasma Levels of Short-Chain Fatty Acids With Gut Microbiota and Clinical Severity in Patients With Parkinson Disease. Neurology. 2022;98(8):e848-e58.

14. Shin C, Lim Y, Lim H, Ahn TB. Plasma Short-Chain Fatty Acids in Patients With Parkinson's Disease. Mov Disord. 2020;35(6):1021-7.

15. Li P, Killinger BA, Ensink E, et al. Gut Microbiota Dysbiosis Is Associated with Elevated Bile Acids in Parkinson's Disease. Metabolites. 2021;11(1).

16. Cirstea MS, Yu AC, Golz E, et al. Microbiota Composition and Metabolism Are Associated With Gut Function in Parkinson's Disease. Mov Disord. 2020;35(7):1208-17.

17. Shao Y, Li T, Liu Z, et al. Comprehensive metabolic profiling of Parkinson's disease by liquid chromatography-mass spectrometry. Mol Neurodegener. 2021;16(1):4.

Chapter 8

General Discussion

Parkinson's disease (PD) is a complex and multifaceted illness linked to many environmental factors. Despite numerous suspected risk factors, few environmental exposures have been definitively proved as a causal factor, primarily due to the scarcity of prospective studies and the lack of objective exposure markers. This thesis contributes to the existing knowledge of environmental exposures in PD etiology within a prospective context, significantly reducing the risk of reverse causality and other biases such as retrospective recall of lifestyle factors.

This thesis explored the utilization of blood samples collected prior to PD diagnosis for exposure assessment through internal markers. The primary focus was on investigating the roles of metals, caffeine, and the gut-brain axis in PD development. Metals and the protein marker of gut microbiota were assessed by conventional methods (ICP-MS and ELISA, respectively). Furthermore, state-of-the-science high-resolution mass spectrometry (HRMS) was applied to measure caffeine metabolites, as well as microbiota-derived metabolites within an exposomic framework. Both liquid chromatography (LC)- and gas chromatography (GC)-HRMS has been applied to measure small molecules in blood, and this thesis is grounded in the LC-HRMS data.

This work not only provides insights into the understanding of PD etiology but also establishes an analytical pipeline for internal exposure assessment. This pipeline involves integrating untargeted HRMS, large chemical libraries, chemoinformatic and in silico tools for metabolite/pathway identification.

The key messages of this thesis include:

Untargeted HRMS provides the opportunity to efficiently detect a wide range of environmental chemicals in biomonitoring studies and has the potential to uncover previously unrecognized chemicals **(Chapters 4 and 5)**.

Caffeine and its metabolites are shown to have neuroprotective effects in PD, shedding light on the PD's etiology and potential preventive measures **(Chapter 5).**

The involvement of metal exposure in PD development remains unproven in current research. When using internal markers, smoking may confound the association with some of the tobacco-related metals **(Chapters 2 and 3)**.

LPS-mediated inflammation is implicated in the gut-brain interaction of PD. Some microbiome-derived metabolites are suggestively associated with PD risk, and individual characteristics may modify the relationship between the gut microbiome and PD **(Chapters 6 and 7)**.

This chapter begins with a summary of the main findings in this thesis. I then discussed application of HRMS in the characterization of the internal exposome and the role of

environmental factors in PD etiology, as well as further methodological considerations specific for the work presented in this thesis. Finally, recommendations for future research opportunities are provided.

Main findings of this thesis

In **Chapter 2**, we systematically reviewed the existing evidence regarding the associations of metal exposures with PD risk. A total of 83 case-control studies and 5 cohort studies were included in the review. Emphasis was placed on assessing study quality, particularly with regard to study design (methodological quality) and exposure assessment methods. The adapted Newcastle-Ottawa scale was utilized to evaluate study design across four domains, including subject selection, comparability of the groups, ascertainment for exposure (for case-control studies) or outcome (for cohort studies), and statistical analysis. Of the included studies, most (70 case-control and 3 cohort studies) were deemed to have low or moderate overall quality. Notably, a large portion of case-control studies (n=64) assessed metal levels in various biosamples after the onset of PD, which was considered of low quality within our self-developed exposure assessment evaluation framework due to potential reverse causation. Studies on copper, iron, manganese and zinc were most prevalent. Meta-analyses were conducted for these metals. Pooled effects indicated lower levels of copper, iron and zinc in blood among PD cases, while magnesium in cerebrospinal fluid and zinc in hair were higher. However, due to methodological limitations, small sample sizes, limited exposure assessment approaches, and high between-study heterogeneity, the current research on PD epidemiology is insufficient to establish associations between these metals and the risk of the disease. Notably, two large-scale and well-designed case-control studies consistently reported increased PD risk related to bone lead levels (a proxy measure for distant lead exposure), suggesting lead as a possible risk factor for PD.

We aimed to deepen the understanding of metals in PD development with an improved study design within EPIC4PD, a prospective population-based cohort spanning six European countries. **Chapter 3** details a case-control study nested within the EPIC4PD cohort, including 362 incident PD cases and 362 age- and sex-matched controls. Metal concentrations in erythrocytes, collected on average eight years prior to PD diagnosis, were measured for eleven metal species. Most assessed metals were not associated with PD risk, except for cadmium and lead. Cadmium had a suggestive negative association with PD (odds ratio (OR) for the highest quartile 0.70, 95% confidence interval (CI) 0.42-1.17), which, however, diminished among never smokers. Among current smokers only, lead was associated with decreased PD risk (OR 0.06, 95% CI 0.01-0.35). Smoking is a recognized inverse risk factor for PD, and cigarette smoke

contributes to cadmium and lead exposure among smokers. In current smokers, the concentrations of these metals were higher compared to never smokers, with geometric mean levels of 1.22 ng/g for cadmium and 92.0 ng/g for lead, as opposed to 0.48 ng/g for cadmium and 67.2 ng/g for lead, respectively. Furthermore, their levels exhibited a positive correlation with smoking intensity among smokers, with a correlation coefficient of 0.50 for cadmium and 0.26 for lead. In summary, our study did not provide further evidence supporting that metals play a role in the pathogenesis of PD. Internal metal measures should be considered carefully as smoking is a potential strong confounder. Even after correction, residual confounding may exist.

In **Chapters 4 and 5**, we applied cutting-edge untargeted HRMS to comprehensively evaluate internal exposures beyond known targeted compounds. In **Chapter 4**, we leveraged a highly unique occupational population with a history of exposure to dioxin(-like) compounds. We measured 29 currently recognized dioxin(-like) compounds by targeted GC-HRMS in blood samples. We connected these targeted compounds to all chemical features with a chlorinated isotopic pattern. Through this procedure we identified 152 dioxin(-like) co-exposures or metabolized products (collectively termed 'dioxin(-like) related compounds'). This chemical-wide approach significantly broadened the scope of suspected environmental chemicals and associated metabolites. Furthermore, by including dioxin(-like) related compounds, we gained a much richer insight into the associated biological responses measured by untargeted LC-HRMS in the metabolome-wide association study (MWAS). Our findings underscore the potential benefits of integrated chemical-wide and metabolome-wide analyses for assessment of the toxicological effects of environmental exposures.

In **Chapter 4**, dioxin(-like) exposures were associated with extensive changes in pathways of amino acids, lipid and fatty acids, carbohydrates, and nucleotides. These metabolic alterations suggest a disruption of redox balance and bioenergetic synthesis, which have been recognized as central mechanisms of PD pathogenesis (1). This finding suggests a potential role of dioxin(-like) components in contributing to PD development. **Chapter 4**, along with previous work on trichloroethylene (TCE) (2), establishes a paradigm for assessing dioxin(-like) and TCE compounds, as well as other halogenated compounds. This framework can be further applied in PD research to explore potential associations with these environmental chemicals.

In **Chapter 5**, we presented another example utilizing a chemical-wide concept, focusing on metabolites in the caffeine pathway. Coffee, long implicated to have a neuroprotective effect on PD, was associated with a reduced risk of PD within the full EPIC4PD cohort (number of subjects=184,024), with a hazard ratio of 0.63 (95% CI 0.46-0.88) for the highest coffee intake group compared to nonconsumers. To validate the

inverse association with objective markers and to unravel the active component in coffee exerting neuroprotection, we measured plasma caffeine metabolites in the nested case-control study within EPIC4PD that was also mentioned in **Chapter 3**, based on 351 case-control pairs (the number of subjects were lower than for the metal analyses in erythrocytes in **Chapter 3** due to the availability of plasma samples). We annotated 22 caffeine metabolites, along with a recently discovered novel metabolite (1,3,7-trimethyldihydrourate) using the untargeted LC-HRMS data. Caffeine and its primary metabolites, paraxanthine and theophylline, were inversely associated with PD risk. The ORs for one standard deviation (SD) increase of log2 ion intensity were 0.80 (95% CI 0.67-0.95), 0.82 (95% CI 0.69-0.96), and 0.78 (95% CI 0.65-0.93), respectively. Importantly, these inverse associations remained robust as early as 10 years before PD diagnosis. This study confirms the neuroprotective action of caffeine and its primary metabolites and provides valuable insights into the etiology of PD and possible leads for prevention.

The gut-brain axis has emerged as an important etiologic hypothesis for PD (3). However, previous research has mainly focused on alterations of microbiota abundance and composition, with limited connections to changes in biological function. In **Chapters 6 and 7**, we explored gut-brain crosstalk using microbiota-related markers with specific biological relevance and functions within the EPIC4PD nested case-control study. LPS-binding protein (LBP) is the reactant of LPS, a component of Gram-negative bacteria and a potent activator of innate immune response. In **Chapter 6**, LBP was found to be associated with increased risk of PD, with an OR of 1.46 (95% CI 0.98-2.19) per one unit increase of log LBP. This positive association was more pronounced among women (OR 2.68, 95%CI 1.40-5.13) and overweight/obese subjects (OR 1.54, 95%CI 1.09-2.18). Elevated LBP indicates excess LPS in blood circulation, which might result from compromised intestinal barrier function (i.e., leaky gut) and directly lead to inflammation. Our results pinpoint the involvement of LPS-mediated inflammation in the gut-brain interaction in PD.

Using the same untargeted LC-HRMS data as described in **Chapter 5**, **Chapter 7** annotated 167 microbial metabolites in plasma by linking to the recently released Exposome-Explorer dataset and the use of authentic standards. Under the nominal criteria of a 0.05 *p*-value, 13 microbial metabolites were associated with PD risk, including amino acids, bile acid, indoles, and hydroxy acid, though none remained significant after multiple comparison corrections. To move beyond individual metabolites towards global insights into pathways, we performed pathway enrichment analyses based on an MWAS utilizing the whole metabolome from the untargeted LC-HRMS. Three microbiota-relevant pathways were enriched based on 1,000 features

associated with PD risk (p-value<0.05). These pathways included valine, leucine and isoleucine degradation, as well as pathways involved with short-chain fatty acids (butanoate and propanoate metabolism). These pathways have been previously associated with dioxin(-like) exposure, as described in **Chapter 4**. From subgroup analyses, differences in metabolite profiles and pathway enrichment related to sex, smoking status and BMI were observed. More PD-associated pathways were identified for men, smokers, and overweight/obese individuals. These findings suggest that personal traits and lifestyles might contribute to involvement of microbiota in PD. Notably, individuals with higher levels of LBP exhibited more enriched microbial pathways compared to those with lower LBP levels. These results might reflect increased influx of microbial metabolites into the blood due to intestinal leakage, supporting the findings in **Chapter 6**.

Application of metabolomics in characterizing the internal exposome

Exposomics is a rapidly growing field of research, with the goal of identifying all non-genetic factors contributing to disease development. Metabolomics has emerged as a powerful tool for measuring environmental chemicals and endogenous biomolecules in biosamples. This thesis centers on the utilization of metabolomics in assessing environmental exposures within epidemiological studies. The following sections provide considerations regarding the choice between targeted and untargeted analyses, development of metabolite identification in untargeted metabolomics, and comparison and integration of metabolomics data.

Choice between targeted and untargeted analyses

Untargeted metabolomics serves as an unbiased method for assessing all small molecules extracted from a sample, unveiling both known and unknown environmental chemicals. This comprehensive assessment of environmental exposures facilitates an exposome-wide association study (ExWAS). An example is a pilot case-control study on a cholestatic liver disease, primary sclerosing cholangitis (PSC) (4). An ExWAS of PSC was conducted using 10,121 chemical features from untargeted GC-HRMS. The GC-HRMS platform enables detection of volatile and semivolatile organic pollutants, many of which have previously been linked to liver toxicity. The study identified 54 disease-associated metabolites, emphasizing the power of studies using untargeted HRMS.

Besides expanding coverage beyond known chemicals untargeted HRMS also allows detailed interrogation of a specific chemical class to study known and unknown related compounds and metabolites. This chemical-wide approach has been exemplified by the exposure assessment of TCE among occupationally exposed populations (2). Beyond

TCE itself, TCE metabolites including those currently unrecognized were identified based on significant statistical associations with parent compounds and the presence of chlorine atoms. This approach was applied in **Chapter 4** of this thesis to identify of dioxin(-like)-related compounds. Currently, this approach is confined to halogenated compounds in specific populations with prior indication of high exposure to these chemical groups. However, with advancements of metabolite identification capabilities through various in-silico tools, this approach is likely to extend to other chemical classes in the future.

This chemical-wide approach, with its expanded scope of exposure markers, has the potential to enhance the assessment of health effects associated with environmental chemicals. The TCE study showed that biological effects exhibited stronger associations with unknown metabolic products of TCE, as opposed to TCE itself and known metabolites (2). Similarly, in **Chapter 4**, richer insights into exposure-associated biological responses were obtained by introducing dioxin(-like)-related compounds. Even for exposure markers that were previously known but less studied, the exposomic framework could provide valuable insights into health impact assessment, as demonstrated in **Chapter 5** with the neuroprotective effects of caffeine metabolites, paraxanthine and theophylline.

While untargeted metabolomics provides a holistic understanding of environmental exposures in disease etiology, the methodology also has inherent limitations. First, untargeted analysis, aiming to broaden the scope of detection coverage, compromises sensitivity in measuring low-abundant environmental chemicals (5). One solution is to remove high-abundant analytes and concentrate low-abundant compounds, mirroring the historical focus of targeted analysis. Second, chemical confirmation poses a formidable challenge, as the identity of most detected metabolites remains unknown, hindering result interpretation. Third, quantitative risk assessment faces challenges because chemicals are reported in terms of intensity counts rather than concentrations.

In addressing these challenges, Trowbridge et al. illustrated a path forward for combining untargeted screening with targeted methods in pregnant individuals. The work started with a suspect screening for environmental chemicals among over 30,000 untargeted features in maternal and cord blood (6-8). Taking into account prioritized criteria including the ubiquity of exposure, potential implications for pregnancy and previous monitoring, nine chemicals (six exogenous and three endogenous) were selected for confirmation and quantification using targeted approach (9). Their associations with pregnancy complications were then evaluated. This study underscores that targeted analysis constitutes an important part of a metabolomics workflow to validate results from untargeted analysis.

The choice between targeted and untargeted metabolomics should align with the study objectives. Hypothesis-driven studies optimized for specific metabolites may benefit from targeted approaches, while exploratory studies aiming to screen understudied chemical exposures may leverage untargeted methods. Integration of both approaches emerges as a future direction to maximize advantages of both platforms.

Development of metabolite identification in untargeted metabolomics

Ion peaks detected by untargeted metabolomics are referred to as features, each characterized by a unique mass-to-charge ratio (*m/z*) of ions and retention time. The process of determining the true chemical identity of these features is known as annotation. Features are best annotated by matching the exact mass, retention time and tandem mass spectrometry (MS/MS) fragmentation spectra against authentic standards of compounds (10). To facilitate such comparisons, extensive compound databases (e.g., PubChem, HMDB, KEGG, ChemSpider) and MS/MS spectral databases (e.g., METLIN, GNPS, MassBank, NIST) have been developed. However, these libraries are small in comparison to the vast array of natural and synthetic chemicals (11). On average, only 10% of untargeted features can be annotated to date (12), making metabolite identification a primary challenge in the field.

For the application of annotated features, Schymanski's criteria are commonly used to communicate the identification confidence of annotations (13). The criteria range from Level 1 (the highest confidence) to Level 5 (the lowest confidence). Levels 1 and 2, with sufficient structure information, are considered acceptable for application. Levels 3-5 are tentative annotations and are usually deemed as unreliable. In this thesis (**Chapters 4, 5 and 7**) we considered therefore annotation at levels 1 and 2 but ignored annotations at levels 3-5.

In cases where chemicals are not archived in current databases, many computational tools have been developed to predict their physicochemical properties, such as MS/MS spectra and retention times, to aid in feature annotation (14). A recent tool for retention time prediction, *Retip* (15), was utilized in **Chapters 5 and 7** for annotation of caffeine metabolites and microbial metabolites. More than half of the candidate metabolites were annotated with assistance of *Retip*. Additionally, many chemicals may be even more abundant in the human body after enzymatic modification, but to date their metabolized products remain largely unrecognized. Various computational tools are available to predict the structures of molecules undergoing biological transformation. *BioTransformer*, an open access software tool (16), was used to predict chemical products after human metabolism for benzene exposure in an occupational study, enabling the annotation of nine benzene metabolites, four of which were previously unknown (17). Another approach to expand the exposure metabolic

products is searching for isotopic patterns specific to targeted chemicals, as adopted for halogenated chemicals in the TCE study (2) and in **Chapter 4** using the *nontarget* R package (18). In summary, these in silico tools can expand the scope of metabolites annotated from untargeted metabolomics and enhance exposure assessment.

When linking environmental chemicals to endogenous biomolecules, pathway mappings help to understand the roles that biomolecules play in relation to each other and in biological aberrations. These include metabolite set enrichment analysis (MSEA) for annotated metabolites (19) and the Mummichog algorithm for unannotated features (20). These enrichment analyses rely on a complete reference library of the full endogenous metabolome. Creating such a library is laborious and currently impossible due to incomplete knowledge of biomolecules. MSEA and Mummichog now refer to public databases, as well as manually curated metabolites from the literature. A more comprehensive and biologically meaningful library of metabolites is warranted for the interpretation of biological pathways in metabolomics studies.

Comparison and integration of metabolomics data

Reproducibility of metabolomics is always a matter of concern, due to the lack of universally accepted gold standard methods. Diverse workflows in sampling, extraction procedures, and instrument settings can lead to distinct coverage of detectable metabolites, rendering it challenging to align metabolomics data obtained from different platforms. Additionally, post-analytical procedures (including feature and sample filtering, data imputation and normalization), and feature selection approaches can induce variations in the selected features associated with disease phenotypes. This further complicates result comparison and evidence synthesis for biological interpretation across studies. To address this, some studies have showed that adopting standardized protocols and instruments, along with using reference materials for data normalization, can enhance inter-laboratory reproducibility (21-23).

In addition to standardization of the analytical procedure, epidemiologists find statistical approaches for pooling metabolomics data particularly intriguing, as the one developed by Viallon et al. (24). This method does not require quality controls and reference assay data for metabolomics normalization and accommodates various biosamples. It relied on identifying major sources of variation in metabolomics data through principal component partial R-square (PC-PR2) analysis. Subsequently, the approach corrects for unwanted variability using mixed models while preserving biological variability. This pipeline has already been applied to targeted metabolomics data obtained from LC-HRMS platforms in eight case-control studies nested in the EPIC cohort. However, its application into other scenarios, especially in untargeted metabolomics, requires further investigation.

Role of environmental factors in Parkinson's disease etiology

The diagnosis of PD relies on the observation of motor symptoms. Unfortunately, by the time these symptoms become noticeable, about 50% of substantia nigra dopamine neurons are lost (25). Furthermore, there is currently no cure or established strategy for slowing the progression of the disease. Therefore, it is crucial to understand the underlying causes of PD and stride toward disease prevention and modification to alleviate the global burden of PD. Various environmental factors, which have been implicated in affecting the risk of PD, hold promise for disease modification, but the available evidence has been inconsistent. This section explores the complex nature of environmental factors and the individual variability in the gut-brain axis involvement in PD, before discussing the therapeutic potential of environmental factors.

Complex nature of environmental factors in PD etiology

Conclusive establishment of causality of most environmental factors remains elusive. One major challenge arises from the prolonged prodromal period of PD, spanning up to 20 years. Throughout this extended stage, various environmental factors may have come into play, influencing the risk and progression of PD (**Box 1** in the **General Introduction**). Some exposures are inherently interconnected or involved in similar biological mechanisms, leading to confounding and effect modification.

Smoking consistently emerges as an inverse risk factor for PD (26). This protective effect has been attributed to nicotine, which influences dopaminergic activity by acting at nicotinic receptors in the brain and could protect against nigrostriatal damage (27). Cigarette smoke is a complex mixture and also serves as an important source of certain metals, such as lead (28), as discussed in **Chapter 3** of this thesis. Lead levels in current smokers were around 1.4 times higher compared to never smokers. An inverse and robust association between lead levels and PD risk was observed among current smokers. However, this effect diminished among never smokers. Lead is a well-recognized neuro-toxicant and exhibited a positive association with PD risk when measured as a cumulative biomarker (bone lead) from previous studies (29, 30). The inverse and smoking-specific association for lead as reported in **Chapter 3** is likely due to the confounding effect of smoking, suggesting that lead might act as a surrogate exposure measure for smoking.

Further illustrating the complexity of environmental interactions is the interplay among caffeine, smoking and estrogen. Both caffeine and estrogen demonstrate neuroprotective effects and share competitive metabolism by a cytochrome P450 enzyme, CYP1A2 (31). Smoking significantly induces the CYP1A2 enzyme, accelerating the metabolism of caffeine and estrogen in smokers (32). Effect modifications among

these three protective factors for PD have been observed in prior research and in **Chapter 5** of this thesis. Smoking was found to enhance the inverse association between caffeine and PD in some cohort studies (33-35) and in **Chapter 5**. Conversely, hormone usage may diminish the protective effects of caffeine on PD, a finding also corroborated in **Chapter 5**.

In summary, caution is essential in epidemiological studies especially considering the intricate interactions when investigating the etiological role of environmental factors in PD. As human observational research is limited in fully elucidating these interactions and their underlying mechanisms, experimental studies are warranted for future in-depth exploration.

Individual variability in the gut-brain axis involvement in PD

Aggregation of α-synuclein in the form of Lewy bodies and Lewy neurites and neuronal loss in the substantia nigra are the pathological hallmarks of PD. Beginning with the discovery of α-synuclein deposits in the enteric nervous system (ENS) and Braak's hypothesis of gut-to-brain spread of PD synucleinopathy, it has been hypothesized that PD pathology might originate from intestinal dysfunction (36). Key events of the gut-brain hypothesis include gut microbial dysbiosis, gut inflammation and hyperpermeability, and propagation of α-synuclein in the ENS (37). Over the past two decades, both supporting and contradicting evidence for this gut-first hypothesis for all PD cases has emerged from human and experimental research (**Box 1**). These inconsistencies have promoted researchers to propose distinct 'body-first' and 'brain-first' subtypes of PD (38). These subtypes are clinically distinguished by the presence or absence of REM sleep behavior disorder (RBD) during the PD prodromal phase (39). It implies that misfolded α-synuclein may originate in the ENS for the body-first group, whereas it originates in the brain for the brain-first group. However, in the EPIC4PD cohort, the distinction of PD subtypes based on the presence of RBD cannot be readily determined due to the absence of documentation regarding pre-clinical symptoms.

While existence of these two phenotypes has been confirmed, underlying determinants remain largely unexplored. This thesis provides some insights on the variability of gut involvement in PD. In **Chapter 6**, elevated LBP levels were associated with an increased risk of PD, especially among women and obese/overweight individuals. In **Chapter 7**, perturbation profiles of microbial metabolites were heterogenous across subgroups of sex, smoking status and BMI. Enriched microbial-related pathways were more pronounced for men, smokers and obese/overweight individuals. These findings suggest that personal traits and lifestyles may contribute to the differentiation of PD subtypes. Furthermore, they may have varying effects on different microbial-related markers, as evidenced by the differential influence of sex on LBP and microbial

metabolites in **Chapters 6 and 7**. Further research is warranted to explore the mechanism underlying diverse disease trajectories in PD.

Moreover, individuals may fall along a continuum between the body-first and brain-first dichotomy, potentially exhibiting both phenotypes simultaneously (40). A deeper exploration of the heterogeneity of gut involvement in PD patients could pave the way for personalized gut-directed therapies.

Box 1. Evidence in favor of and against the gut-first hypothesis in PD cases (adapted from Tan et al. 2022 (3))

Evidence in favor	Evidence against
Epidemiological studies	
Gastrointestinal symptoms such as constipation can precede PD motor symptoms by several decades	Preceding of gastrointestinal symptoms because the ENS is less able to compensate to neurodegeneration than CNS
Full truncal vagotomy is associated with a reduced risk of PD	Neuroprotective effect of vagotomy is not consistent
Increased intestinal inflammation and permeability found in PD patients.	Results from enteric inflammation and hyperpermeability were heterogenous.
Neuropathological studies	
Synucleinopathy in the ENS in almost all PD patients	Gastrointestinal α-synuclein accumulation in neurologically healthy individuals
Gastrointestinal α-synuclein deposition up to 20 years before PD diagnosis	No notable neuronal loss was detected in the ENS in PD patients
Animal studies	
Gut-to-brain α-synuclein transmission, accompanied by dopaminergic neurodegeneration and PD-like symptoms	Animal models involving the injection of α-synuclein into the gastrointestinal tract might not accurately recapitulate human PD
Hemivagotomy prevented both α-synuclein accumulation in vagus nerve and brain	Brain-to-gut α-synuclein transfer has been demonstrated
α-synuclein pathology and PD-related dysregulation in the ENS before striatal dopaminergic abnormalities	Bidirectional gut-to-brain and brain-to-gut α-synuclein propagation was independent of vagus nerve
Exposure to microbial amyloid proteins in the gastrointestinal tract increased α-synuclein pathology and inflammation in the gut and brain, with corresponding motor deficits.	CNS disorders can also alter gut function; nigral pathology in rats resulted in enteric dysmotility, inflammation and microbiome alterations

ENS, enteric nervous system; CNS, central nervous system.

Therapeutic potential of environmental factors for PD

The challenge of drug development for PD is partly attributed to the uncertainty surrounding the disease etiology. While accumulating evidence points towards environmental factors that either increase or decrease the risk of PD, it is still unclear how these environmental agents mediate subsequent pathological changes. Another obstacle is the lack of early biomarkers for PD diagnosis, with motor signs typically emerging only after about 50% of substantia nigra dopamine neurons are lost (25). Early detection is crucial for administering therapeutic treatments before the onset of noticeable symptoms.

Observations of consistent inverse associations between smoking and coffee consumption and the risk of PD have prompted investigations into their therapeutic potency. Nicotine has been proposed as the active component responsible for the neuroprotective effect. Through the activation of nicotinic acetylcholine receptors, nicotine could modulate dopamine transmission and reduce levodopa-induced dyskinesias, as shown in animal models (41, 42). **Chapter 5** in this thesis has confirmed the inverse association between pre-diagnostic caffeine intake and PD. Caffeine, an adenosine 2A receptor antagonist, could attenuate neuronal death, dopamine depletion and motor impairment, as shown in parkinsonian rodents (43-45). However, clinical trials evaluating the efficacy of nicotine and caffeine in PD patients have yielded inconclusive results (**Box 2**). Recent double-blinded, placebo-controlled trials with sizable PD patient cohorts and extended follow-up periods failed to demonstrate the effectiveness of nicotine and caffeine in addressing motor deficits (46, 47). One can speculate that the lack of efficacy is due to the intervention being applied after the clinical PD diagnosis (with an average disease duration 4-11 years across trials). At this stage, pathological changes and movement impairment are largely irreversible.

Additionally, novel therapeutic approaches targeting the gut microbiota have been proposed, given the potential involvement of gut-related factors in PD. This connection has been suggested by the finding of increased chronic inflammation and altered microbial metabolism, as shown in **Chapters 6 and 7** in this thesis. Current therapies include diet intervention to increase prebiotics (substrates utilized by microbiome, such as high-fiber foods), microbiota modulation by faecal microbiota transplantation and antimicrobials, as well as the use of postbiotics and small-molecule drugs (e.g., short-chain fatty acids, tight junction modulators) (3). Although these approaches have shown potential benefits for PD patients with regards to motor symptoms and bowel functions like constipation in small pilot studies, double-blind controlled trials in PD patient populations are limited (48, 49).

In summary, the therapeutic potential of nicotine, caffeine and microbiota-targeted treatments merits validation through well-designed randomized controlled trials, ideally initiated early in the disease course. This can be achieved by administering candidate agents to individuals with rapid eye movement (REM) sleep behavior disorder (RBD), which is by far the strongest indicator of prodromal PD. The interval between development/diagnosis of RBD and defined PD averages 10-15 years (50). Determining specific treatment regimens, especially regarding the timing of treatment, and adopting personalized modalities based on the pathways driving PD are essential steps to optimize the success of disease modification.

Box 2. Double-blind controlled trials of clinical efficacy of nicotine and caffeine on PD

n. of patients (ref)	Treatment, dose	Duration of treatment	Primary endpoint	Efficacy*
Nicotine				
48 (51)	Nicotine gum, 2mg/gum	3 times at 2h intervals	UPDRS Part III	0
32 (52)	Transdermal nicotine, up to 35mg daily	3 weeks	CURS part D and E score	0
77 (53)	Oral selective nAChR agonist, up to 40mg twice daily	4 weeks	UPDRS Part I to III, cognitive test	0
65 (54)	Oral nicotine, up to 24mg daily	10 weeks	UPDRS Part II and III, UDysRS	+
163 (47)	Transdermal nicotine, up to 28mg daily	52 weeks	Total UPDRS score	0
Caffeine				
61 (55)	Oral caffeine, up to 200mg twice daily	6 weeks	UPDRS Part III	+
121 (46)	Oral caffeine, up to 200mg twice daily	≥ 6 months	UPDRS Part III	0

UPDRS, unified Parkinson's disease rating scale; CURS, Columbia University rating scale; nAChR, nicotinic acetylcholine receptor; UDysRS, unified dyskinesia rating scale

* +, efficacy found; 0, no efficacy found.

Methodological considerations of this thesis

PD research faces unique challenges related to reverse causality as compared to other chronic diseases like cancer and diabetes. Over the course of decades in the prodromal phase of PD, various pre-clinical symptoms manifest, including sleep disorders, constipation, and lost sense of olfactory and taste. These symptoms may prompt future PD patients to modify their diet habits and lifestyles, thereby altering relevant exposures. Unlike PD, other chronic diseases typically have a less evident prodromal phase, making pre-clinical symptoms less apparent and may therefore not affect diet habits and lifestyles prior to disease as much.

The main strength of this thesis lies in reducing reverse causality by utilizing the prospective EPIC4PD cohort and its biobank established at baseline. The assessments of internal exposure markers through biomonitoring occurred in samples collected prior to PD diagnosis. However, certain methodological aspects within this thesis warrant discussion, particularly the biomonitoring based on a single moment in time, the captured exposure window for environmental factors, and tracking microbiome dysbiosis with blood biomarkers.

One-time biomonitoring of exposures

In this thesis, we adopted biomonitoring to evaluate exposures to metals, dioxin(-like) compounds, caffeine metabolites and microbial markers. However, it is essential to recognize that exposure levels measured in blood at a single time point only offer a snapshot of an individual's exposure history, only reflecting exposures proximal to the time of blood sample collection. The alignment of one-time biomonitoring results with long-term exposure status depends on the chemical properties of the assessed exposures and the trajectory of exposure states throughout an individual's lifetime.

For environmental chemicals with decades-long biological half-lives, such as dioxin(-like) compounds and other persistent organic pollutants, blood concentrations can effectively mirror exposures spanning the preceding 20-25 years (56). Moreover, due to their exclusive anthropogenic sources, these chemicals are less likely to be affected by natural or physiological processes. Conversely, in the case of nutrients and metals, with half-lives varying from a few hours to a few months in the blood, the reliability of biomonitoring in representing long-term exposure hinges on the stability of exposure status. One-time biomonitoring could be inaccurate when exposures fluctuate over a lifetime due to changes in lifestyles and environmental sources.

Cigarette smoke is a major source of cadmium intake among smokers. In **Chapter 3**, we observed that cadmium concentrations in former smokers resembled those in never smokers, but that levels in current smokers were higher. This demonstrates that a single

measurement may not adequately reflect past exposure status. Another example is lead; in the general population, blood lead levels have declined by over 90% since the 1980s-1990s due to reduced lead emissions from petrol (57). As blood samples in the EPIC4PD nested case-control study were collected during the 1990s, with a median age of subjects at 60, the assessed lead levels may not have accurately represented high exposures during subjects' early life. Notably, exposure misclassification resulting from a one-time measurement occurs equally in both the case and control groups, leading to an underestimate of the true strength of the association between exposure and disease. Consequently, while false negatives may occur, the false positive rate is typically low. Despite uncertainties regarding extrapolation to chronic and past exposures, biomonitoring remains a promising approach to quantify exposure gradients, enabling characterization of exposure-response analyses.

Relevant exposure windows for environmental factors

Most epidemiologic studies on PD thus far have focused on exploring potential environmental risk factors that are present during mid or late-life. Despite PD being recognized as a neurodegenerative disorder typically presenting after the age of 60, the possibility that PD could be driven by lifelong environmental factors has been proposed (58). Smoking and coffee consumption have consistently been identified to have protective associations, and these behaviors generally start in early adulthood. **Chapter 5** in this thesis has revealed that inverse associations with caffeine metabolites are evident as early as 12 years before PD diagnosis. A prior report from the EPIC4PD cohort indicated that individuals who started to smoke before age 16 had a stronger reduction in PD risk (59). Furthermore, some studies have highlighted potential roles of perinatal and neonatal factors, including birthweight, birth season and place, multiple birth, suggesting that early-life exposures might also be important contributors for late-onset PD (60, 61).

It is plausible that environmental factors have a long latency or a slow effect on PD development. Therefore, a life course approach, examining the causal relationship between exposures during gestation, childhood, adolescence and young adulthood, and subsequent PD, has been emphasized. In the EPIC4PD nested case-control study, blood samples were gathered on average 8 years prior to disease diagnosis of PD cases, around their age of 60. The absence of associations with blood exposure markers, as demonstrated in **Chapter 3** for metals, might be attributed to a misalignment between exposure assessment and the relevant etiological time window, and this cannot rule out their etiological roles during earlier stages of life. The characterization of early exposures necessitates extended follow-up periods and repeated exposure measurements for existing cohorts. Rather than focusing solely on PD occurrence,

utilizing early disease indicators as observational outcomes, such as the presence of RBD, which typically precedes PD onset by 10-15 years, could help shorten prolonged follow-up periods.

Tracking gut microbiome dysbiosis with blood biomarkers

Direct measurements of gut microbial composition and diversity are usually infeasible in large cohort studies where no faecal samples have been collected. In instances of microbiome dysbiosis, the permeabilization of the blood-gut barrier allows microbiota-derived products to cross into systemic circulation. Therefore, blood-based measures including markers of host immune response (e.g., LPS-binding protein (LBP in **Chapter 6**), soluble CD14, endotoxin core antibody) and microbial metabolic products (e.g., microbial metabolites in **Chapter 7**) offer practical alternatives for gut microbiome functions.

When utilizing blood biomarkers, it is crucial to consider the role of the intestinal barrier permeability. Short chain fatty acids (SCFAs), such as propionate, butyrate and acetate, are produced from the metabolism of digested fibers by bacteria. Low faecal levels of butyrate and acetate have been associated with worsened PD symptoms (62, 63). Conversely, plasma levels of acetate and propionate levels were found to be higher in PD patients than in control individuals (64, 65). Some researchers postulated that this discrepancy could be attributed to the leakage of gut SCFAs into systemic circulation due to intestinal barrier disruption and reduced clearance of these SCFAs by the colonic epithelium (64). On the other hand, certain microbial metabolites, such as amino acids, can also be produced by human tissues or directly derived from the diet (66). Thus, metabolites primarily or exclusively synthesized by the microbiota provide a better insight on microbial function in the complex pathogenesis of PD. The classification for metabolites based on their origin (in this thesis in **Chapter 7**) has been adopted by the Exposome-Explorer database (66). This database categorized evidence from the literature into *in vitro* and *in vivo* types regarding microbial origin. Metabolites substantiated by both types of evidence are deemed as largely or exclusively produced by the gut microbiota.

With the emergence of advanced analytical techniques, microbial DNA sequencing in blood, which is historically considered sterile, holds promise for studying gut microbial translocation and would extend the current work on gut metabolites. Exploratory work has highlighted the heterogenous richness of blood bacterial genera between PD patients and healthy controls, with specific genera showing associations with clinical characteristics of PD (67).

Future directions

In this thesis, HRMS-based metabolomics provides an avenue to comprehensively characterize diverse environmental exposures in a high-throughput manner. This approach has been applied for the measurement of dioxin(-like) compounds (**Chapter 4**), caffeine metabolites (**Chapter 5**), and microbial metabolites (**Chapter 7**). The detailed characterization of these exposures largely facilitates subsequent research on health impacts. The next step in exposure assessment by metabolomics involves broadening the coverage of detectable or annotated chemicals. This expansion is crucial for enhancing exposome-wide association studies, contributing to a deeper exploration of PD etiology. With integration of metabolomics and other omics, a holistic view on PD can be obtained.

Exposome-wide association study for PD

As outlined in the **General Introduction**, various environmental chemicals have been proposed to be linked with an altered risk of PD. Among these, pesticides stand out as one of consistently reported risk factors in the literature (68). However, limitations in previous exposure assessment methods hindered the accurate estimation of pesticide exposures. Most studies relied on self-reported contact with pesticides, treating them either as a single entity or categorizing them into sub-classes such as insecticides, herbicides or fungicides. However, these sub-classes comprise numerous active ingredients with different chemical properties and action mechanisms, complicating the identification of specific pesticides as potential etiological factors for PD. A recent study overcame this limitation by estimating ambient exposure to 288 specific pesticides in California using a geographic information systems-based model (69). The study further identified ten pesticides directly toxic to human dopaminergic neurons. In addition to external exposure assessment, blood metabolomics provides a unique opportunity for the comprehensive characterization of internal pesticide levels. The LC-HRMS data (used in this thesis) and GC-HRMS data (currently under analysis) from the EPIC4PD project would allow for the identification of hundreds of specific pesticides, facilitating a pesticide-wide association study for the risk of PD.

Beyond pesticides, metabolomics, particularly with GC-HRMS, enables the detection of numerous other environmental chemicals. This is particularly relevant as most environmental chemicals are hydrophobic, semi-volatile and poorly ionize with LC-HRMS methods. Within our collaborative network, an analytical pipeline utilizing untargeted GC-HRMS has been established to operationalize the human exposome (70). This approach facilitates a comprehensive screening of various chemical groups, including polychlorinated biphenyls (PCBs), organic acids, per-and polyfluoroalkyl substances (PFASs), and dioxins, some of which are suspected risk factors for PD. Such

broad screening empowers an exposome-wide association analysis for PD in the near future. Furthermore, utilizing the methodology outlined in **Chapter 4** on dioxin(-like) compounds and the previous work on TCE, a comprehensive chemical-wide analysis, including both parent compounds and their metabolites, can be applied to thoroughly explore the potential effects of particular chemical groups.

Multi-omics data integration

Although this thesis exclusively discusses environmental factors, the prevailing view acknowledges that most PD cases likely arise from a combination of environment exposures and genetic susceptibility. Growing evidence suggests that epigenetic changes are key mechanism at the interface of the gene-environment interaction. Many *in vitro* and *in vivo* studies have shown that PD-relevant toxins, 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP), as well as pesticides rotenone and paraquat, can induce histone acetylation and DNA methylation -- two common types of epigenetic changes (71, 72). A recent systematic review identified ten human studies investigating epigenetic regulation in PD acting upon pesticides, metals, smoking and coffee consumption (73). The research on this topic is just emerging, and definitive conclusions regarding any of the aforementioned exposure categories remain elusive. Notably, anti-parkinsonian therapy has been found to influence DNA methylation (74), potentially introducing confounding factors into associations derived from prevalent PD patients. The EPIC4PD cohort, with its pre-diagnostic blood samples and relatively large sample size, proves particularly valuable for conducting epigenome-wide association studies. Furthermore, the integration with environmental exposures assessed by metabolome analysis can be utilized to explore the mechanism underlying gene-environment interactions in PD.

Beyond the realms of epigenome and metabolome, high-throughput transcriptome and proteome analyses have begun to be applied to in human PD research (75, 76). These omics platforms collectively offer a holistic perspective to understanding the disease mechanisms of PD across various molecular layers. With the ongoing development of data integration algorithms, the crossover analysis of multi-omics holds tremendous promise for uncovering novel disease markers and pathogenic pathways.

Concluding remarks

PD stands out as a complex neurodegenerative disorder featuring a wide spectrum of motor and non-motor symptoms, intertwined with multifaceted environmental causes. The exploration of PD etiology in human observational research encounters challenges, particularly the issue of reverse causality due to the extended pre-disease periods. This thesis provides valuable insights into the impact of environmental exposures on PD

development, leveraging a relatively large sample size and pre-diagnostic biosamples from the EPIC4PD cohort. The findings of this thesis, in line with previous studies, consistently highlight a protective effect of coffee consumption on PD risk. This effect is predominantly attributed to caffeine and its primary metabolites. Additionally, chronic systemic inflammation induced by the bacterial product, LPS, is implicated prior to PD diagnosis, and the metabolic profiles of microbial metabolites show variations influenced by individual characteristics. These results underscore the involvement of the gut in the pathogenesis of PD. However, the conclusive roles of metal exposures remain elusive, with no definitive reports from either the existing literature or analyses within the EPIC4PD cohort.

This thesis particularly emphasizes the characterization of internal exposures through untargeted metabolomics. Leveraging this advanced analytic platform enables the detection of a broad spectrum of exposure biomarkers, encompassing both known and currently unidentified chemicals. The integration of metabolomics into epidemiological research facilitates the identification of body burden of environmental agents and their relevance with biological outcomes. Looking forward, the application of multi-omics holds a major promise with the EPIC4PD cohort being one of the largest bio- and omics-data repositories for PD worldwide, offering entirely new possibilities for biomarker discovery work.

References

1. Anandhan A, Jacome MS, Lei S, et al. Metabolic Dysfunction in Parkinson's Disease: Bioenergetics, Redox Homeostasis and Central Carbon Metabolism. Brain Res Bull. 2017;133:12-30.

2. Walker DI, Uppal K, Zhang L, et al. High-resolution metabolomics of occupational exposure to trichloroethylene. Int J Epidemiol. 2016;45(5):1517-27.

3. Tan AH, Lim SY, Lang AE. The microbiome-gut-brain axis in Parkinson disease - from basic research to the clinic. Nat Rev Neurol. 2022;18(8):476-95.

4. Walker DI, Juran BD, Cheung AC, et al. High-Resolution Exposomics and Metabolomics Reveals Specific Associations in Cholestatic Liver Diseases. Hepatol Commun. 2021.

5. Clayton S. Bloszies OF. Using untargeted metabolomics for detecting exposome compounds. Current Opinion in Toxicology. 2018.

6. Jiang T, Wang M, Wang A, et al. Large-Scale Implementation and Flaw Investigation of Human Serum Suspect Screening Analysis for Industrial Chemicals. J Am Soc Mass Spectrom. 2021;32(9):2425-35.

7. Panagopoulos Abrahamsson D, Wang A, Jiang T, et al. A Comprehensive Non-targeted Analysis Study of the Prenatal Exposome. Environ Sci Technol. 2021;55(15):10542-57.

8. Wang A, Abrahamsson DP, Jiang T, et al. Suspect Screening, Prioritization, and Confirmation of Environmental Chemicals in Maternal-Newborn Pairs from San Francisco. Environ Sci Technol. 2021;55(8):5037-49.

9. Trowbridge J, Abrahamsson D, Bland GD, et al. Extending Nontargeted Discovery of Environmental Chemical Exposures during Pregnancy and Their Association with Pregnancy Complications-A Cross-Sectional Study. Environ Health Perspect. 2023;131(7):77003.

10. Alseekh S, Aharoni A, Brotman Y, et al. Mass spectrometry-based metabolomics: a guide for annotation, quantification and best reporting practices. Nature Methods. 2021;18(7):747-56.

11. Cui Y, Balshaw DM, Kwok RK, Thompson CL, Collman GW, Birnbaum LS. The Exposome: Embracing the Complexity for Discovery in Environmental Health. Environ Health Perspect. 2016;124(8):A137-40.

12. de Jonge NF, Mildau K, Meijer D, et al. Good practices and recommendations for using and benchmarking computational metabolomics metabolite annotation tools. Metabolomics. 2022;18(12):103.

13. Schymanski EL, Jeon J, Gulde R, et al. Identifying small molecules via high resolution mass spectrometry: communicating confidence. Environ Sci Technol. 2014;48(4):2097-8.

14. Blazenovic I, Kind T, Ji J, Fiehn O. Software Tools and Approaches for Compound Identification of LC-MS/MS Data in Metabolomics. Metabolites. 2018;8(2).

15. Bonini P, Kind T, Tsugawa H, Barupal DK, Fiehn O. Retip: Retention Time Prediction for Compound Annotation in Untargeted Metabolomics. Anal Chem. 2020;92(11):7515-22.

16. Djoumbou-Feunang Y, Fiamoncini J, Gil-de-la-Fuente A, Greiner R, Manach C, Wishart DS. BioTransformer: a comprehensive computational tool for small molecule metabolism prediction and metabolite identification. J Cheminform. 2019;11(1):2.

17. Rothman N, Vermeulen R, Zhang L, et al. Metabolome-wide association study of occupational exposure to benzene. Carcinogenesis. 2021;42(11):1326-36.

18. Loos M. Nontarget: detecting isotope, adduct and homologue relations in LC-MS. R package version. 2016.

19. Xia J, Wishart DS. MSEA: a web-based tool to identify biologically meaningful patterns in quantitative metabolomic data. Nucleic Acids Res. 2010;38(Web Server issue):W71-7.

20. Li S, Park Y, Duraisingham S, et al. Predicting network activity from high throughput metabolomics. PLoS computational biology. 2013;9(7):e1003123.

21. Allwood JW, Erban A, de Koning S, et al. Inter-laboratory reproducibility of fast gas chromatography-electron impact-time of flight mass spectrometry (GC-EI-TOF/MS) based plant metabolomics. Metabolomics. 2009;5(4):479-96.

22. Benton HP, Want E, Keun HC, et al. Intra- and interlaboratory reproducibility of ultra performance liquid chromatography-time-of-flight mass spectrometry for urinary metabolic profiling. Anal Chem. 2012;84(5):2424-32.

23. Izumi Y, Matsuda F, Hirayama A, et al. Inter-Laboratory Comparison of Metabolite Measurements for Metabolomics Data Integration. Metabolites. 2019;9(11).

24. Viallon V, His M, Rinaldi S, et al. A New Pipeline for the Normalization and Pooling of Metabolomics Data. Metabolites. 2021;11(9).

25. Ross GW, Petrovitch H, Abbott RD, et al. Parkinsonian signs and substantia nigra neuron density in decendents elders without PD. Ann Neurol. 2004;56(4):532-9.

26. Li X, Li W, Liu G, Shen X, Tang Y. Association between cigarette smoking and Parkinson's disease: A meta-analysis. Arch Gerontol Geriatr. 2015;61(3):510-6.

27. Quik M, Perez XA, Bordia T. Nicotine as a potential neuroprotective agent for Parkinson's disease. Mov Disord. 2012;27(8):947-57.

28. Talhout R, Schulz T, Florek E, van Benthem J, Wester P, Opperhuizen A. Hazardous compounds in tobacco smoke. Int J Environ Res Public Health. 2011;8(2):613-28.

29. Coon S, Stark A, Peterson E, et al. Whole-body lifetime occupational lead exposure and risk of Parkinson's disease. Environ Health Perspect. 2006;114(12):1872-6.

30. Weisskopf MG, Weuve J, Nie H, et al. Association of cumulative lead exposure with Parkinson's disease. Environ Health Perspect. 2010;118(11):1609-13.

31. Kolahdouzan M, Hamadeh MJ. The neuroprotective effects of caffeine in neurodegenerative diseases. CNS Neurosci Ther. 2017;23(4):272-90.

32. Benowitz NL, Peng M, Jacob P, 3rd. Effects of cigarette smoking and carbon monoxide on chlorzoxazone and caffeine metabolism. Clin Pharmacol Ther. 2003;74(5):468-74.

33. Ross GW, Abbott RD, Petrovitch H, et al. Association of coffee and caffeine intake with the risk of Parkinson disease. Jama. 2000;283(20):2674-9.

34. Ascherio A, Zhang SM, Hernán MA, et al. Prospective study of caffeine consumption and risk of Parkinson's disease in men and women. Ann Neurol. 2001;50(1):56-63.

35. Liu R, Guo X, Park Y, et al. Caffeine intake, smoking, and risk of Parkinson disease in men and women. Am J Epidemiol. 2012;175(11):1200-7.

36. Braak H, Del Tredici K, Rüb U, de Vos RA, Jansen Steur EN, Braak E. Staging of brain pathology related to sporadic Parkinson's disease. Neurobiol Aging. 2003;24(2):197-211.

37. Klann EM, Dissanayake U, Gurrala A, et al. The Gut-Brain Axis and Its Relation to Parkinson's Disease: A Review. Front Aging Neurosci. 2021;13:782082.

38. Borghammer P, Van Den Berge N. Brain-First versus Gut-First Parkinson's Disease: A Hypothesis. Journal of Parkinson's disease. 2019;9(s2):S281-S95.

39. Horsager J, Andersen KB, Knudsen K, et al. Brain-first versus body-first Parkinson's disease: a multimodal imaging case-control study. Brain. 2020;143(10):3077-88.

40. Nuzum ND, Loughman A, Szymlek-Gay EA, Teo WP, Hendy AM, Macpherson H. To the Gut Microbiome and Beyond: The Brain-First or Body-First Hypothesis in Parkinson's Disease. Front Microbiol. 2022;13:791213.

41. Quik M, Mallela A, Ly J, Zhang D. Nicotine reduces established levodopa-induced dyskinesias in a monkey model of Parkinson's disease. Mov Disord. 2013;28(10):1398-406.

42. Quik M, Wonnacott S. α6β2* and α4β2* nicotinic acetylcholine receptors as drug targets for Parkinson's disease. Pharmacol Rev. 2011;63(4):938-66.

43. Chen JF, Xu K, Petzer JP, et al. Neuroprotection by caffeine and A(2A) adenosine receptor inactivation in a model of Parkinson's disease. J Neurosci. 2001;21(10):Rc143.

44. Joghataie MT, Roghani M, Negahdar F, Hashemi L. Protective effect of caffeine against neurodegeneration in a model of Parkinson's disease in rat: behavioral and histochemical evidence. Parkinsonism Relat Disord. 2004;10(8):465-8.

45. Khadrawy YA, Salem AM, El-Shamy KA, Ahmed EK, Fadl NN, Hosny EN. Neuroprotective and Therapeutic Effect of Caffeine on the Rat Model of Parkinson's Disease Induced by Rotenone. J Diet Suppl. 2017;14(5):553-72.

46. Postuma RB, Anang J, Pelletier A, et al. Caffeine as symptomatic treatment for Parkinson disease (Café-PD): A randomized trial. Neurology. 2017;89(17):1795-803.

47. Oertel WH, Müller H-H, Unger MM, et al. Transdermal Nicotine Treatment and Progression of Early Parkinson's Disease. NEJM Evidence. 2023;2(9):EVIDoa2200311.

48. Phillips MCL, Murtagh DKJ, Gilbertson LJ, Asztely FJS, Lynch CDP. Low-fat versus ketogenic diet in Parkinson's disease: A pilot randomized controlled trial. Mov Disord. 2018;33(8):1306-14.

49. Tan AH, Lim SY, Chong KK, et al. Probiotics for Constipation in Parkinson Disease: A Randomized Placebo-Controlled Study. Neurology. 2021;96(5):e772-e82.

50. Postuma RB. Neuroprotective Trials in REM Sleep Behavior Disorder: The Way Forward Becomes Clearer. Neurology. 2022;99(7 Suppl 1):19-25.

51. Clemens P, Baron JA, Coffey D, Reeves A. The short-term effect of nicotine chewing gum in patients with Parkinson's disease. Psychopharmacology (Berl). 1995;117(2):253-6.

52. Vieregge A, Sieberer M, Jacobs H, Hagenah JM, Vieregge P. Transdermal nicotine in PD: a randomized, double-blind, placebo-controlled study. Neurology. 2001;57(6):1032-5.

53. Group TPS. Randomized placebo-controlled study of the nicotinic agonist SIB-1508Y in Parkinson disease. Neurology. 2006;66(3):408-10.

54. Lieberman A, Lockhart TE, Olson MC, et al. Nicotine Bitartrate Reduces Falls and Freezing of Gait in Parkinson Disease: A Reanalysis. Front Neurol. 2019;10:424.

55. Postuma RB, Lang AE, Munhoz RP, et al. Caffeine for treatment of Parkinson disease: a randomized controlled trial. Neurology. 2012;79(7):651-8.

56. Paustenbach D, Galbraith D. Biomonitoring and biomarkers: exposure assessment will never be the same. Environ Health Perspect. 2006;114(8):1143-9.

57. von Storch H, Costa-Cabral M, Hagner C, et al. Four decades of gasoline lead emissions and control policies in Europe: a retrospective assessment. Sci Total Environ. 2003;311(1-3):151-76.

58. Logroscino G. The role of early life environmental risk factors in Parkinson disease: what is the evidence? Environ Health Perspect. 2005;113(9):1234-8.

59. Gallo V, Vineis P, Cancellieri M, et al. Exploring causality of the association between smoking and Parkinson's disease. Int J Epidemiol. 2019;48(3):912-25.

60. Gardener H, Gao X, Chen H, Schwarzschild MA, Spiegelman D, Ascherio A. Prenatal and early life factors and risk of Parkinson's disease. Mov Disord. 2010;25(11):1560-7.

61. Liu B, Chen H, Fang F, Tillander A, Wirdefeldt K. Early-Life Factors and Risk of Parkinson's Disease: A Register-Based Cohort Study. PLoS One. 2016;11(4):e0152841.

62. Unger MM, Spiegel J, Dillmann KU, et al. Short chain fatty acids and gut microbiota differ between patients with Parkinson's disease and age-matched controls. Parkinsonism Relat Disord. 2016;32:66-72.

63. Aho VTE, Houser MC, Pereira PAB, et al. Relationships of gut microbiota, short-chain fatty acids, inflammation, and the gut barrier in Parkinson's disease. Mol Neurodegener. 2021;16(1):6.

64. Chen SJ, Chen CC, Liao HY, et al. Association of Fecal and Plasma Levels of Short-Chain Fatty Acids With Gut Microbiota and Clinical Severity in Patients With Parkinson Disease. Neurology. 2022;98(8):e848-e58.

65. Shin C, Lim Y, Lim H, Ahn TB. Plasma Short-Chain Fatty Acids in Patients With Parkinson's Disease. Mov Disord. 2020;35(6):1021-7.

66. Neveu V, Nicolas G, Amara A, Salek RM, Scalbert A. The human microbial exposome: expanding the Exposome-Explorer database with gut microbial metabolites. Sci Rep. 2023;13(1):1946.

67. Qian Y, Yang X, Xu S, et al. Detection of Microbial 16S rRNA Gene in the Blood of Patients With Parkinson's Disease. Front Aging Neurosci. 2018;10:156.

68. van der Mark M, Brouwer M, Kromhout H, Nijssen P, Huss A, Vermeulen R. Is pesticide use related to Parkinson disease? Some clues to heterogeneity in study results. Environ Health Perspect. 2012;120(3):340-7.

69. Paul KC, Krolewski RC, Lucumi Moreno E, et al. A pesticide and iPSC dopaminergic neuron screen identifies and classifies Parkinson-relevant pesticides. Nat Commun. 2023;14(1):2803.

70. Hu X, Walker DI, Liang Y, et al. A scalable workflow to characterize the human exposome. Nat Commun. 2021;12(1):5575.

71. Park G, Tan J, Garcia G, Kang Y, Salvesen G, Zhang Z. Regulation of Histone Acetylation by Autophagy in Parkinson Disease. J Biol Chem. 2016;291(7):3531-40.

72. Song C, Kanthasamy A, Jin H, Anantharam V, Kanthasamy AG. Paraquat induces epigenetic changes by promoting histone acetylation in cell culture models of dopaminergic degeneration. Neurotoxicology. 2011;32(5):586-95.

73. Tsalenchuk M, Gentleman SM, Marzi SJ. Linking environmental risk factors with epigenetic mechanisms in Parkinson's disease. NPJ Parkinson's disease. 2023;9(1):123.

74. Henderson-Smith A, Fisch KM, Hua J, et al. DNA methylation changes associated with Parkinson's disease progression: outcomes from the first longitudinal genome-wide methylation analysis in blood. Epigenetics. 2019;14(4):365-82.

75. Kaiser S, Zhang L, Mollenhauer B, et al. A proteogenomic view of Parkinson's disease causality and heterogeneity. NPJ Parkinson's disease. 2023;9(1):24.

76. Irmady K, Hale CR, Qadri R, et al. Blood transcriptomic signatures associated with molecular changes in the brain and clinical outcomes in Parkinson's disease. Nat Commun. 2023;14(1):3956.

Appendices

Summary

Samenvatting

Publication List

Acknowledgement

Curriculum Vitae

Summary

Parkinson's disease (PD) is the second most common neurodegenerative disorder after Alzheimer's disease. Disability and death due to PD are rapidly increasing globally. PD symptoms get worse over time, and there is no cure for the disease due to an incomplete understanding of its etiology and mechanism. Genetic and environmental factors may lead to PD, with genetic factors only accounting for up to 10% of PD. Emerging research suggests associations between PD and various environmental factors, including pesticides, smoking, coffee consumption and metals. However, identifying specific causal agents has been challenging, primarily due to the scarcity of prospective studies and the absence of objective exposure markers.

This thesis aims to investigate the etiology of multiple environmental exposures, particularly metals, caffeine, and the role of the gut-brain axis in PD development. The strengths of the approaches applied in this thesis are 1) the prospective study design with blood samples obtained prior to PD diagnosis, 2) exposure assessment of blood-based markers, and 3) the application of metabolomics techniques, primarily via high-resolution mass spectrometry (HRMS), for measuring small molecular metabolites.

In this thesis, most chapters (**Chapters 3, 5, 6 and 7**) are based on a European prospective cohort, EPIC4PD, spanning >20 years of follow-up and identifying >700 incident PD cases. Blood samples, detailed dietary intake, and lifestyle information were collected from participants at recruitment. Within the EPIC4PD cohort, a nested case-control study was conducted, matching incident PD cases with controls based on age and sex. Erythrocyte and plasma samples, collected on average eight years prior to disease diagnosis for PD cases, were sourced from the EPIC4PD biobank for internal exposure assessment. The number of samples utilized in the study was contingent upon availability (362 case-control pairs for erythrocytes and 351 pairs for plasma). Furthermore, this thesis made use of an occupational population highly exposed to dioxin(-like) compounds and HRMS to establish an analytical pipeline of a chemical-wide approach (**Chapter 4**).

Impact of metal exposures on PD Risk

In **Chapter 2**, a systematic review was conducted, including 83 case-control studies and five cohort studies. Most of these studies exhibited limitations in their study design, primarily focusing on prevalent PD cases post-diagnosis, thereby potentially introducing reverse causality. Despite indications of pooled effects from meta-analyses for certain metal species, such as copper, iron, and zinc, significant heterogeneity was observed across studies. In **Chapter 3** we aimed to address these limitations by employing an

improved study design and by analyzing pre-diagnostic erythrocyte samples to investigate the relationship between erythrocyte metal levels and PD risk. The measurements involved eleven metal species, but the majority showed no significant association with PD risk. Notably, an inverse association was detected between lead levels and PD risk only among current smokers. However, this finding is likely confounded by smoking, a recognized inverse risk factor for PD, as cigarette smoke contributes to lead exposure among smokers. Despite these efforts, **Chapters 2 and 3** did not yield definitive conclusions regarding the roles of metal exposures in PD risk.

Establishment of a chemical-wide framework and investigating caffeine effect in PD

In Chapter 4 we established a comprehensive framework to explore the health impacts of a specific chemical group. The study subjects comprised a unique population with a history of exposure to dioxin(-like) compounds. Targeted gas chromatography (GC)-HRMS was employed to quantify the 29 currently recognized dioxin(-like) compounds in plasma. Moreover, untargeted GC-HRMS was applied to identify all potential chlorinated chemicals. Through correlating these compounds to 29 targeted compounds, 152 new dioxin(-like) related compounds were identified. Incorporating dioxin(-like) related compounds in a metabolome-wide association study provided richer insights into the biological responses associated with such exposures. This chemical-wide approach was applied in **Chapter 5** to investigate the effect of caffeine on PD within the EPIC4PD cohort. Notably, a 40% lower risk of PD was observed among the highest consumers of coffee within the EPIC4PD cohort. Untargeted HRMS characterized caffeine and 14 other metabolites in the caffeine pathway in the EPIC4PD nested case-control study. Caffeine and its primary metabolites, paraxanthine and theophylline, exhibited inverse associations with PD risk. These findings reinforce the long-implicated protective role of caffeine in PD. **Chapters 4 and 5** underscore the utility of untargeted metabolomics in detecting a broad spectrum of exposure markers, encompassing those that are less studied or currently unknown, thereby significantly aiding in the assessment of health impacts of environmental chemicals.

Exploration of the gut-brain axis in PD development

In Chapters 6 and 7 we investigated the relationship between the gut-brain axis in PD by blood-based microbiota-relevant markers. Lipopolysaccharide (LPS) is the outer membrane component of Gram-negative bacteria and a potent activator of innate immune response. LPS-binding protein (LBP) serves as a reactant of LPS in the blood and is utilized as a marker of LPS. In **Chapter 6**, elevated LBP levels were associated with an increased risk of PD in the nested case-control study. This positive association was notably more pronounced among women and overweight/obese individuals. These findings demonstrate the involvement of LPS-mediated inflammation in the gut-brain

crosstalk in PD. **Chapter 7** focuses on the identification of microbial metabolites in plasma, utilizing the aforementioned untargeted HRMS data in **Chapter 5**. Out of the 167 microbial metabolites identified, 13 were associated with PD risk under the criteria of a 0.05 p-value. These metabolites include various amino acids, bile acid, indoles, and hydroxy acid. Furthermore, through an analysis including the entire metabolome and subsequent pathway enrichment, three microbiota-relevant pathways—i) valine, leucine and isoleucine degradation, ii) butanoate metabolism, and iii) propanoate metabolism—were associated with PD risk. Importantly, personal traits and lifestyles might contribute to the variability in the involvement of the gut-brain axis in PD, with observations indicating more PD-associated pathways for men, smokers, and overweight/obese individuals.

In Chapter 8 we discuss the broader implications of the thesis findings, contextualizing the application of metabolomics in measuring the internal exposome, understanding the role of environmental factors in PD etiology, and addressing methodological considerations. The complex relationship between environmental factors and PD risk remains a paramount area of research. Metabolomics emerges as a promising tool due to its extensive coverage in measuring internal metabolite markers within health impact studies. Moving forward, the next step involves expanding the spectrum of environmental chemicals to facilitate an exposome-wide association study. Furthermore, integrating the metabolome with other omics platforms holds great promise in discovering novel PD biomarkers. This integration is poised to illuminate avenues for disease prevention, early detection, and treatment strategies.

Samenvatting

De ziekte van Parkinson is de op één na meest voorkomende neurodegeneratieve ziekte na de ziekte van Alzheimer. Invaliditeit en sterfte als gevolg van Parkinson nemen wereldwijd snel toe. De symptomen van Parkinson verslechteren in de loop van de tijd en er is geen genezing voor de ziekte vanwege onvolledig begrip van de etiologie en mechanismen. Een combinatie van genetische factoren en omgevingsfactoren kan leiden tot Parkinson, waarbij genetische factoren slechts tot 10% van de Parkinsongevallen verklaren. Een groeiend aantal onderzoeken suggereert associaties tussen Parkinson en verschillende omgevingsfactoren, waaronder pesticiden, roken, koffieconsumptie en metalen. Het identificeren van specifieke causale factoren is echter uitdagend gebleken, voornamelijk vanwege het gebrek aan prospectieve studies en het ontbreken van objectieve markers van blootstelling.

Dit proefschrift beoogt de etiologische rol van blootstelling aan verschillende omgevingsfactoren, met name metalen, cafeïne, en de zogenaamde “darm-hersen-as”, te onderzoeken bij de ontwikkeling van Parkinson. De kracht van dit proefschrift ligt in 1) het prospectieve studieontwerp waarin bloedmonsters zijn verkregen vóór de Parkinsondiagnose, 2) de beoordeling van markers in het bloed en 3) de toepassing van metabolomics-technieken, voornamelijk via hoge resolutie massaspectrometrie (HRMS), voor het meten van kleine moleculaire metabolieten.

In dit proefschrift zijn de meeste hoofdstukken (**Hoofdstuk 3, 5, 6 en 7**) gebaseerd op een Europese prospectieve cohortstudie, EPIC4PD, met meer dan 20 jaar follow-up en meer dan 700 mensen met incidente Parkinson. Bloedmonsters en gedetailleerde informatie over voedingsinname en levensstijl werden verzameld bij de werving van deelnemers. Binnen het EPIC4PD-cohort werd een geneste patiënt-controle studie uitgevoerd, waarbij mensen met incidente Parkinson werden gekoppeld met controles op basis van leeftijd en geslacht. Erytrocyten- en plasmamonsters, die gemiddeld acht jaar vóór de diagnose zijn verzameld, werden verkregen uit de EPIC4PD-biobank voor interne blootstellingsevaluatie. Het aantal monsters dat in de studie werd gebruikt, was afhankelijk van de beschikbaarheid (362 patiënt-controle paren voor erytrocyten en 351 paren voor plasma). Ook wordt in dit proefschrift gebruik gemaakt van een studie van een hoog blootgestelde beroepspopulatie aan dioxine(-achtige) verbindingen en HRMS om een analytische pijplijn op te zetten voor brede chemische screening (**Hoofdstuk 4**).

Het effect van metaalblootstellingen op het risico op Parkinson

In **Hoofdstuk 2** werd een systematische review uitgevoerd van 83 patiënt-controle studies en vijf cohortstudies. De meerderheid van deze studies vertoonde beperkingen in hun onderzoeksdesign, waarbij ze zich vooral richtten op mensen met Parkinson na de diagnose, waardoor er mogelijk sprake is van omgekeerde causaliteit. Ondanks aanwijzingen van gepoolde effecten uit meta-analyses voor bepaalde metaalsoorten, zoals koper, ijzer en zink, werd significante heterogeniteit waargenomen tussen studies. In **Hoofdstuk 3** beoogden we deze beperkingen aan te pakken door een verbeterd studiedesign te gebruiken en pre-diagnostische erytrocytenmonsters te analyseren om de relatie tussen erytrocytenmetaalniveaus en het risico op Parkinson te onderzoeken. De metingen omvatten elf metaalsoorten, maar de meerderheid vertoonde geen significante associatie met het Parkinsonrisico. Opvallend genoeg werd alleen bij huidige rokers een omgekeerde associatie waargenomen tussen loodniveaus en het risico op Parkinson. Deze bevinding is echter waarschijnlijk vertekend door roken, een erkende beschermende factor voor Parkinson, aangezien sigarettenrook bijdraagt aan de blootstelling aan lood bij rokers. Ondanks deze inspanningen hebben **Hoofdstuk 2 en 3** geen definitieve conclusies opgeleverd over de rol van metaalblootstellingen in het risico op Parkinson.

Opzet van een brede chemische screening raamwerk en onderzoek naar het effect van cafeïne op Parkinson

Hoofdstuk 4 richt zich op het opzetten van een uitgebreid raamwerk om de gezondheidseffecten van een specifieke chemische groep te onderzoeken. De studiepopulatie bestond uit een unieke populatie die in het verleden is blootgesteld aan dioxine(-achtige) stoffen. Gerichte gaschromatografie (GC)-HRMS werd gebruikt om de 29 momenteel erkende dioxine(-achtige) verbindingen in het plasma te kwantificeren. Bovendien werd ongerichte GC-HRMS toegepast om alle mogelijke gechloreerde chemicaliën te identificeren. Door deze te correleren met de 29 bekende verbindingen, werden 152 nieuwe dioxine(-achtige) gerelateerde verbindingen geïdentificeerd. Het opnemen van dioxine(-achtige) gerelateerde verbindingen in een metaboloom-brede-associatiestudie leverde meer inzicht op in de biologische reacties die gepaard gaan met dergelijke blootstellingen. Deze brede chemische screening benadering werd toegepast in **Hoofdstuk 5** om het effect van cafeïne op Parkinson binnen het EPIC4PD-cohort te onderzoeken. Opvallend genoeg werd een 40% lager risico op Parkinson waargenomen bij de grootste koffiedrinkers binnen het EPIC4PD-cohort. Ongerichte HRMS karakteriseerde cafeïne en 14 andere metabolieten in de cafeïne-route in de geneste EPIC4PD patiënt-controle studie. Cafeïne en zijn primaire metabolieten, paraxanthine en theofylline, verlaagde het risico op Parkinson. Deze

bevindingen versterken de al lang veronderstelde beschermende rol van cafeïne bij Parkinson. **Hoofdstuk 4 en 5** benadrukken het nut van ongerichte metabolomics bij het detecteren van een breed spectrum van blootstellingsmarkers, waaronder markers die minder bestudeerd of momenteel onbekend zijn, en draagt hiermee aanzienlijk bij aan de beoordeling van de gezondheidseffecten van chemicaliën in de omgeving.

Verkenning van de darm-hersen-as in de ontwikkeling van Parkinson

In Hoofdstuk 6 en 7 onderzochten we de relatie tussen de darm-hersen-as in Parkinson aan de hand van microbiota-gerelateerde bloedmarkers. Lipopolysaccharide (LPS) is het buitenmembraancomponent van Gram-negatieve bacteriën en een sterke activator van het aangeboren immuunrespons. Lipopolysaccharide-bindend (LBP) eiwit is een eiwit in de acute fase met het vermogen om bacterieel LPS (of endotoxine) te binden en door te geven. LBP kan gebruikt worden als een marker van LPS. In **Hoofdstuk 6** bleken verhoogde LBP-niveaus geassocieerd te zijn met een verhoogd risico op Parkinson in de geneste patiënt-controle studie. Deze positieve associatie was met name meer uitgesproken bij vrouwen en personen met overgewicht/obesitas. Deze bevindingen tonen de betrokkenheid van LPS-gemedieerde inflammatie in de darmen in Parkinson aan. **Hoofdstuk 7** richt zich op de identificatie van microbiële metabolieten in plasma, waarbij gebruik wordt gemaakt van de eerdergenoemde ongerichte HRMS-gegevens uit **Hoofdstuk 5**. Van de 167 geïdentificeerde microbiële metabolieten waren er 13 geassocieerd met het risico op Parkinson onder de criteria van een p-waarde van 0,05. Deze metabolieten omvatten verschillende aminozuren, galzuur, indolen en hydroxyzuur. Bovendien werden, door een analyse van het hele metaboloom en daaropvolgende routes, drie microbieel relevante routes — i) valine, leucine en isoleucinedegradatie, ii) butanoaatmetabolisme, en iii) propanoaatmetabolisme — geassocieerd met het Parkinsonrisico. Belangrijk is dat persoonlijke kenmerken en levensstijl bijdragen aan de variabiliteit in de betrokkenheid van de darm-hersen-as in Parkinson, met observaties die wijzen op meer Parkinson-geassocieerde biologische routes voor mannen, rokers en personen met overgewicht/obesitas.

In Hoofdstuk 8 bespreken we de bredere implicaties van de bevindingen van het proefschrift, waarbij de toepassing van metabolomics bij het meten van het interne exposoom, het begrijpen van de rol van omgevingsfactoren in de etiologie van Parkinson en het aanpakken van methodologische overwegingen in context wordt geplaatst. De complexe relatie tussen omgevingsfactoren en het risico op Parkinson blijft een belangrijk onderzoeksgebied. Metabolomics komt naar voren als een veelbelovend instrument vanwege de uitgebreide dekking bij het meten van interne metabole markers in onderzoeken naar gezondheidseffecten. Vooruitkijkend is de volgende stap het uitbreiden van het aantal exogene chemische verbindingen in bloed

om een exposoombrede associatiestudie mogelijk te maken. Bovendien is de integratie van het metaboloom met andere omics-platforms veelbelovend voor de ontdekking van nieuwe biomerkers voor Parkinson. Deze integratie biedt mogelijkheden voor ziektepreventie, vroegtijdige detectie en behandelingsstrategieën.

List of publications

Yujia Zhao, Yunjia Lai, Hilde Konijnenberg, et al. Association of Coffee Consumption and Prediagnostic Caffeine Metabolites with Incident Parkinson Disease in a Population-Based Cohort. *Neurology*, 2024. DOI: 10.1212/WNL.0000000000209201.

Yujia Zhao, Jeroen Meijer, Douglas I Walker, et al. Dioxin(-like)-Related Biological Effects through Integrated Chemical-wide and Metabolome-wide Analyses. *Environmental Science & Technology*, 2024. DOI: 10.1021/acs.est.3c07588.

Yujia Zhao, Douglas I Walker, Christina M Lill, et al. Lipopolysaccharide-binding Protein and Future Parkinson's Disease Risk: A European Prospective Cohort. *Journal of Neuroinflammation*, 2023. DOI: 10.1186/s12974-023-02846-2.

Yujia Zhao, Anushree Ray, Karin Broberg, et al. Prediagnostic Blood Metal Levels and the Risk of Parkinson's Disease: A Large European Prospective Cohort. *Movement Disorders*, 2023. DOI: 10.1002/mds.29602.

Yujia Zhao, Anushree Ray, Lützen Portengen, et al. Metal Exposure and Risk of Parkinson Disease: A Systematic Review and Meta-Analysis. *American Journal of Epidemiology*, 2023. DOI: 10.1093/aje/kwad082.

Yujia Zhao, Yuze Qi, Qingru Li, et al. Connexin43 Inhibition Attenuated Dopaminergic Neuronal Loss in the Lipopolysaccharide-induced Mice Model of Parkinson's Disease. *Neuroscience Letter*, 2022. DOI: 10.1016/j.neulet.2022.136471.

Dan Liu, **Yujia Zhao**, Yuze Qi, et al. Benzo(a)pyrene Exposure Induced Neuronal Loss, Plaque Deposition, and Cognitive Decline in APP/PS1 Mice. *Journal of Neuroinflammation*, 2020. DOI: 10.1186/s12974-020-01925-y.

Acknowledgements

It was an amazing journey when reflecting on the past four and half years. Never did I imagine I would embark on such an adventure, leaving my home thousands of miles away to pursue a PhD in the Netherlands at Utrecht University. This decision is an important moment in my both personal and professional life.

In the winter of 2018, I was planning to continue my study and desperately searching for European PhD vacancies from Beijing. It was then that my boyfriend, who had already been doing his PhD at IRAS for one year, reached out to his supervisor if he knew anyone who can take in new PhDs in the neuroscience field. Out of our surprise, the supervisor replied, *'She can do a PhD at our institute working with us…to do metabolomics on Parkinson's patients in a large biobank study'*. This unconventional recruitment process marked the beginning of my journey with the supervision team, **Prof. Roel Vermeulen** and **Dr. Susan Peters**. Throughout the subsequent years, I never regret joining the group and conducting research under their guidance. They are teachers, mentors, and role models in my professional career.

Roel, I extend my heartfelt appreciation for offering me the opportunity to contribute to this excellent EPIC project. I received comprehensive training in epidemiology and on the way being an independent researcher. In our meetings, you always quickly captured the most insightful messages from the vast amount of output, preventing me from losing my sight of the bigger picture and becoming bogged down in trivial details. Your input and feedback were consistently invaluable, guiding me through the next steps in my work. Regardless of busy schedule, your careful review and constructive comments on my papers were instrumental in honing my scientific and critical thinking. Your unwavering passion and attention to details are truly motivating for me.

Susan, I am deeply grateful for your support and encouragement, particularly during the challenging moments. Your kindness, patience, and steadfast support were invaluable during my PhD. I always felt reassured knowing that you had my back during times of stress. Your dedicated efforts and coordination were instrumental in ensuring the smooth progression of our project. Thank you for providing me with the space to explore and conduct research at my own pace. Your guidance, both through sharing your own experiences and connecting me with other experts, has been invaluable. I eagerly anticipate the opportunity to continue our collaboration in the future, alongside Roel.

I also would like to thank other supervisors at IRAS for their generous support me during my PhD. **Jelle**, I am thankful for your guidance in the field of molecular epidemiology at

the very beginning and for your contribution to my first metabolomics paper, which laid the foundation for my entire PhD work. **Lützen**, your tutorials on the methodology of epidemiology and statistics were immensely beneficial, and your vast knowledge and open-mindedness were truly inspiring.

I would like to thank my collaborators and EPIC co-authors for their invaluable contribution and constructive feedback. **Dr. Douglas Walker**, your dedicated work on the measurement of metabolome and protein markers for the dioxin project and EPIC project has been instrumental. Your papers serve as precious references on my desk. **Prof. Gary Miller**, I am deeply thankful for the excellent collaboration with your lab on the EPIC project. **Dr. Yunjia Lai**, your dedication on metabolomics annotation is appreciated, along with your constructive comments on my papers and thesis. **Prof. Christina Lill**, thank you for your support and feedback as the group leader of EPIC-Neuro. **Prof. Karin Broberg** and **Dr. Maria Kippler**, your collaboration on metal measurement and critical comments on my paper have been invaluable. **Prof. Bastiaan Bloem** and **Dr. Sirwan Darweesh**, your insightful inputs from the perspective of neurologists always enhanced the reasoning of my papers.

The success of my PhD and thesis is also attributed to the welcoming working environment at IRAS and the supportive team. **Christina**, thank you for handling all practical matters for me. **Louisa**, I appreciated all your work in scheduling meetings and negotiating defense date. **Mieke**, thanks for guiding me through courses and registration, as well as organizing all the PhD events. **Jack**, I cannot imagine how I would have managed blood sample aliquoting without your help. To **Gerard, Anke, Ulrike, Myrna, and Robert**, thanks for your willingness to practice with me and for your valuable contributions to the defense preparation.

I would like to express my gratitude to the friends I've made in Utrecht. The transition of moving to a new country would not be easy without you. **Anran**, you are such a good listener and a considerable friend. Thank you for all laughter, surprises, companionship you bring to me. You are always the first person I turn to, whether for good news or bad. Many thanks to **Fleur** and **Youchen** for being my paranymphs of my defense. Fleur, your warm introduce to the office and colleagues eased my nerves as I began my studies in a foreign country. I am grateful for inviting me to join dog walks and your birthday party. Youchen, thanks for your presence for Chinese holidays, and your encouragement during the final phase of my writing.

Jie, you have been a senior role model for me, generously sharing your experience on life, research and career. I cherish the memories of our times as neighbors. **Haykanush**, thank you for inviting me to experience Armenian and Italian food, as well as the King's day city walk with you and Leo. **Eugenio**, our conversations and dinners are truly

enjoyable and helped alleviate shared confusions and difficulties. **Dorothee**, I am grateful for your hospitality during the gap month between two renting leases. Your place always feels like home to me. Thanks to you, I had the opportunity to immerse myself in the local culture.

I am also thankful for the colleagues who have enriched my life over the years. To my 'old-320c' roommates—**Mariana, Warner, Jeroen, Judith, Laura and Xinhan**—I missed the camaraderie and shared moments during lunch and tea breaks. **Ilse**, thank you for organizing the writing retreat and inviting me. **Tara and Jolien**, your assistance in reviewing my Dutch summary for this thesis is greatly appreciated. To my fellow colleagues for small talks and lunch breaks—**Daniel, Femke, Xuan, Amanda, Yifeng**—your company has made my social life more enjoyable. **Wenxin and Max**, the boat trip in Utrecht was so cool, thanks for the invitation. Special thanks to the IRAS party committee—**Esther, Pepijn, Lora-Sophie, Kimberly, David, Ceder, Beatrice**—for the opportunity to organize events together. Thanks to my friends **Dan** in Germany, **Zhebin and Minjia** in Sweden, **Yao** in Utrecht, for all the happy times, nice meals and common misery we shared. To friends in China—**齐芪, 龚丹, 蔚晓阳, 陈娟, 黄焕焕**—although we are separated by distance, I know your heart is always with me.

My infinite gratitude goes to my family—**Mom and Dad**—for always respecting and supporting my decisions in life. I know how much you missed me during the three years of separation due to Covid. However, you always maintained positive and never allowed me to feel guilty about pursuing my dream. With you, I feel safe and loved. Special thanks to my sister, **赵雪纯**, for taking care of Mom and Dad in my absence. I am privileged to have you as my sister.

Lastly, to my partner, **Yongliang**, you were the main reason I moved to the Netherlands. Despite the stress of your own PhD, you have been unwavering in your support for me. I feel incredibly fortunate to have a partner who understands both the joys and challenges of pursuing a PhD. To my cat, **Edward**, your presence brings me endless joy and comfort.

Curriculum Vitae

Yujia Zhao was born on November 19, 1993 in Wenshui, Shanxi Province, China. She started her bachelor's studies in Preventive Medicine at West China College of Public Health, Sichuan University, China, in 2011. Subsequently, in 2016, she followed a master's degree in Environmental Toxicology at Peking University Health Center. During her master's study, she participated several projects focused on the impact of environmental exposures on neurodegenerative diseases such as Alzheimer's disease and Parkinson's disease. In 2019, she relocated to the Netherlands and started her PhD at the Institute for Risk Assessment Sciences (IRAS), Utrecht University, under the supervision of Prof. Roel Vermeulen and Dr. Susan Peters. She has been engaged in the Stichting ParkinsonFonds-financed project 'Metabolome and exposome profiling of Parkinson's disease', which is integrated into the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort. Her research passion lies in utilizing internal markers and metabolome to elucidate the etiology of Parkinson's disease. During her time at IRAS, Yujia also served as a member of the IRAS Party Committee and contributed to the Diversity, Safety, Inclusiveness workgroup. Since April 2024, she has been working as a postdoctoral researcher at IRAS, involved in the EPIC4ND project and the VHP4Safety project.