



**HAL**  
open science

# Risk factors of Parkinson's disease in women from the E3N cohort : role of physical activity, adiposity, and diabetes

Berta Portugal

► **To cite this version:**

Berta Portugal. Risk factors of Parkinson's disease in women from the E3N cohort : role of physical activity, adiposity, and diabetes. Santé publique et épidémiologie. Université Paris-Saclay, 2023. English. NNT : 2023UPASR005 . tel-04213574

**HAL Id: tel-04213574**

**<https://theses.hal.science/tel-04213574v1>**

Submitted on 21 Sep 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Risk factors of Parkinson's disease in women from the E3N cohort: role of physical activity, adiposity, and diabetes

*Facteurs de risque de la maladie de Parkinson chez les femmes de la cohorte E3N : rôle de l'activité physique, d'adiposité et du diabète*

## Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 570, Santé Publique EDSP  
Spécialité de doctorat : Epidémiologie  
Graduate School : Santé Publique. Référent : Faculté de médecine

Thèse préparée dans la unité de recherche CESP – Exposome et hérédité (Université Paris-Saclay, UVSQ, Inserm), sous la direction de **Alexis ELBAZ**, MD PhD

Thèse soutenue à Paris-Saclay, le 2 février 2023, par

**Berta PORTUGAL**

## Composition du Jury

Membres du jury avec voix délibérative

### **Christine TRANCHANT**

Professeur des Universités,  
Université de Strasbourg

Présidente

### **Cécilia SAMIERI**

Directrice de recherche, INSERM,  
Université de Bordeaux

Rapporteur & Examinatrice

### **Marc WEISSKOPF**

Professeur des Universités,  
Harvard T.H. Chan School of Public Health

Rapporteur & Eamineur

### **Bastiaan BLOEM**

Professeur des Universités,  
Radboud University

Examineur



*"Education is the most powerful weapon  
which you can use to change the world"*

Nelson Mandela



## **Research Unit**

Centre de recherche en épidémiologie et santé des populations (CESP), INSERM U1018, équipe « Exposome, hérité, cancer et santé », Hôpital Paul Brousse, 16 avenue Paul Vaillant Couturier, Villejuif, France.

## **Funding**

This thesis was financed by a doctoral scholarship from the French Ministry of Higher Education, Research and Innovation (MESRI).



# Acknowledgements

First and foremost I want to thank my thesis advisor Dr. Alexis Elbaz. It was an honor to be under his supervision for the past three and a half years. I have great admiration for his work and how he does it. I will always remember him as the one who taught me how to do epidemiological research. I am forever grateful for the opportunity and mentorship.

In addition I would like to thank Christine Tranchant, president and examiner, Cécilia Samieri and Marc Weisskopf, reviewers, and Bastiaan Bloem, examiner, for their acceptance to evaluate my thesis.

I am also grateful to Cécile Proust-Lima for helping us with the implementation and interpretation of some of the statistical models applied in the thesis.

I would like to thank the Doctoral School of Public Health (EDSP, N°570) together with the Minister of Higher Education, Research and Innovation (MESRI) for funding my PhD.

I am thankful to the members of the team Exposome and Heredity for the pleasant and stimulating working environment. A very special thanks to Fanny who always helped during my PhD.

Furthermore, I would like to thank all women from the E3N cohort study who made possible this study.

At last I would like to thank Alexandre for the support and motivation in the past years. My life is much better and less ordinary with you. To my parents who are my foundation, for their love and support. To my closest family for the support and to my friends who are always there for me no matter the distance.





# TABLE OF CONTENTS

<b>LIST OF ACRONYMS</b> .....	<b>13</b>
<b>LIST OF TABLES</b> .....	<b>15</b>
<b>LIST OF FIGURES</b> .....	<b>17</b>
<b>LIST OF ANNEXES</b> .....	<b>18</b>
<b>1. INTRODUCTION: PARKINSON'S DISEASE</b> .....	<b>19</b>
1.1. Clinical characteristics .....	21
1.2. Descriptive epidemiology .....	23
1.3. Etiology: risk and protective factors.....	24
1.3.1. Exercise and physical activity.....	29
1.3.1.1. Role of exercise and physical activity in PD patients.....	29
1.3.1.2. Relation between physical activity and incidence of PD.....	30
1.3.2. Body mass index and abdominal adiposity .....	37
1.3.3. Diabetes .....	44
<b>2. OBJECTIVES</b> .....	<b>51</b>
<b>3. PARTICIPANTS AND METHODS</b> .....	<b>55</b>
3.1. E3N cohort study .....	57
3.2. Ascertainment of Parkinson's disease .....	57
3.3. Covariates .....	60
3.4. Statistical Analysis.....	62
<b>4. Objective 1: Association between physical activity and Parkinson's disease</b> .....	<b>65</b>
4.1. Introduction .....	67
4.2. Methods .....	68
4.2.1. Assessment of physical activity .....	68
4.2.2. Nested case-control study .....	69
4.3. Statistical Analysis.....	72
4.3.1. Linear mixed models .....	72
4.3.2. Latent process mixed models .....	72
4.3.3. Latent process mixed models for multivariate longitudinal markers .....	73
4.3.4. Construction of the latent physical activity variable in E3N.....	74
4.3.5. Trajectories of latent physical activity .....	76

4.3.6.	Survival Analysis .....	76
4.4.	Results .....	78
4.4.1.	Trajectories of physical activity preceding Parkinson's disease .....	82
4.4.2.	Physical activity and Parkinson's disease incidence .....	82
4.5.	Discussion .....	88
<b>5.</b>	<b>Objective 2: Association between adiposity and Parkinson's disease.....</b>	<b>93</b>
5.1.	Introduction .....	95
5.2.	Methods .....	96
5.2.1.	Assessment of body mass index and abdominal adiposity .....	96
5.2.2.	Nested case-control study .....	97
5.3.	Statistical Analysis.....	97
5.3.1.	Survival Analysis .....	97
5.3.2.	BMI trajectories in PD patients and controls .....	99
5.4.	Results .....	100
5.4.1.	BMI and Parkinson's disease incidence .....	104
5.4.2.	Abdominal adiposity and Parkinson's disease incidence.....	110
5.4.3.	Trajectories of BMI in PD patients and controls.....	113
5.5.	Discussion .....	115
<b>6.</b>	<b>Objective 3: Association between diabetes and Parkinson's disease .....</b>	<b>121</b>
6.1.	Introduction .....	123
6.2.	Methods .....	124
6.2.1.	Ascertainment of diabetes .....	124
6.2.2.	Nested case-control study .....	127
6.3.	Statistical Analysis.....	127
6.3.1.	Survival Analysis .....	127
6.3.2.	Trajectories.....	128
6.3.3.	Meta-analysis .....	129
6.3.4.	Statistical power .....	129
6.4.	Results.....	130
6.4.1.	Diabetes and Parkinson's disease incidence.....	134
6.4.2.	Trajectories of diabetes in PD patients and controls.....	134
6.4.3.	Meta-analysis .....	135
6.5.	Discussion .....	142

<b>7. Main findings, Perspectives, and Conclusion .....</b>	<b>147</b>
7.1. Main findings.....	149
7.2. Perspectives and conclusions .....	150
<b>REFERENCES.....</b>	<b>153</b>
<b>ANNEXES.....</b>	<b>175</b>

# SCIENTIFIC PRODUCTION

## Articles related to the thesis

### **Published**

Portugal B, Artaud F, Domenighetti C, Roze E, Degaey I, Canonico M, Elbaz A. Body mass index and incidence of Parkinson's disease in French women from the E3N cohort study. *Neurology* 2023 Jan 17; 100(3):e324-e335. DOI:10.1212/WNL.0000000000201468.

### **Submitted**

Portugal B, Artaud F, Degaey I, Roze E, Fournier A, Severi G, Canonico M, Proust-Lima C, Elbaz A. Physical activity and Parkinson's disease in women over 29-years of follow-up: E3N cohort study. Under second revision after response to comments from the reviewers.

### **In preparation**

Portugal B, Artaud F, Degaey I, Roze E, Fournier A, Laouali N, Canonico M, Elbaz A. Diabetes and Parkinson's disease in women followed for 29 years: E3N cohort study.

## Other publications

### **Published**

Domenighetti C, Sugier PE, Kumar-Sreelatha AA, Schulte C, Grover S, Mohamed O, Portugal B, *et al.* Mendelian Randomisation Study of Smoking, Alcohol, and Coffee Drinking in Relation to Parkinson's Disease. *J Parkinsons Dis* 2021; 12(1):267-282. DOI: 10.3233/JPD-212851.

Domenighetti C, Sugier PE, Kumar-Sreelatha AA, Schulte C, Grover S, Mohamed O,

Portugal B, et al. Dairy intake and Parkinson's disease, a Mendelian randomization study. *Mov Disord* 2022; 37(4):857-864. DOI: 10.1002/mds.28902.

Grover S, Kumar-Sreelatha AA, Pihlstom L, Domenighetti C, Schulte C, Sugier PE, Radivojkov-Blagojevic M, Lichtner P, Mohamed, Portugal B, et al. Genome-wide meta-analysis and replication provides evidence of impact of a novel BST1 loci on age-at-onset in Parkinson's disease. *Neurology* 2022;99(7):e698-e710. DOI: 10.1212/wnl.0000000000200699.

Canonico M, Artaud F, Degaey I, Moisan F, Kabore R, Portugal B, et al. Incidence of Parkinson's disease in French women from the E3N cohort study over 27 years of follow-up. *Eur J Epidemiol* 2022; 37(5):513-523. DOI: 10.1007/s10654-022-00851-y.

Pesce G, Artaud F, Roze E, Degaey I, Portugal B, Nguyen TTH, Fournier A, Marie-Christine Boutron-Ruault MC, Severi G, Elbaz A, Canonico M. Reproductive characteristics, use of exogenous hormones, and Parkinson's disease in women from the E3N cohort study. *Brain* 2022.

### **Submitted**

Domenighetti C, Sugier PE, Kumar-Sreelatha AA, Schulte C, Grover S, Mohamed O, Portugal B, et al. Body mass index and Parkinson's disease : A bidirectional Mendelian randomization study.

### **Book chapter**

Elbaz A, Artaud F, Canonico M, Carcaillon-Bentata L, Domenighetti C, Moisan F, Perrin L, Portugal B, Sugier PE, Vlaar T. Chapitre Epidémiologie descriptive et étiologique de la maladie de Parkinson du livre "Pathologies du mouvement". Doin 2020.

## **Presentations**

Portugal B, Artaud F, Domenighetti C, Roze E, Degaey I, Canonico M, Elbaz A. Body mass index and incidence of Parkinson's disease in French women from the E3N cohort study. Movement Disorders Society Virtual Congress, September 2021.

Portugal B, Artaud F, Domenighetti C, Roze E, Degaey I, Canonico M, Elbaz A. Trajectoires de l'indice de masse corporelle avant la maladie de Parkinson : étude cas-témoins nichée dans la cohort de femmes E3N. Journées de Neurologie de Langue Française (JNLF), April 2022, Strasbourg (France).

Portugal B, Artaud F, Degaey I, Roze E, Severi G, Canonico M, Proust-Lima C, Elbaz A. Physical activity and incidence of Parkinson's disease in French women from the E3N cohort study. Movement Disorders Society Congress, September 2022, Madrid (Spain).

## LIST OF ACRONYMS

<b>α-Syn</b>	Alpha-synuclein
<b>AIC</b>	Akaike's information criterion
<b>APDs</b>	Antiparkinsonian drugs
<b>AMSAR-2</b>	Assessing the Methodological quality of Systematic Reviews 2
<b>ATC</b>	Anatomical Therapeutic Chemical
<b>AUC</b>	Area under the curve
<b>BDNF</b>	Brain-derived neurotrophic factor
<b>BMI</b>	Body mass index
<b>CI</b>	Confidence interval
<b>DBP</b>	Diastolic blood pressure
<b>DM</b>	Diabetes mellitus
<b>DPP4</b>	Dipeptidyl peptidase-4
<b>EPIC</b>	European Prospective Investigation into Cancer and Nutrition
<b>FU</b>	Follow-up
<b>GBD</b>	Global Burden of Disease
<b>GDNF</b>	Glial cell derived neurotrophic factor
<b>GEE</b>	Generalized estimating equations
<b>GLP-1</b>	Glucagon-like peptide-1 agonists
<b>GP</b>	General practitioner
<b>H</b>	Hours
<b>HbA1c</b>	Hemoglobin A1c
<b>HC</b>	Hip circumference
<b>HR</b>	Hazard ratio
<b>ICD</b>	International classification of disease
<b>INSEE</b>	Institut national de la statistique et des études économiques ( <i>National Institute of Statistics and Economic Studies</i> )
<b>IGF1</b>	Insulin-like growth factor one
<b>ICD</b>	International classification of disease
<b>IR</b>	Incidence rates
<b>Kcal</b>	Kilocalorie
<b>LPMM</b>	Latent process mixed models
<b>M</b>	Mean
<b>METs</b>	Metabolic equivalent of tasks
<b>mg</b>	Milligram
<b>MGEN</b>	Mutuelle Générale de l'Éducation Nationale ( <i>Insurance of the national education</i> )



<b>MR</b>	Mendelian randomization
<b>NR</b>	Not reported
<b>NSAIDs</b>	Non-steroidal anti-inflammatory drugs
<b>OR</b>	Odds ratio
<b>PA</b>	Physical activity
<b>PD</b>	Parkinson's disease
<b>Q</b>	Questionnaire
<b>RBD</b>	Rapid eye movement sleep behavior disorder
<b>RR</b>	Relative risk
<b>SBP</b>	Systolic blood pressure
<b>SD</b>	Standard deviation
<b>T1DM</b>	Type 1 diabetes mellitus
<b>T2DM</b>	Type 2 diabetes mellitus
<b>TZD</b>	Thiazolidinediones
<b>WC</b>	Waist circumference
<b>WHO</b>	World Health Organization
<b>WHR</b>	Waist-hip ratio
<b>WHtR</b>	Waist-height ratio
<b>Y</b>	Years

# LIST OF TABLES

Table 1: Potential risk and protective factors of Parkinson’s disease: summary of meta-analyses..... 27

Table 2: Cohort studies on the association between physical activity and Parkinson’s disease incidence..... 33

Table 3: Cohort studies on the association between midlife body mass index (BMI) and Parkinson’s disease incidence. .... 39

Table 4: Cohort studies on the association between diabetes and Parkinson’s disease incidence..... 47

Table 5: Assessment of self-reported characteristics in E3N questionnaires over the follow-up..... 63

Table 6: Assessment of self-reported recreational and household physical activities in E3N questionnaires over the follow-up..... 71

Table 7: Participant’s characteristics at baseline (1990-Q1)..... 80

Table 8: Participant’s characteristics at baseline (1990-Q1) according to Parkinson’s disease status at the end of the follow-up..... 81

Table 9: Association of time-varying physical activity with Parkinson’s disease incidence. .... 84

Table 10: Association of physical activity (Q5-1997) with Parkinson’s disease incidence (N=81,777): analyses adjusted for diet (Q3-1993)..... 86

Table 11: Association of physical activity with Parkinson’s disease incidence: exclusion of physical activity assessed at baseline (Q1)..... 87

Table 12: Association of physical activity with Parkinson’s disease incidence: analyses stratified by median age. .... 87

Table 13: Participants’ characteristics at baseline (1990-Q1) overall and according to body mass index. .... 102

Table 14: Participant’s characteristics at baseline (1990) according to Parkinson’s disease status at the end of the follow-up..... 103

Table 15: Association of time-varying body mass index with Parkinson’s disease incidence over the follow-up..... 105

Table 16: Association of body mass index (BMI) assessed at baseline in 1990 (Q1) with Parkinson’s disease incidence over the follow-up..... 107

Table 17: Association of body mass index (BMI) measured in 1997 (Q5) with Parkinson’s disease incidence over the follow-up: analyses adjusted for diet measured in 1993 (Q3). .... 108

Table 18: Association of time-varying body mass index with Parkinson’s disease incidence over the follow-up: adjustment for potential mediators (hypercholesterolemia, cardiovascular diseases, and diabetes). .....	109
Table 19: Association of time-varying adiposity measures (waist circumference, waist-hip ratio, and waist-height ratio) with Parkinson’s disease incidence over the follow-up. ....	111
Table 20: Association of adiposity measures (waist circumference, waist-hip ratio, waist-height ratio) measured in 1995 (Q4) with Parkinson’s disease incidence over the follow-up.....	112
Table 21: Participants’ characteristics at baseline (1990-Q1) overall and according to diabetes at the end of the follow-up.....	132
Table 22: Participants’ characteristics at baseline (1990) according to Parkinson’s disease status at the end of the follow-up.....	133
Table 23: Association between time-varying diabetes (prevalent and incident cases) and PD incidence over the follow-up.....	136
Table 24: Lagged analyses of the association between time-varying diabetes and PD incidence over the follow-up.....	137
Table 25: Association between incident diabetes (1990-2016) and PD incidence over the follow-up.....	138
Table 26: Lagged analyses of the association between time-varying incident diabetes (1990-2016) and PD incidence over the follow-up. ....	139

# LIST OF FIGURES

Figure 1: Hypothetical causal relation between three exposures and PD..... 53

Figure 2: Incidence rates of Parkinson’s disease in women from E3N cohort study and from Western Europe..... 60

Figure 3: Relationship between the 11 activities (METs-h/week) and the latent process. .... 75

Figure 4: Time-varying physical activity measures with no lag (A) and with 10-year lag (B). .... 77

Figure 5: Flow chart for inclusion into the study on the relation between physical activity and Parkinson’s disease incidence in survival analyses. .... 79

Figure 6: Trajectories of the latent physical activity (LPA) in cases with Parkinson’s disease and matched controls up to 29 years before the index date. .... 83

Figure 7: Hazard ratios of Parkinson’s disease in relation to latent physical activity (10-year lag) modelled with restricted cubic splines..... 85

Figure 8: Time-varying BMI measures with no lag (A) and with 5-year lag (B)..... 98

Figure 9: Flow chart for inclusion into the study (5-year, 10-year, and 20-year lag).. 101

Figure 10: Hazard ratios of Parkinson’s disease in relation to time-varying body mass index modelled with restricted cubic splines. .... 106

Figure 11: Trajectories of the frequency of the four categories of body mass index in cases and controls up to 29 years before the index date..... 114

Figure 12: Ascertainment of diabetes in the E3N cohort study..... 126

Figure 13: Prevalent cases and incident cases of diabetes..... 126

Figure 14: Flow-chart for inclusion into the study..... 131

Figure 15: Trajectories of the frequency of diabetes in cases and controls. .... 139

Figure 16: Meta-analysis of the association between diabetes and Parkinson’s disease. .... 140

Figure 17: Meta-analysis of the association between diabetes and Parkinson’s disease accounting for lagged analysis or duration of diabetes..... 141

# LIST OF ANNEXES

Annex 1: Résumé substantial..... 177

Annex 2: Results from the linear mixed model for the trajectories of physical activity with a retrospective time scale in the nested case-control study (1,196 PD cases, 23,879 controls). ..... 193

Annex 3: Results from the multinomial GEE logistic model for trajectories of the categories of body mass index with a backward time-scale (1,196 patients, 23,876 controls). ..... 194

Annex 4: Trajectories of the frequency of the four categories of body mass index in cases and controls up to 29 years before the index date: time modelled using restricted cubic splines. .... 195

Annex 5: Results from the GEE logistic model for trajectories of diabetes with a backward time-scale (1,200 patients, 25,200 controls). ..... 196

# **1. INTRODUCTION: PARKINSON'S DISEASE**



In 1817, Dr. James Parkinson (1755-1824) wrote *An essay on the shaking palsy*.<sup>[1]</sup> He described for the first time a disease called “paralysis agitans”, characterized by resting tremor, flexed posture, and festination. In 1872, this disease was later renamed by Jean-Martin Charcot (1825-1893) as Parkinson’s disease (PD) to honor its discoverer. Charcot was particularly important in improving the disease description and disseminating it internationally.<sup>[2]</sup>

## **1.1. Clinical characteristics**

Motor symptoms, including rest tremor, rigidity, bradykinesia, and postural instability, represent the core features of PD.<sup>[3, 4]</sup> In addition, non-motor symptoms, such as hyposmia, constipation, rapid eye movement sleep behavior disorder (RBD), urinary dysfunction, and depression can also be present and even precede the onset of motor symptoms by several years or decades in some patients.<sup>[5-9]</sup> The presence of non-motor symptoms before the diagnosis of PD characterize the prodromal phase of PD, defined by the onset of symptoms caused by neurodegeneration, without the presence of typical motor symptoms essential for diagnosis.<sup>[10]</sup> PD is characterized by marked heterogeneity between patients, in terms of clinical presentation, including age at onset, non-motor and motor symptoms, rate of progression, and treatment response.<sup>[2]</sup>

Mitochondrial dysfunction, oxidative stress, and neuro-inflammation are considered to be the main mechanisms involved in the etiology of PD.<sup>[11]</sup> PD is caused by the loss of dopaminergic neurons in the substantia nigra, characterized by abnormal aggregation and deposits of alpha-synuclein ( $\alpha$ -Syn) in Lewy bodies and Lewy neurites that represent the pathological hallmarks of the disease.<sup>[12]</sup> Lewy body pathology can start in the brain, the peripheral autonomic nervous system, and/or the olfactory bulb. The region where it starts may contribute to better understand the heterogeneity of the disease and define different PD subtypes with distinct prodromal and clinical



features.[13-15] The observation that the initial pathological aggregation and propagation of  $\alpha$ -Syn can occur in the brain or the periphery has led to the hypothesis that PD comprises two main subtypes: a body-first subtype, in which  $\alpha$ -Syn pathology originates in the autonomic and enteric nervous system and invades the central nervous system via the vagus nerve and sympathetic connectome, and a brain-first subtype, in which pathology arises in the brain itself, most often in the limbic system or sometimes secondary to entry via the olfactory bulb, and descends through the brainstem and into the periphery.

At the time of the clinical diagnosis of PD, it is estimated that ~40%-60% dopaminergic neurons from the substantia nigra have been lost.[7, 16] Backward extrapolation of the rate of nigrostriatal loss in autopsy or neuroimaging studies suggests that the premotor phase begins ~5-6 years before clinical diagnosis,[5, 17] in agreement with studies showing that subtle motor symptoms and fall-related fractures can be present over this period and before the diagnosis of PD.[6, 18, 19] In addition, epidemiologic studies suggest that prodromal symptoms such as constipation, anxiety, or RBD can be present 20 years or even earlier before PD diagnosis in some patients.[5, 20] RBD is tightly linked to the body-first type, since ascending  $\alpha$ -Syn pathology affects pontine structures before the nigral dopamine neurons. By contrast, patients with the brain-first subtype usually do not have RBD when motor symptoms emerge, although top-down RBD can develop once the pons becomes damaged.

In the absence of curative treatment, the treatment of PD is mainly symptomatic. Since the 70's, levodopa remains the first-line therapy of motor symptoms. Despite its remarkable benefits, long-term use of levodopa is accompanied by side effects, including motor fluctuations, dyskinesia, and hallucinations.[2] More advanced therapies, such as deep brain stimulation and magnetic resonance-guided focused ultrasound, are increasingly used in patients with complications that do not respond to drug adjustments.[21] However, not all PD patients are eligible for these advanced therapies.[22]

Increasingly, non-pharmacological therapies, in particular exercise, have gained recognition as effective interventions in PD patients.[23-27] Given its low cost and limited side effects, exercise may lead to benefits for most PD patients.[28] Clinical trials have showed that different types of aerobic exercise may stabilize and/or improve motor symptoms,[29, 30] sleep quality,[31] and cognitive function.[32, 33] Observational studies consistently showed that physical activity contributed to stabilize motor symptoms,[34-36] and was associated with lower mortality in PD patients.[35, 37] Additional studies are needed to better understand the mechanisms underlying the beneficial effects of exercise/physical activity and their potential long-term effects in PD.

Despite treatment, as the disease progresses, patients face an increased risk of falls, cognitive decline, and disability that contribute to poorer quality of life and reduced life expectancy in PD patients compared to persons without PD.[38-40] Given the lack of pharmacological curative treatments, alternative strategies are urgently needed in order to prevent or delay the onset of PD.[41, 42]

## **1.2. Descriptive epidemiology**

PD is an age-related disease; it is rare before the age of 50 years, and its incidence increases markedly after that age.[43] It is the second most prevalent neurologic disorder after Alzheimer's disease, and the fastest growing globally, in terms of prevalence, disability, and death.[44] Between 1990 and 2016, the age-standardized prevalence rate of PD increased by 22%[44] compared to 2%[45] for Alzheimer's disease and other dementias. According to the Global Burden of Disease (GBD) project, the number of PD patients increased from 2.5 million PD in 1990 to 6.1 million in 2016[44] and 8.5 million in 2019.[46] Projections foresee that in 2040 there will be 14.2 million PD cases worldwide. This increase is mainly explained by increasing life expectancy leading to a larger number of persons at risk. In addition, PD patients are living longer with the disease than before, which also contributes to explain increasing

prevalence.[44, 47]

In France, in 2018, the prevalence rate of PD was of 3.4 per 1,000 persons and the incidence rate was of 0.51 per 1,000 person-years. The number of PD patients increased from ~150,000 in 2010 to ~175,000 in 2018.[48] Projections estimate that in 2030 there will be 260,000 PD cases, corresponding to a 73% increase since 2010.[47]

Age-standardized PD incidence rates of PD are about 1.5 higher in men than in women.[49, 50] It is hypothesized that this difference may be explained by distinct occupational exposures in men and women,[51-53] and/or by a neuroprotective effect of estrogens.[54-56]

### **1.3. Etiology: risk and protective factors**

PD is a multifactorial disorder caused in most cases by a combination of genetic and environmental factors. The majority of PD cases are considered idiopathic cases. Only a small percentage of PD cases are explained by Mendelian genes with autosomal or recessive inheritance.[57, 58]

One recent overview of systematic reviews and meta-analysis on PD risk and protective factors included 46 articles published between 2011 and 2020.[59] In this review, each meta-analysis included  $\geq 5$  studies and  $\geq 388$  PD patients. Potential risk and protective exposures were categorized as habits, environmental agents, dietary factors, medical history and comorbid diseases, drugs, and biomarkers. The authors used the AMSAR-2 (Assessing the Methodological quality of Systematic Reviews 2) tool to analyze the quality of the studies. It included 16 items, such as use of appropriate methods for statistical combination of results, adequate investigations of publication bias and discussion of its impact on results, list of excluded studies and the justification for exclusion, account for the risk of bias of individual studies when interpreting or discussing the results of the review, registration of a protocol, and use a comprehensive literature search strategy. Based on the AMSAR-2 tool, no meta-analysis had high or

moderate quality, seven had low methodological quality and 39 had critically low quality.

Table 1 presents 32 exposures that were significantly associated with PD; of these, 16 were associated with lower PD risk (hazard ratios [HRs] comprised between 0.44 for smoking and 0.86 for welding exposure), while 18 were associated with increased risk (HRs comprised between 1.12 for aspirin use and 1.76 for pesticides exposure). We did not include in this table associations with prodromal symptoms of the disease (e.g., depression, constipation). Studies on diabetes yielded inconsistent results in case-control and cohort studies, and this issue will be later discussed in the Introduction (§1.3.3) and in the corresponding chapter (§6).

It is generally considered that systematic reviews and meta-analysis help to provide more reliable estimates of associations and to generate hypotheses for future research.[60] However, they also have limitations. Several studies combined results based on different study designs (i.e., cohort studies, case-control studies, cross-sectional studies) and different study designs have different sources of bias. For instance, case-control studies are more likely to be prone to recall bias and selection bias compared with cohort studies. Combining them will increase heterogeneity which limits interpretation and generalizability of results.[61] Differences in participants' selection (e.g., prevalent vs incident cases), exposure measurement (especially for time-varying exposures), and diagnostic criteria of PD may also increase heterogeneity.

Given the long prodromal phase of PD, bias due to reverse causation may contribute to some of the associations previously reported in epidemiologic studies. Reverse causation refers to situations in which the undiagnosed outcome precedes and leads to changes in the exposure instead of the other way around.[62] For instance, non-motor symptom of PD such as constipation may lead to changes in diet and weight; similarly, subtle motor symptoms may lead some persons to reduce the level of physical activity in the years preceding the diagnosis of PD. This issue represents a

major threat for epidemiological studies, in particular for studies with a short follow-up and those that assessed exposures close to the diagnosis of PD.[8]

Appropriate statistical methods are needed to take into account the potential for reverse causation. In particular, survival analyses allow to estimate the HR of PD over the follow-up by comparing the hazard of PD in exposed and non-exposed participants; including an exposure lag allows to estimate risk at time  $t$  in relation to exposures assessed earlier in time, i.e., at time  $t - lag$ . [63-65] However, few cohort studies have a sufficiently long follow-up in order to do this while maintaining a sufficient number of cases. In case-control studies, information on exposures assessed long before diagnosis are needed, but recall bias and exposure misclassification may be an issue in this context. [66, 67] Analyses of trajectories represent a complementary approach and allow to compare changes in the frequency of exposure before diagnosis in PD cases and matched controls, and to examine whether reverse causation may lead to changes in exposures in PD cases prior to disease onset, but few studies have used this method in the field of PD research. [68-70]

**Table 1: Potential risk and protective factors of Parkinson's disease: summary of meta-analyses.**

<b>Risk/protective factors</b>	<b>Category</b>	<b>Reference</b>	<b>Cases (n)</b>	<b>Studies (n cohorts)</b>	<b>Effect size (95% CI)</b>	<b>Heterogeneity (I<sup>2</sup>, %)</b>	<b>P</b>
<b>Habits</b>							
Alcohol intake	High vs low	Zhang, 2014	9,994	32 (8)	0.75 (0.66-0.85)	52.3	NR
Coffee intake	Yes vs no	Noyce, 2012	5,801	19 (6)	0.67 (0.58-0.76)	42.9	0.03
Tea intake	High vs low	Qi, 2014	1,929	8 (4)	0.63 (0.49-0.81)	51.7	NR
Caffeine	High vs low	Qi, 2014	2,659	7 (4)	0.55 (0.43-0.71)	53.0	NR
Smoking	Ever vs never	Noyce, 2012	17,856	67 (6)	0.64 (0.60-0.69)	49.6	<0.01
	Current vs never	Noyce, 2012	13,271	33 (7)	0.44 (0.39-0.50)	33.8	0.03
Physical Activity	High vs low	Fang, 2018	2,192	8 (7)	0.79 (0.68-0.91)	0.0	NR
<b>Dietary factors</b>							
Dairy products	High vs low	Jiang, 2014	1,083	5 (5)	1.40 (1.20-1.63)	8.2	0.37
Milk	High vs low	Hughes, 2017	1,725	4 (4)	1.56 (1.30-1.88)	0.0	0.50
Carbohydrate intake	High vs low	Wang, 2015	1,482	8 (4)	1.24 (1.05-1.48)	0.0	0.45
Energy intake	High vs low	Wang, 2015	1,553	8 (4)	1.39 (1.01-1.92)	83.8	<0.01
<b>Comorbidity and medical history</b>							
Head trauma	Yes vs no	Jafari, 2013	18,344	22 (1)	1.57 (1.35-1.83)	61.0	<0.01
Hypertension	Yes vs no	Noyce, 2012	5,993	12 (2)	0.74 (0.61-0.90)	76.5	<0.01
Infection of <i>Helicobacter pylori</i>	Yes vs no	Wang, 2020	NR	9 (1)	1.65 (1.43-1.92)	0.7	0.43
Infection of Hepatitis C virus	Yes vs no	Wang, 2020	NR	7 (5)	1.20 (1.01-1.41)	79.0	<0.01
Diabetes (case-control studies)	Yes vs no	Noyce, 2012	NR	9 (0)	0.72 (0.54-0.97)	64.9	<0.01
Diabetes (cohort studies)	Yes vs no	Noyce, 2012	NR	4 (4)	1.31 (1.10-1.57)	34.7	0.20
<b>Drugs</b>							
NSAIDs <sup>a</sup>	Use vs no use	Noyce, 2012	9,064	9 (4)	0.83 (0.72-0.95)	50.9	0.04
Ibuprofen	Use vs no use	Gao, 2011	2,168	5 (3)	0.73 (0.62-0.85)	NR	<0.01
Aspirin	Use vs no use	Gao, 2011	2,779	6 (3)	1.12 (1.01-1.23)	NR	<0.01
Calcium channel blockers	Use vs no use	Lang, 2015	6,709	5 (3)	0.76 (0.68-0.84)	10.7	0.35
Statins	Use vs no use	Bai, 2016	21,011	11 (6)	0.81 (0.71-0.92)	64.5	<0.01
Thiazolidinediones (TZDs)	Use vs no use	Zhu, 2019	2,644	5 (5)	0.70 (0.51-0.96)	86.0	<0.01

<b>Risk/protective factors</b>	<b>Category</b>	<b>Reference</b>	<b>Cases (n)</b>	<b>Studies (n cohorts)</b>	<b>Effect size (95% CI)</b>	<b>Heterogeneity (I<sup>2</sup>, %)</b>	<b>P</b>
<b>Environmental agents</b>							
Welding exposure	Yes vs no	Mortimer, 2012	3,001	9 (2)	0.86 (0.80-0.92)	0.0	0.82
Hydrocarbon exposure	Yes vs no	Palin, 2015	4,483	14 (0)	1.36 (1.13-1.63)	29.0	0.15
Solvents exposure	Yes vs no	Pezzoli, 2013	NR	16 (0)	1.35 (1.09-1.67)	35.5	NR
Farming	Yes vs no	Pezzoli, 2013	NR	24 (0)	1.30 (1.14-1.49)	43.2	NR
Rural living	Yes vs no	Pezzoli, 2013	NR	30 (0)	1.32 (1.15-1.51)	75.2	NR
Well drinking water	Yes vs no	Pezzoli, 2013	NR	37 (0)	1.34 (1.16-1.55)	66.4	NR
Pesticides exposure	Yes vs no	Pezzoli, 2013	NR	51 (0)	1.76 (1.56-2.04)	67.3	NR
Herbicides exposure	Yes vs no	Pezzoli, 2013	NR	19 (0)	1.33 (1.08-1.65)	55.0	NR
Insecticides exposure	Yes vs no	Pezzoli, 2013	NR	18 (0)	1.53 (1.12-2.08)	78.8	NR
<b>Biomarkers</b>							
Serum vitamin D	Low vs normal	Shen, 2015	966	6 (1)	1.50 (1.31-1.71)	55.9	0.05
Serum urate	High vs low	Shen, 2013	578	6 (1)	0.68 (0.50-0.91)	42.1	0.10

NR, not reported; CI, confidence intervals.

<sup>a</sup> NSAIDs, Non-steroidal anti-inflammatory drugs.

Source: Chen et al, 2021.

## **1.3.1. Exercise and physical activity**

### **1.3.1.1. Role of exercise and physical activity in PD patients**

Physical activity is one of the most promising disease-modifying lifestyle intervention candidates, with few adverse effects and low cost.[27] Two recent well-designed clinical trials confirmed that aerobic exercise improves motor function in PD patients. The home-based Park-in-Shape is a double blind, randomized clinical trial performed in a single-center, that included 130 PD patients (mean age=59 years) newly diagnosed with PD (diagnosed ~40 months before the start of the trial). Participants were randomly assigned to cycle on a stationary home-trainer ( $n=65$ , aerobic intervention group) or stretching ( $n=65$ , control group).[71] Those who practiced high-intensity cycling (30-45 min, 3 times a week, during 6 months) had their motor symptoms stabilized and better cardiovascular fitness compared with those in the control group.[71] The Parkinson Disease of Exercise (SPARX), a randomized phase 2 clinical trial, included 128 PD patients (mean age=64.9 years) within 5 years of diagnosis. Participants were not exercising at moderate intensity greater than 3 times/week and not expected to need dopaminergic medication within 6 months. They were randomly assigned to high-intensity treadmill exercise ( $n=43$ , 4 days a week, during 6 months), moderate-intensity treadmill exercise ( $n=45$ , 4 days a week, during 6 months) or to the control group ( $n=40$ , usual care). Those who practiced high-intensity exercise had a less pronounced motor decline compared with those in the usual care group, while moderate-intensity exercise showed no significant benefits.[72] Other types of exercise, such as progressive resistance training and dance therapy, have also showed beneficial effects on motor and non-motor symptoms.[25, 73] Nevertheless, most studies had a short duration (<6 months), and a small number of participants.

One recent observational study (569 early PD patients) with 6 years of follow-up



and repeated measures of physical activity showed that regular physical activity was not associated with motor function at baseline, but that higher physical activity over the follow-up was associated with slower deterioration of motor symptoms and activities of daily living.[34] These results should be carefully interpreted since postural and gait instability may lead patients to reduce exercise, so that results may be explained, at least in part, by reverse causation. In addition, the PASE questionnaire assesses physical activity based on a 1-week recall, which is less reliable than other questionnaires or objective measures (e.g., accelerometers).[74]

### **1.3.1.2. Relation between physical activity and incidence of PD**

Results from eight prospective studies on the role of physical activity before PD onset are summarized in Table 2. In all these studies, physical activity was assessed through questionnaires only once, at baseline; therefore, change in physical activity over the follow-up was not examined. These studies included different types of activities (e.g., household, commuting, occupational, recreational) and they generally expressed physical activity level as metabolic equivalent of task per hour per week (MET-h/week) categorized as light (<3 METs; e.g., walking slowly, cooking), moderate (3-6 METs; e.g., heavy cleaning) and/or vigorous (>6 METs; e.g., bicycling fast) physical activity.

Based on 2,192 PD cases from eight studies, a meta-analysis showed that physical activity at baseline was associated with lower PD incidence, in analyses with men and women combined, with a pooled HR of 0.79 (95% Confidence intervals[CI]=0.68-0.91) for those with higher levels of total physical activity compared to those with the lowest levels ( $I^2=0$ ).[75]

In sex-stratified analyses, in men, the pooled HR was of 0.90 (95% CI=0.85-0.95) per increase of 10 MET-h/week in total physical activity, and 0.83 (95% CI=0.76-0.90) per increase of 10 MET-h/week in moderate-to-vigorous physical activity (948 PD cases

from four cohort studies and one case-control study).[75] In women (604 PD cases from four cohort studies), neither total physical activity (HR per 10 MET-h/week=0.95, 95% CI=0.87-1.04), nor moderate-to-vigorous physical activity (HR per 10 MET-h/week=0.88, 95% CI=0.71-1.09) were significantly associated with PD.[75] The lack of association in women may be explained by lack of statistical power due to a smaller number of studies and PD patients compared to men, or by different physiological responses to physical activity.[76] Therefore, in order to develop appropriate PD interventions, additional studies are needed to determine if physical activity has beneficial effects in women.

Multiple mechanisms could be explain the inverse association between physical activity and PD, including neuroprotection of dopaminergic neurons, regulation of neurotransmitter transmission, modulation of neuro-inflammation and oxidative stress, and brain plasticity.[77-82]

The potential for reverse causation represents a key issue for the interpretation of the results of studies on the relation between physical activity and the incidence of PD, in particular for those studies with a relatively short follow-up.[8] During the prodromal phase of PD, it is possible that patients who will later develop PD reduce physical activity as a consequence of non-motor and subtle motor symptoms. To address this issue, some cohort studies excluded cases over the first 4 to 10 years of the follow-up,[69, 83-86] and only one study with a 4-year lag showed a significant inverse association, overall and in sex-stratified analyses (Table 2).[86] Another study reported a significant inverse association between higher physical activity in early adulthood and later incidence of PD in men but not in women.[69] Lag times of 4 to 10 years may be too short to adequately address the potential for reverse causation. Among the previous cohort studies, four had a follow-up period shorter than 10 years,[83, 86-88] three had a follow-up between 10 to 15 years,[69, 84] and only one had a longer follow-up of more than 15 years.[85] However, this last study was based on a small sample size ( $n=6,715$ ), and in main analysis without a lag there were only 3

PD cases among those who practiced high levels of physical activity.

*Physical activity has been identified as one of the most promising interventions in terms of its potential for PD prevention.[42] If the inverse association between physical activity and PD was not explained by reverse causation, it would have important implications for designing interventions. In order to address reverse causation bias, larger cohorts with long follow-ups, such as E3N, are needed. Moreover, analyses of physical activity trajectories prior to PD diagnosis would help understand the temporal relation between physical activity and PD.*

**Table 2: Cohort studies on the association between physical activity and Parkinson's disease incidence.**

Study (Year)	Data source (country)	Follow-up (mean, years); Participants (% women)	PD Definition	Type of PA (assessment, unit)	PA level	Cases (n)	Baseline Analysis HR (95% CI)	Lagged Analysis HR (95% CI)	Other PA exposures	
Chen <i>et al.</i> [69] (2005)	Health Professionals Follow-Up Study (HPFS, US)	1986-2000 (NR); 48,574 (0)	Self-report confirmed by neurologist/internists, medical records and validation	<b>Total PA<sup>a</sup></b> (self-report, MET-h/wk)	<b>Men</b>					Moderate and vigorous PA. PA at high school, college, ages 30-40, and early in life.
						<b>252</b>		<b>Lag=6 years</b>		
					<2.8	46	1.00 (Ref.)	1.00 (Ref.)		
					2.8-7.7	68	1.3 (0.9-1.9)	1.1 (0.6-2.1)		
					7.8-16.9	58	1.1 (0.7-1.6)	0.6 (0.3-1.3)		
	17-34	44	0.8 (0.6-1.3)	0.5 (0.2-1.1)						
	≥34	36	0.7 (0.5-1.1)	0.5 (0.2-1.1)						
	Nurses' Health Study (NHS, US)	1986-1998 (NR); 77,254 (100)	Self-report confirmed by neurologist/internists, medical records and validation	<b>Total PA<sup>b</sup></b> (self-report, MET-h/wk)	<b>Women</b>					Moderate and vigorous PA.
						<b>135</b>		<b>Lag=4 years</b>		
					<2.1	18	1.00 (Ref.)	1.00 (Ref.)		
2.1-4.6					23	1.1 (0.6-2.0)	0.8 (0.4-1.7)			
4.7-10.4					43	2.0 (1.1-3.5)	1.9 (1.0-3.7)			
10.5-21.7	22	1.0 (0.5-1.8)	0.9 (0.5-1.9)							
>21.7	29	1.3 (0.7-2.3)	1.1 (0.5-2.2)							
Logroscino <i>et al.</i> [87] (2006)	Harvard Alumni Health Study (US)	1988-1997 (NR); 10,714 (0)	Self-report, death certificates and validation	<b>Total energy expenditure<sup>c</sup></b> (self-report, kcal/wk)	<b>Men</b>					Moderate and vigorous sports/recreation, PA in college, and average energy expenditure (1962/ 1966/ 1977/ 1988).
						<b>101</b>		-		
					<1000	42	1.00 (Ref.)			
					1000-1999	28	1.15 (0.71-1.88)			
1999-2999	14	0.92 (0.50-1.71)								
≥3000	17	0.63 (0.36-1.12)								
Thacker <i>et al.</i> [83] (2008)	Cancer Prevention study II Nutrition Cohort (US)	1992-2001 (NR); 141,339 (55.8)	Self-report confirmed by neurologist/internists/general practitioners, medical records	<b>Total recreational PA<sup>d</sup></b> (self-report, MET-h/wk)	<b>Men</b>					-
						<b>264</b>		-		
					0	28	1.00 (Ref.)			
					3.5	72	1.0 (0.7-1.6)			
					4-13.5	45	0.9 (0.6-1.5)			
14-22.5	67	0.9 (0.6-1.5)								
≥23	52	0.8 (0.5-1.2)								

Study (Year)	Data source (country)	Follow-up (mean, years); Participants (% women)	PD Definition	Type of PA (assessment, unit)	PA level	Cases (n)	Baseline Analysis HR (95% CI)	Lagged Analysis HR (95% CI)	Other PA exposures				
			and validation		<b>Women</b>					-	-		
						<b>145</b>							
					0	10	1.00 (Ref.)						
					3.5	40	1.2 (0.6-2.5)						
					4.0-8.5	32	1.8 (0.8-3.7)						
					9.0-18.0	37	1.3 (0.6-2.6)						
					18.5	26	1.0 (0.5-2.2)						
					<b>Men and women combined</b>								Total recreational PA and moderate to vigorous activity at age 40.
						<b>409</b>		<b>Lag=4 years</b>					
						38	1.00 (Ref.)	1.0 (Ref.)					
						112	1.1 (0.7-1.6)	1.0 (0.5-1.9)					
						77	1.2 (0.6-2.2)	1.1 (0.7-1.9)					
						104	1.0 (0.7-1.5)	1.2 (0.7-1.9)					
	78	0.8 (0.6-1.2)	0.8 (0.5-1.3)										
Xu <i>et al.</i> [86] (2010)	NIH-AARP Diet and Health Study Cohort (US)	2000-2006 (NR); 213,701 (NR)	Self-report, medical records and validation	<b>Moderate to vigorous activities<sup>e</sup></b> (self-report, h/wk)	<b>Men</b>					-	-		
								<b>Lag=4 years</b>					
					Never/rarely			1.00 (Ref.)					
					<1			0.94 (0.67-1.32)					
					1-3	-	-	0.92 (0.70-1.21)					
					4-7			0.86 (0.65-1.12)					
					>7			0.68 (0.51-0.90)					
					<b>Women</b>								-
								<b>Lag=4 years</b>					
					Never/rarely			1.00 (Ref.)					
					<1			1.21 (0.74-1.97)					
					1-3	-	-	0.69 (0.44-1.08)					
					4-7			0.74 (0.48-1.15)					
>7			0.60 (0.37-0.95)										

Study (Year)	Data source (country)	Follow-up (mean, years); Participants (% women)	PD Definition	Type of PA (assessment, unit)	PA level	Cases (n)	Baseline Analysis HR (95% CI)	Lagged Analysis HR (95% CI)	Other PA exposures	
					<b>Men and women combined</b>					
					Never/rarely	-	-	<b>Lag=4 years</b>	Light, moderate to vigorous activities at ages 15-18, 19-29 and 35-39. Daily routines at work. Changes of PA between ages 35-39 and past 10 years.	
				<1	1.00 (Ref.)					
				1-3	1.02 (0.77-1.34)					
				4-7	0.85 (0.67-1.07)					
				>7	0.82 (0.65-1.03)					
							0.65 (0.51-0.83)			
Sääksjärvi <i>et al.</i> [85] (2014)	Finnish Mobile Clinic Health Examination Survey (Finland)	1973-1994 (NR); 6,715 (NR)	Drug reimbursement, hospital records and validation	<b>Leisure-time PA<sup>f</sup></b> (self-report, h/wk)	<b>Men and women combined</b>					
						<b>97</b>		<b>Lag=10 years</b>		
					None	32	1.00 (Ref.)	1.00 (Ref.)		
					Light	62	0.73 (0.47-1.13)	0.65 (0.38-1.10)		
					Heavy	3	0.27 (0.08-0.90)	0.25 (0.06-1.09)		
Yang <i>et al.</i> [84] (2015)	Swedish National March Cohort (Sweden)	1997-2010 (12.6); 43,368 (64.3)	Swedish Patient Register	<b>General PA<sup>g</sup></b> (self-report, MET-h/wk)	<b>Men</b>					
						<b>158</b>		<b>Lag=8 years</b>	Household and community activity, leisure time exercise, total PA in a 24-h day (Energy expenditure questionnaire), occupational physical activity.	
					<3.9	57	1.00 (Ref.)	1.00 (Ref.)		
					3.9-6.6	61	1.04 (0.72-1.52)	1.11 (0.65-1.90)		
					≥6.6	40	0.53 (0.33-0.85)	0.69 (0.36-1.29)		
					<b>Women</b>					
						<b>126</b>		<b>Lag=8 years</b>		
					<4.0	45	1.00 (Ref.)	1.00 (Ref.)		
					4.0-6.2	44	0.81 (0.52-1.25)	1.03 (0.60-1.77)		
					≥6.2	37	0.85 (0.54-1.34)	0.81 (0.44-1.49)		
					<b>Men and women combined</b>					
						<b>286</b>		<b>Lag=8 years</b>		
					<3.9	104	1.00 (Ref.)	1.00 (Ref.)		
					3.9-6.4	105	0.88 (0.66-1.18)	1.05 (0.71-1.55)		
					≥6.4	77	0.72 (0.53-0.99)	0.78 (0.51-1.20)		

Study (Year)	Data source (country)	Follow-up (mean, years); Participants (% women)	PD Definition	Type of PA (assessment, unit)	PA level	Cases (n)	Baseline Analysis HR (95% CI)	Lagged Analysis HR (95% CI)	Other PA exposures
Llamas-Velasco <i>et al.</i> [88] (2021)	Neurological Disorders in Central Spain study (Spain)	1994-1998 (3.4); 2,943 (57.1)	Self-report, neurological examination	<b>Daily PA<sup>h</sup></b> (self-report, hours)	<b>Men and women combined</b>				
						<b>22</b>			
					≤ 15.6	-	1.00 (Ref.)	-	-
> 19.4	-	0.42 (0.71-1.04)							

PA, physical activity; h/wk, hours per week; NR, not reported.

<sup>a</sup> Total PA: cumulative sum of average time spent per week on walking/hiking outdoors, jogging, running, bicycling, lap swimming, tennis, squash or racket ball, calisthenics or rowing, and stairs climbed.

<sup>b</sup> Total PA: cumulative sum of average time spent per week on calisthenics/aerobics/aerobic dance/rowing machine, and stairs climbed.

<sup>c</sup> Total energy expenditure: cumulative sum of average time spent per week on walking, stair climbing activities, sports and recreational activities.

<sup>d</sup> Total recreational PA: cumulative sum of average time spent per week on walking, jogging/running/lap swimming, tennis or racquetball, bicycling/stationary bike, aerobics/calisthenics, and dancing.

<sup>e</sup> Moderate to vigorous activities: cumulative sum of time spent on tennis, biking, swimming, and heavy housework.

<sup>f</sup> Leisure-time PA: None; light PA – at least 4h/week doing activities such as walking and cycling; heavy PA –at least 3h/week doing activities such as jogging, skiing, vigorous gardening.

<sup>g</sup> General PA: cumulative sum of average time spent per week on household, community activity and leisure exercise.

<sup>h</sup> Daily PA: Sedentary (≤15.6h doing only minimal house chores or short walks at home), light PA (≤17.6h doing regular house chores, walking independently at home), high PA (>19.4h performing heavy housework, walking more than 1km or practicing any sport regularly).

### **1.3.2. Body mass index and abdominal adiposity**

Weight loss is common in PD patients who have lower average weight than controls, with a difference that increases with disease duration.[89] A combination of factors is likely to be involved, including increased energy expenditure, appetite loss, nausea, anosmia, and gastrointestinal function impairment.[90]

The relation between body mass index (BMI) and the incidence of PD is complex and previous studies provided inconsistent results on their association (Table 3). All these studies assessed weight at baseline, and did not take into account changes in weight over the follow-up.

A meta-analysis of 10 cohort studies showed no association between continuous baseline BMI and PD, both overall (HR per 5 kg/m<sup>2</sup>=1.00, 95% CI=0.89-1.12; heterogeneity I<sup>2</sup>=64.5%, P=0.003), and in men (HR=1.03, 95% CI=0.90-1.18) and women (HR=1.04, 95% CI=0.83-1.30).[91] In another meta-analysis, overweight and obesity were not associated with PD, while underweight was associated with a 20% increase in PD risk compared with those in the normal range of BMI.[92] Differences in the duration of follow-up across studies partly explained the between-study heterogeneity. In addition, two Mendelian randomization (MR) studies, which are less likely to be affected by reverse causation and confounding than observational studies, reported an inverse association between genetically-predicted BMI and PD risk, thus suggesting that increasing BMI is associated with lower risk of PD.[93, 94]

Studies based on measures of abdominal adiposity (waist circumference, WC; waist-hip ratio, WHR; waist-height ratio, WHtR) also yielded inconsistent findings.[95-99] While most studies found no association between WC and PD risk,[95-98] higher WC was associated with higher PD risk in two studies from South Korea based on the same healthcare databases with follow-ups comprised between 6 and 9 years.[99, 100] Alternatively, a MR study that examined nine different adiposity traits reported an



inverse association between adiposity and PD.[101]

Reverse causation may contribute to explain some of the previous results. Prodromal symptoms such as constipation and olfactory dysfunction may lead to changes in diet and caloric intake and, as a consequence, to changes in weight before diagnosis.[8, 102] To address this issue, four studies used a lag of  $\leq 5$  years,[95-97, 99, 103-105] two used a lag of 10 years[85, 106] and one a lag of 15 years (Table 3).[85] Therefore, larger cohorts with longer follow-ups are needed to assess whether reverse causation contributes to the inverse association between adiposity and PD. Trajectories prior to PD diagnosis would also help to understand the temporality of the relationship.

Given that a set of assumptions are met, Mendelian randomisation (MR) uses genetic instruments associated with an exposure to estimate exposure-disease associations that are not biased by confounding or reverse causation. Two MR studies reported that obesity was associated with lower PD incidence.[93, 94] Furthermore, a MR study of nine different adiposity traits reported the same inverse association.[101]. The mechanisms underlying this puzzling association are not well understood, but it has been hypothesized that hyperinsulinemia[102, 107-109] or changes in microbiota related to obesity[110, 111] may be involved.

*Due to inconsistent findings between observational and MR studies, given the relation between PA and adiposity, and in order to examine the impact of reverse causation, we selected adiposity as our second exposure of interest.*

**Table 3: Cohort studies on the association between midlife body mass index (BMI) and Parkinson's disease incidence.**

Study (Year)	Data source (country)	Participants (% women)	Follow-up (mean, years)	BMI assessment	BMI measure	Parkinson's Disease definition	BMI (kg/m <sup>2</sup> )	Cases (n)	Adjusted HR (95% CI)	Lagged analysis Adjusted HR (95% CI)
Chen <i>et al.</i> [96] (2004)	Health Professionals Follow-Up Study (HPFS, US)	48,574 (0)	1986-2000 (NR)	Self-report	Baseline	Self-report confirmed by neurologist/ internists, medical records and validation	<b>Men</b>			
								<b>249</b>		
							<23	44	1.00 (Ref.)	-
							23-24.9	76	1.1 (0.8-1.7)	
							25-26.9	64	1.0 (0.7-1.5)	
							27-29.9	48	1.1 (0.8-1.7)	
	≥30	17	1.0 (0.6-1.8)							
	<b>Women</b>									
		<b>202</b>								
	<23	93	1.00 (Ref.)	-						
	23-24.9	49	1.0 (0.7-1.5)							
	25-26.9	29	1.0 (0.7-1.6)							
	27-29.9	17	0.7 (0.4-1.2)							
	≥30	14	0.7 (0.4-1.2)							
<b>Men and women combined</b>										
			<b>Lag=4 years</b>							
<23	-	1.00 (Ref.)	1.00 (Ref)							
23-24.9	-	1.0 (0.9-1.4)	-							
25-26.9	-	1.0 (0.8-1.4)	-							
27-29.9	-	1.0 (0.7-1.3)	-							
≥30	-	0.8 (0.6-1.2)	0.8 (0.5-1.2)							
Hu <i>et al.</i> [103] (2006)	Six Finnish population Surveys (Finland)	45,806 (51.2)	1972-2002 (18.8)	Nurses' examination	Baseline	Drug reimbursement	<b>Men</b>			
								<b>272</b>		
							<23	22	1.00 (Ref.)	-
							23-24.9	61	1.97 (1.21-3.22)	
25-26.9	61	1.83 (1.12-2.99)								

Study (Year)	Data source (country)	Participants (% women)	Follow-up (mean, years)	BMI assessment	BMI measure	Parkinson's Disease definition	BMI (kg/m <sup>2</sup> )	Cases (n)	Adjusted HR (95% CI)	Lagged analysis Adjusted HR (95% CI)	
							27-29.9	83	2.34 (1.45-3.78)		
							≥30	45	2.44 (1.44-4.15)		
							<b>Women</b>				
								<b>254</b>			
							<23	33	1.00 (Ref.)		
							23-24.9	43	1.50 (0.95-2.37)		
							25-26.9	50	1.65 (1.05-2.59)		
							27-29.9	61	1.79 (1.15-2.80)		
							≥30	67	1.77 (1.12-2.78)		
							<b>Men and women combined</b>				
								<b>526</b>			
							<b>Lag=5 years</b>				
							<23	55	1.00 (Ref.)		1.00 (Ref)
							23-24.9	104	1.70 (1.23-2.37)		1.67 (1.18-2.36)
25-26.9	111	1.70 (1.23-2.37)	1.58 (1.11-2.23)								
27-29.9	144	2.02 (1.46-2.79)	1.90 (1.36-2.67)								
≥30	112	2.03 (1.44-2.85)	1.88 (1.31-2.70)								
Logroscino <i>et al.</i> [104] (2007)	Harvard Alumni Health Study (US)	10,812 (0)	1988-1998 (NR)	Self-report	Baseline	Self-report, death certificates and validation	<b>Men</b>				
								<b>106</b>		<b>Lag=4 years</b>	
							<22.5	40	1.51 (0.95-2.40)	NR	
							22.5-24.9	34	1.00 (Ref.)		
≥25.0	32	0.86 (0.53-1.41)									
Palacios <i>et al.</i> [97] (2011)	Cancer Prevention study II Nutrition Cohort (US)	147,096 (54.4)	1992- 2005 (NR)	Self-report	Baseline	Self-report confirmed by neurologist/ internists/family physicians, medical records	<b>Men</b>				
								<b>413</b>		<b>Lag=5 years</b>	
							18.5-23	62	1.00 (Ref.)	NR	
							23-24.9	88	0.93 (0.67-1.28)		
							25-26.9	122	1.17 (0.86-1.59)		
27-29.9	95	1.06 (0.77-1.47)									

Study (Year)	Data source (country)	Participants (% women)	Follow-up (mean, years)	BMI assessment	BMI measure	Parkinson's Disease definition	BMI (kg/m <sup>2</sup> )	Cases (n)	Adjusted HR (95% CI)	Lagged analysis Adjusted HR (95% CI)				
						and validation	≥30	46	0.98 (0.66-1.46)					
							<b>Women</b>							
								<b>226</b>		<b>Lag=5 years</b>				
							18.5-23	72	1.00 (Ref.)	NR				
							23-24.9	52	1.08 (0.76-1.55)					
							25-26.9	28	0.73 (0.47-1.14)					
							27-29.9	41	1.18 (0.80-1.74)					
							≥30	33	1.02 (0.66-1.56)					
							<b>Men and women combined</b>							
							18.5-23	-	1.00 (Ref.)	NR				
							23-24.9	-	1.00 (0.78-1.26)					
							25-26.9	-	0.95 (0.60-1.50)					
							27-29.9	-	1.11 (0.86-1.42)					
≥30	-	1.00 (0.75-1.34)												
Kyrozis <i>et al.</i> [112] (2013)	EPIC Greece (Greece)	25,407 (59.3)	1993-2010 (NR)	Self-report	Baseline	Self-report and phone interview validation	<b>Men and women combined</b>							
							Per 10	88	0.85 (0.53-1.39)	-				
Sääksjärvi <i>et al.</i> [85] (2014)	Finnish Mobile Clinic Health Examination Survey (Finland)	6,715 (52.4)	1973-1994 (NR)	Nurses' examination	Baseline	Drug reimbursement, hospital records and validation	<b>Men and women combined</b>							
								<b>97</b>		<b>Lag=10 years</b>				
							<23	13	1.00 (Ref.)	1.00 (Ref.)				
							23-24.9	16	1.04 (0.50-2.17)	1.17 (0.50-2.76)				
							25-27.4	21	0.91 (0.45-1.83)	0.69 (0.29-1.64)				
							27.5-29.9	26	1.46 (0.74-2.87)	1.18 (0.52-2.69)				
							≥30	21	1.09 (0.54-2.21)	1.09 (0.47-2.51)				
							<b>Lag=15 years</b>							
							<27.5	-	-	1.00 (Ref.)				
≥27.5	-	-	3.21 (1.42-7.28)											

Study (Year)	Data source (country)	Participants (% women)	Follow-up (mean, years)	BMI assessment	BMI measure	Parkinson's Disease definition	BMI (kg/m <sup>2</sup> )	Cases (n)	Adjusted HR (95% CI)	Lagged analysis Adjusted HR (95% CI)
Sääksjärvi <i>et al.</i> [106] (2015)	Mini-Finland Health Survey (Finland)	6,641 (NR)	1978-2007 (NR)	Nurses' examination	Baseline	Drug reimbursement, hospital records and validation	<b>Men and women combined</b>			
								<b>86</b>		<b>Lag= 10 years</b>
							<25	29	1.00 (Ref.)	1.00 (Ref.)
							≥25	57	1.18 (0.75-1.87)	1.75 (1.00-3.07)
Roos <i>et al.</i> [105] (2018)	Swedish National March Cohort (Sweden)	41,638 (64.2)	1997-2010 (NR)	Self-report	Baseline	Swedish Patient Register and death certificates	<b>Men and women combined</b>			
								<b>198</b>		<b>Lag=2, 3, and 5 years</b>
							<25	108	1.00 (Ref.)	NA
							[25-30[	74	0.96 (0.68-1.38)	
≥30	16	1.13 (0.60-2.12)								
Jeong <i>et al.</i> [99] (2018)	National health Insurance Service (South Korea)	6,800,601 (49.9)	2009-2017 (7.3)	Health examination	Baseline	NHIS registration program for rare intractable diseases (RID)	<b>Men and women combined</b>			
								<b>33,443</b>		<b>Lag=5 years</b>
							<18.5	1,122	1.28 (1.21-1.36)	1.16 (1.06-1.26)
							[18.5-23[	11,626	1.00 (Ref.)	1.00 (Ref.)
							[23-25[	8,891	0.97 (0.94-0.99)	0.98 (0.95-1.02)
							[25-30[	10,726	0.90 (0.88-0.93)	0.91 (0.88-0.94)
≥30	1,078	0.77 (0.72-0.82)	0.80 (0.74-0.86)							
Riso <i>et al.</i> [95] (2019)	NeuroEPIC4P D study (Europe)	220,347 (62.2)	1992-2012 (12)	Self-report	Baseline	Self-report, clinical records, mortality certificates and validation	<b>Men</b>			
								<b>378</b>		<b>Lag=5 years</b>
							≤24.9	127	1.00 (Ref.)	1.00 (Ref.)
							]24.9-29.9]	191	0.85 (0.68-1.08)	0.85 (0.64-1.12)
							>29.9	60	0.84 (0.60-1.17)	0.71 (0.46-1.08)
							<b>Women</b>			
								<b>356</b>		<b>Lag=5 years</b>
							≤24.9	138	1.00 (Ref.)	1.00 (Ref.)
]24.9-29.9]	139	0.99 (0.77-1.28)	0.89 (0.65-1.22)							
>29.9	79	1.11 (0.81-1.53)	1.00 (0.67-1.50)							

Study (Year)	Data source (country)	Participants (% women)	Follow-up (mean, years)	BMI assessment	BMI measure	Parkinson's Disease definition	BMI (kg/m <sup>2</sup> )	Cases (n)	Adjusted HR (95% CI)	Lagged analysis Adjusted HR (95% CI)
Park <i>et al.</i> [98] (2021)	National health Insurance Service (South Korea)	2,815,135 (33.2)	2008-2016 (~4.4)	Health examination	Baseline	NHIS registration program for rare intractable diseases (RID)	<b>Men and women combined</b>			-
								<b>1,962</b>		
							<18.5	82	1.12 (0.89-1.40)	
							18.5-22.9	1,065	1.00 (Ref.)	
							23-24.9	852	0.997 (0.91-1.09)	
							25-29.9	934	0.96 (0.88-1.05)	
≥30	94	1.12 (0.91-1.39)								

NR, not reported.

### 1.3.3. Diabetes

Observational studies suggest that PD patients with type 2 diabetes mellitus (T2DM) may be more prone to have worse motor symptoms, cognitive impairment, and reduced response to treatment than PD patients without diabetes.[113]

Results from prospective studies on the relation between diabetes before PD onset and subsequent incidence of PD are summarized in Table 4. A meta-analysis of seven prospective studies (including 20,628 PD patients) reported that diabetes increased PD risk with a pooled effect size of 1.29 (95% CI=1.19-1.40,  $I^2=51%$ ,  $P$ -heterogeneity=0.057).[114] Similar results were reported in another meta-analysis based on nine prospective studies (22,825 PD patients, RR=1.29, 95% CI=1.15-1.45,  $I^2=93.9%$ ,  $P$ -heterogeneity $\leq 0.001$ ).[115] Between-study heterogeneity was highly significant and may be explained by differences in PD ascertainment, diabetes assessment, duration of follow-up, and adjustment for different confounders.

Alternatively, a meta-analysis of 14 case-control studies reported an inverse association between diabetes and PD risk (21,395 PD patients, OR=0.75, 95% CI=0.58-0.98,  $I^2=75%$ ,  $P$ -heterogeneity= <0.001).[116]

The reasons for different results in cohort and case-control studies are unclear. It has been argued that case-control studies are more prone to selection and recall biases than cohort studies; the source of controls (e.g., hospital or population-based) may also have a strong impact on their findings. In addition, survival bias may be an issue in case-control studies for exposures that are strongly associated with survival. Diabetic patients have a higher mortality risk, and those who survived until old age, when PD is usually diagnosed, may be different compared to those who did not.[116, 117] This may also be an issue in most cohort studies that included patients with prevalent diabetes at baseline, who may be different from those who developed diabetes during the follow-up. One prospective cohort study[70] and one study based

on health care databases[118] included both prevalent diabetic patients at baseline and incident cases of diabetes over the follow-up, but did not perform analyses restricted to incident cases. Another study based on health care databases compared PD incidence in a cohort of incident diabetic patients to PD incidence in a cohort of matched participants without diabetes who were followed for a maximum of 11 years.[119]

Surveillance or detection bias may also be an issue. During the prodromal phase of PD, participants may have more healthcare contacts due to non-motor and subtle motor symptoms, and therefore may be more likely to be screened for and diagnosed with diabetes.[70] This hypothesis is supported by a cohort study in men showing that the association between diabetes and PD was present only among those with short diabetes duration (<10 years) and without complications of diabetes, and that the diagnosis of diabetes tended to cluster around the diagnosis of PD.[70] Alternatively, another study reported that only diabetes with long duration ( $\geq 10$  years) was associated with higher PD risk,[120] and two other studies showed stronger associations in those with diabetic complications than in those without.[121, 122] Nevertheless, the importance of taking into account use of medical care is strongly supported by a case-control in Medicare beneficiaries showing that not taking into account in the analyses use of medical care tended to bias the associations between medical conditions and PD.[123] For instance, smoking and smoking-related cancers were positively associated with PD without adjustment for use of care, but became inversely associated with PD, as expected, with adjustment for use of care; for constipation, depression, anosmia, and RBD, associations were positive but decreased after adjustment for use of care by 8% to 58%, depending on the condition. Although the authors did not specifically examine diabetes, these findings support the importance of examining the role of use of care in analyses of the association between comorbidities and PD.

Last, we cannot exclude that reverse causation may also be involved, especially



in those studies with short follow-up. There is some evidence showing that dopaminergic neuronal loss alters glycaemic control[70] and mitochondrial function, thus increasing the risk of insulin resistance and subsequent development of T2DM.[124] Larger cohorts with longer follow-ups are needed to assess whether reverse causation explain prior associations between diabetes and PD. In addition, analyses of trajectories would help to understand the temporal relation between diabetes and PD diagnosis.

*Given inconsistent results from previous studies and the methodological issues described above, and because of the relation of diabetes with PA and adiposity, we selected diabetes as our third exposure of interest.*

**Table 4: Cohort studies on the association between diabetes and Parkinson’s disease incidence.**

Study (Year)	Data source (country)	Follow-up (mean, years)	Participants (% women)	PD Definition	Type of diabetes; Definition	Diabetes status	Cases (n)	Baseline Analysis RR (95% CI)	Adjustments
Hu <i>et al.</i> [125] (2007)	Six independent cross-sectional population surveys (Finland)	1972-2002 (18)	51,552 (51.2)	National social insurance Institution’s register on reimbursement for PD	T2DM; Self-report, national hospital discharge register data, and national social insurance Institution’s register on reimbursement for antidiabetic drugs	<b>Men</b>			Age, Study year, BMI, SBP, Cholesterol, Education, PA, Smoking, Coffee and tea consumption, Alcohol consumption.
							<b>324</b>		
						<b>No</b>	311	1.00 (Ref.)	
						<b>Yes</b>	13	1.78 (1.01-3.12)	
						<b>Women</b>			
							<b>309</b>		
						<b>No</b>	298	1.00 (Ref.)	
						<b>Yes</b>	11	1.91 (1.04-3.52)	
						<b>Men and women combined</b>			
							<b>633</b>		
<b>No</b>	609	1.00 (Ref.)							
<b>Yes</b>	24	1.83 (1.21-2.76)							
Simon <i>et al.</i> [126] (2007)	Health Professionals Follow-up Study (HPFS, US)	1986-2000 (12.6)	50,833 (0)	Self-report confirmed by neurologist/ internists, medical records, and validation	T2DM/T1DM not distinguished; Self-report	<b>Men and women combined</b>			Age; Smoking, Caffeine intake, Total calorie intake, Alcohol consumption, PA, NSAID use, BMI.
						<b>530</b>			
	<b>No</b>	493	1.00 (Ref.)						
	<b>Yes</b>	37	1.04 (0.74-1.46)						
Driver <i>et al.</i> [70] (2008)	Physicians’ Health study (US)	1982-2006 (NR)	21,841 (0)	Self-report, death certificates, and validation	T2DM; Self-report	<b>Men</b>			Age, smoking, alcohol use, BMI, PA, hypertension, cholesterol.
							<b>556</b>		
						<b>No</b>	498	1.00 (Ref.)	
						<b>Yes</b>	58	1.34 (1.01-1.77)	
Palacios <i>et al.</i> [97] (2011)	Cancer Prevention study II	1992-2005 (NR)	147,096 (54.4)	Self-report confirmed by neurologist/	T2DM/T1DM not distinguished; Self-report	<b>Men</b>			Age, smoking, alcohol intake, calories, dairy
							<b>362</b>		
						<b>No</b>	338	1.00 (Ref.)	

Study (Year)	Data source (country)	Follow-up (mean, years)	Participants (% women)	PD Definition	Type of diabetes; Definition	Diabetes status	Cases (n)	Baseline Analysis RR (95% CI)	Adjustments
	Nutrition Cohort (US)			internists/general practitioner, medical records, and validation		Yes	24	0.87 (0.57-1.33)	intake, pesticides exposure, PA and education.
						<b>Women</b>			
							<b>202</b>		
						No	191	1.00 (Ref.)	
						Yes	11	0.90 (0.48-1.66)	
						<b>Men and women combined</b>			
							<b>564</b>		
						No	529	1.00 (Ref.)	
Yes	35	0.88 (0.62-1.25)							
Xu <i>et al.</i> [120] (2011)	NIH-AARP Diet and Health Study Cohort (US)	1995-2006 (NR)	288,662 (40.4)	Self-report, medical records, and validation	T2DM/T1DM not distinguished; Self-report	<b>Men and women combined</b>			Age, sex, race, education, smoking, coffee, BMI, PA.
							<b>1,565</b>		
						No	1,393	1.00 (Ref.)	
						Yes	172	1.41 (1.20-1.66)	
Sun <i>et al.</i> [127] (2012)	Taiwan's national Health Insurance database (Taiwan)	2000-2008 (NR)	1,075,604 (NR)	Outpatients' claims or hospitalization records	T2DM/T1DM not distinguished; NHI claims dataset	<b>Men</b>			Age, sex, geographic area, urbanization status, hypertension, hyperlipidemia, and cardiovascular disease.
							<b>1,098</b>		
						No	387	1.00 (Ref.)	
						Yes	711	1.51 (1.44-1.57)	
						<b>Women</b>			
							<b>1,324</b>		
						No	422	1.00 (Ref.)	
						Yes	902	1.70 (1.63-1.77)	
						<b>Men and women combined</b>			
	<b>2,422</b>								
No	809	1.00 (Ref.)							
Yes	1,613	1.61 (1.56-1.66)							

Study (Year)	Data source (country)	Follow-up (mean, years)	Participants (% women)	PD Definition	Type of diabetes; Definition	Diabetes status	Cases (n)	Baseline Analysis RR (95% CI)	Adjustments
Pupillo <i>et al.</i> [128] (2016)	Italian longitudinal patient database collected from GP (HSD, Italy)	2002-2013 (NR)	923,356	Database of general practitioners	T2DM/T1DM not distinguished; Database of general practitioners	<b>Men and women combined</b>			NR
						<b>No</b>	-	1.00 (Ref.)	
						<b>Yes</b>	361	1.32 (1.18-1.48)	
Yang <i>et al.</i> [119] (2017)	National Health Insurance Research Database (NHIRD, Taiwan)	2000-2011 (7.3)	145,176 (46.6)	Records of ambulatory and inpatient care data from NHIRD	T2DM; Outpatient and inpatient care data from NHIRD	<b>Men</b>			Age, gender, insurance premium, urbanization level, residential area, type of occupation, comorbidity, flunarizine, metoclopramide and zolpidem use, and outpatients claims.
							<b>907</b>		
						<b>No</b>	639	1.00 (Ref.)	
						<b>Yes</b>	268	1.12 (0.97-1.30)	
						<b>Women</b>			
							<b>875</b>		
						<b>No</b>	593	1.00 (Ref.)	
						<b>Yes</b>	282	1.29 (1.12-1.49)	
						<b>Men and women combined</b>			
	<b>1,782</b>								
<b>No</b>	1,232	1.00 (Ref.)							
<b>Yes</b>	550	1.19 (1.08-1.32)							
De Pablo-Fernandez <i>et al.</i> [121] (2018)	English national Hospital Episode Statistics and mortality data (HES, England)	1999-2011 (NR)	8,190,323 (64.3)	HES records and death certificates	T2DM; HES records	<b>Men and women combined</b>			Age, sex, year of cohort entry, region of residence, patients' deprivation score
						<b>No</b>	-	1.00 (Ref.)	
						<b>Yes</b>	-	1.32 (1.29-1.35)	
Kizza <i>et al.</i> [129] (2019)	China Kadoorie Biobank study (CKB, China)	2008-NR (9)	503,497 (59.2)	Health insurance records from hospitalization	T2DM/T1DM not distinguished; Self-reported and clinical examination	<b>Men and women combined</b>			Age, region, income, education, occupation,
							<b>521</b>		
						<b>No</b>	481	1.00 (Ref.)	
<b>Yes</b>	40	0.93 (0.67-1.29)							

Study (Year)	Data source (country)	Follow-up (mean, years)	Participants (% women)	PD Definition	Type of diabetes; Definition	Diabetes status	Cases (n)	Baseline Analysis RR (95% CI)	Adjustments
									alcohol consumption, PA, sex, smoking, BMI, SBP and DBP.
Kummer <i>et al.</i> [130] (2019)	Medicare insurance datasets	2008-2015 (5.2)	1,035,536 (59.2)	≥2 outpatient Medicare claims	T2DM/T1DM not distinguished; Algorithms with Medicare claims	<b>Men and women combined</b>			Age, sex, race/ethnicity, Charlson comorbidities
							<b>15,531</b>		
						<b>No</b>	11,245	1.00 (Ref.)	
						<b>Yes</b>	4,286	1.17 (1.11-1.24)	
Rhee <i>et al.</i> [118] (2020)	National Health Insurance (NHI, South Korea)	2009-2016 (6.3)	8,443,351 (50.5)	NHIS registration program for rare diseases	T2DM/T1DM not distinguished; Health screening	<b>Men and women combined</b>			Age, sex, BMI, smoking, drinking, PA, competing risk of death.
							<b>31,577</b>		
						<b>No</b>	23,299	1.00 (Ref.)	
						<b>Yes</b>	8,278	1.34 (1.30-1.37)	
Sanchez-Gomez <i>et al.</i> [131] (2021)	Information system for Research in Primary Care (SIDIAP, Spain)	2006-2018 (7.3)	3,104,460 (51.8)	SIDIAP data	T2DM; SIDIAP data	<b>Men</b>			Age, sex, BMI, smoking status and socioeconomic status.
							<b>6,562</b>		
						<b>No</b>	5,552	1.00 (Ref.)	
						<b>Yes</b>	1,010	1.11 (1.04-1.20)	
						<b>Women</b>			
							<b>5,975</b>		
						<b>No</b>	5,196	1.00 (Ref.)	
						<b>Yes</b>	779	1.27 (1.18-1.38)	
						<b>Men and women combined</b>			
	<b>12,537</b>								
<b>No</b>	10,748	1.00 (Ref.)							
<b>Yes</b>	1,789	1.19 (1.13-1.25)							

NR, Not reported; T2DM, diabetes mellitus type 2; T1DM, diabetes mellitus type 1; BMI, body mass index; PA, physical activity; GP, general practitioner; SBP, systolic blood pressure; DBP, diastolic blood pressure.

## **2. OBJECTIVES**



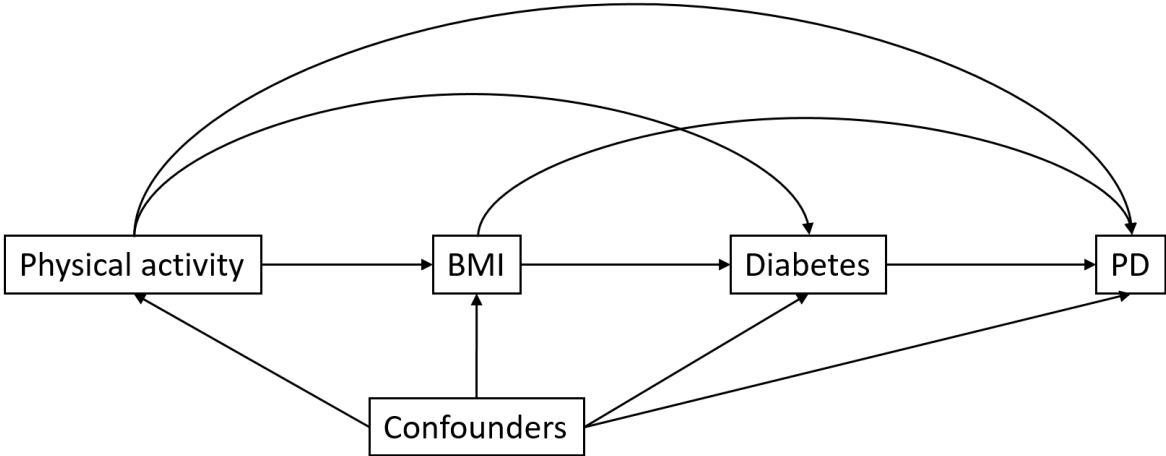
PD is the fastest growing neurological disorder in terms of prevalence, disability, and death worldwide. At the present time, PD has no cure and identifying preventative strategies allowing to prevent or delay the onset of PD is an urgent public health need.

Epidemiologic studies may help identify modifiable exposures and contribute to the definition of preventative strategies that could be tested in appropriate populations.[42] However, the long prodromal phase of PD is a major challenge for epidemiologic studies due to the potential for reverse causation. Large studies with a long follow-up and appropriate statistical methods are needed to address this issue.

In addition, because PD is more frequent in men than women, women represent an understudied group in PD research, and specific studies in women are needed.[132]

In this thesis, the aim was to examine the relation between potentially modifiable health behaviours (**first and second objectives**: physical activity, adiposity) and comorbidities (**third objective**: diabetes) and the incidence of PD in ~100,000 women from the French E3N cohort study followed since 1990, while accounting for the potential for reverse causation. Figure 1 shows the hypothetical causal relation of these three exposures in relation to PD.

**Figure 1: Hypothetical causal relation between three exposures and PD.**







### **3. PARTICIPANTS AND METHODS**



### **3.1. E3N cohort study**

E3N is an ongoing prospective French cohort study of 98,995 women, born between 1925-1950 and recruited in 1990, who were affiliated with a French national health insurance plan that covers mostly teachers (*Mutuelle Générale de l'Éducation Nationale*, MGEN).[133] At baseline, participants who gave written informed consent (approximately 20% of invited women) completed a self-administered questionnaire on lifestyle, environment, and medical history (Q1). The informed consent was in compliance with the rules of the French National Commission for Data Protection and Privacy, from which approval was obtained. Follow-up questionnaires have been sent every 2-3 years thereafter. At the present date, 11 waves of data collection are available (last in 2014, Q11), with an average response rate of 80% at each questionnaire, and a very low rate of participants lost to follow-up (<3%).

In addition, since January 2004, women have been passively followed through databases from MGEN that include drug claims and medical consultations (updated every trimester). These databases include detailed information about drugs delivered to the participants until December 2018 and information on medical contacts (including date and specialty of the physicians). Causes of death are also available for women who died during the follow-up.

### **3.2. Ascertainment of Parkinson's disease**

Potential PD patients were identified using two data sources.[134] The first source consisted of PD self-reports by E3N participants in the follow-up E3N questionnaires. The second source was MGEN drug claims databases that were used to identify women who were reimbursed for any drug that can be used to treat PD (Anatomical Therapeutic Chemical [ATC] code N04)[135] between January 2004 and December 2018. To exclude patients with drug-induced parkinsonism, we identified women who used neuroleptics (ATC code N05A except lithium) over the same period. In addition,

we searched death certificates for PD codes (International classification of disease: ICD-9-332.0, ICD-10-G20).

Starting in 2010, we followed a three-phase approach. At stage 1, women who self-reported PD in any questionnaire or had used antiparkinsonian drugs (APDs) and were still alive were contacted by mail using a standardized questionnaire, in order to confirm the diagnosis of PD or parkinsonism, and to obtain further information: reasons for prescription of APDs and date of diagnosis; contact information for their treating neurologist or general practitioner; hospitalizations. Women who responded that they were not treated for PD or parkinsonism were considered as free of PD thereafter.

At stage 2, we sent a standardized questionnaire to neurologists (or general practitioners if not available) of women who responded that they were treated for PD or parkinsonism at stage 1, did not respond at stage 1, or were deceased, in order to obtain detailed clinical information: date of onset and diagnosis, cardinal motor signs, other symptoms (e.g., dementia, dysautonomia, supranuclear palsy, falls), use of neuroleptics, treatment, responsiveness to treatment, and diagnosis. We also requested copies of outpatient visits and hospitalizations reports, and results of brain imaging exams.

At stage 3, all the information that were collected were reviewed by an adjudication committee, including two movement disorders specialists, one epidemiologist, and one neurologist-epidemiologist.

We assigned PD diagnoses by consensus according to the following classification: definite PD, probable PD, possible PD, no PD. We retained a diagnosis of definite PD when there were at least 2 out of 4 cardinal signs (rest tremor, rigidity, bradykinesia, impaired postural reflexes), with a positive and significant response to treatment at a sufficient dose, without prominent or early signs of more extensive nervous system involvement and drug-induced parkinsonism.[136] When the same criteria were fulfilled but there was insufficient information regarding treatment

response, we attributed a diagnosis of probable PD. Women were included in the possible PD group when the clinical description was suggestive of PD, but there was insufficient information in the medical records to assign a definite or probable PD diagnosis. The remaining women were included in the no PD group which included a variety of conditions (e.g., other causes of parkinsonism such as dementia with Lewy bodies, multiple system atrophy, or progressive supranuclear palsy, vascular, or drug-induced; restless leg syndrome; essential tremor). Regarding the presence of dementia, we considered dementia to be an exclusion criteria if it was present before or at the time of diagnosis of parkinsonism. We did not consider dementia that developed early during the course of the disease as an exclusion criteria for PD, if all other features were compatible with this diagnosis and if there were no other red flags. The same criteria were used throughout the validation process for all cases.

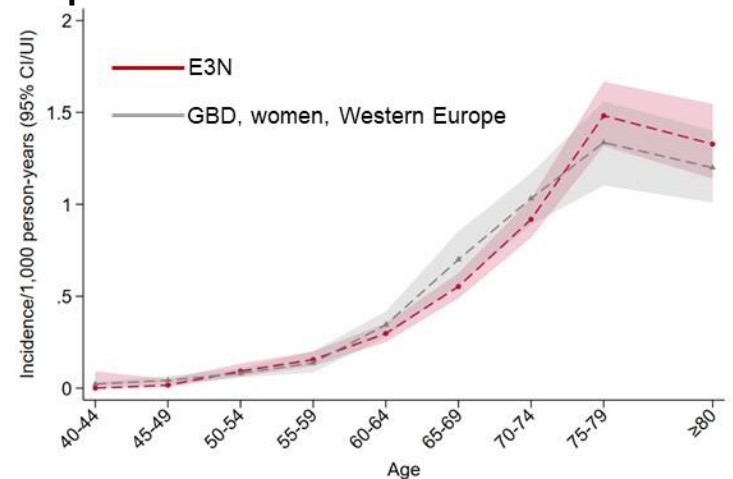
When no medical records were available, we predicted PD status using an algorithm based on drug claims databases that was previously developed and validated against a clinical PD diagnosis.[137] This algorithm allows to estimate the probability that a person is treated for PD based on 12 variables that characterize the annual pattern of APD use, the proportion of time treated, and the number of neurologist/general practitioner's visits. We used a modified version of this algorithm that is characterized by a high area under the curve (AUC=0.957, 95% CI=0.949-0.964), and allows to predict PD status with 94% sensitivity and 88% specificity for a probability cut-off of 0.235 in E3N.[134] Women who had used APDs, who did not use neuroleptics in the 10 years preceding the first prescription of APDs with a probability predicted by the algorithm above this cut-off were considered as having PD. Women with a lower probability or who used neuroleptics in the 10 years preceding the first prescription of APDs were considered as not having PD.

The year of PD diagnosis was defined as the year of diagnosis available in medical records or, in decreasing order of priority, self-reported year of diagnosis, year of first APDs use, and year of the first questionnaire where PD was self-reported. The

proportion of PD diagnoses based on medical records and the algorithm was 62% and 38%, respectively. Definite and probable PD patients were retained in the analyses, while patients with possible PD were excluded.

A previous study showed that PD incidence rates in E3N were in agreement with those in women from Western Europe according to the Global Burden of Disease over the same period; however, both curves showed an inflexion after 80 years, thus suggesting that PD may be under-diagnosed in the oldest age groups. (Figure 2).[134, 138]

**Figure 2: Incidence rates of Parkinson’s disease in women from E3N cohort study and from Western Europe.**



Source: Canonico *et al.*[134] The figure shows the incidence rates per age groups in women from E3N (red) and Western Europe according to the Global Burden of disease project (grey) between 1992 and 2018. Dashed lines represent the point estimates and shaded areas the 95% CI.

### 3.3. Covariates

Self-reported characteristics were collected through follow-up questionnaires (Table 5).

At baseline (Q1, 1990), the following characteristics were available:

- Birth date;
- Education level (<high school, ≥high school): previous studies suggested that higher education may be association with higher PD risk;[139]
- Region of residence (French commune): we categorized French

communes into rural or urban according to the French *Institut national de la statistique et des études économiques* (INSEE)[140] as a proxy for environmental exposure to pesticides; previous studies showed that pesticide exposure is associated with higher PD risk;[141]

- Age at menarche ( $\leq 11$  years, [12-13 years],  $\geq 14$  years): analyses in E3N show that age at menarche at  $\leq 11$  or  $\geq 14$  years was associated with higher PD incidence, compared to age at menarche at 12-13 years;[142]
- Parity (nulliparous, one child, two children,  $\geq$  three children): analyses in E3N show that women with two or more children had higher risk of PD.[142]

Time-varying covariates included the following characteristics:

- Menopausal status was collected in each questionnaire until Q9 (2008) and categorized into four categories (premenopausal, natural, artificial, unknown type of menopause): analyses in E3N show that menopausal status was not associated with PD, but artificial menopause was associated with higher PD risk compared with natural menopause.[142]
- Information on cigarette smoking (never, ex, current) was collected from Q1 to Q8: there is a robust inverse association between cigarette smoking and PD,[141] and MR studies are in favor of a causal association.[143] Despite years of research, the mechanisms underlying the relation between smoking and PD are not well understood. One study found inhibition of the enzyme monoamine oxidase B (MAO-B) in the brains of smokers.[144] Another study showed that among five compounds in cigarette smoke (anabasine, cotinineone, nicotine, hydroquinone, nor nicotine), nicotine and hydroquinone inhibited the aggregation of  $\alpha$ -synuclein.[145] Finally, an interaction between smoking and the *HLA-DRB1* gene was replicated in two large studies; these findings suggest that smoking may reduce the affinity between the proteins encoded by



the *HLA-DRB1* gene and  $\alpha$ -synuclein protein and therefore reduce an immune response that is involved in the etiology of PD.[146, 147]

- Diet was assessed using a validated 208-item semi-quantitative questionnaire in 1993 (Q3). The questionnaire was validated using twelve 24-hour recalls carried out monthly, with good 1-year reproducibility.[148] Intakes and frequencies of 208 food items were converted into total intakes of energy and nutrients using food composition databases from the French national databases.[149] We used scores of adherence to the prudent diet obtained by principal component analysis as a proxy for a Mediterranean diet.[150, 151] We also included total intake of caffeine (mg per day) and total intake of lactose (g per day), and categorized them in quartiles. Previous studies showed that caffeine intake[141] and Mediterranean diet[152, 153] were associated with reduced risk of PD. Alternatively, dairy intake has been shown to be associated with higher risk of PD,[59] and a MR study is in favor of a causal relationship.[154]

For all covariates, missing values were coded as specific categories to retain the same number of participants in all analyses.

Detailed information regarding the assessment of the exposures of interest in this thesis are provided in the corresponding chapters.

### **3.4. Statistical Analysis**

Statistical analyses were conducted using SAS 9.4 (SAS Institute Inc, Cary, NC), R (R Foundation for Statistical Computing, Vienna, Austria), or Stata 15.1 (StataCorp, College Station, TX). Statistical analyses are described for each exposure of interest in the corresponding chapter.

**Table 5: Assessment of self-reported characteristics in E3N questionnaires over the follow-up.**

Characteristics	Questionnaires											Variable recorded as	
	Q1 (1990)	Q2 (1992)	Q3 (1993)	Q4 (1995)	Q5 (1997)	Q6 (2000)	Q7 (2002)	Q8 (2005)	Q9 (2008)	Q10 (2011)	Q11 (2014)		
Birth date	✓												Continuous.
Education	✓												Categorical: <high School; ≥high School.
Residence	✓												Categorical: urban; rural.
Age at menarche	✓												Categorical with 3 categories: ≤11y; [12-13y]; ≥14y.
Parity	✓												Categorical with 4 categories: nulliparous; 1 child; 2 children; ≥3 children.
Menopausal status	✓	✓	✓	✓	✓	✓	✓	✓	✓				Categorical with 4 categories: premenopausal; natural; artificial; unknown type of menopause.
Weight	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	Continuous.
Height	✓			✓		✓	✓	✓	✓	✓	✓	✓	Continuous (standardized).
BMI	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	Categorical with 4 categories: <18.5; [18.5-25.0]; [25.0-30.0]; ≥30.0 kg/m <sup>2</sup> .
Smoking status	✓	✓	✓	✓	✓	✓	✓	✓					Categorical: never; ex; current.
Total caloric intake			✓										Categorical: quartiles.
Mediterranean diet			✓										Categorical: quartiles.
Total intake of caffeine			✓										Categorical: quartiles.
Total intake of lactose			✓										Categorical: quartiles.
Physical Activity	✓		✓		✓		✓	✓				✓	Categorical: quartiles.
Diabetes	✓	✓	✓	✓	✓	✓	✓						Dichotomous: Yes; No.
Hypercholesterolemia	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	Dichotomous: Yes; No.
Hypertension	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	Dichotomous: Yes; No.
Cardiovascular diseases	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	Dichotomous: Yes; No.

y, years.



## **4. Objective 1: Association between physical activity and Parkinson's disease**



## 4.1. Introduction

Growing evidence from clinical trials and observational studies support the beneficial effects of exercise on motor and non-motor PD symptoms,[25, 34, 71, 72] and exercise is considered as the most effective non-pharmacological intervention in PD.[26, 27, 42]

These observations raise the question about the potential of physical activity to prevent or delay the onset of PD. Previous cohort studies examined the association between physical activity and PD incidence (§1.3.1, Table 2). A meta-analysis showed that physical activity at baseline was associated with lower PD incidence, in analyses with men and women combined, with a pooled HR of 0.79 (95% CI=0.68-0.91) for those with higher levels of total physical activity compared to those with lowest levels (2,192 PD cases from 8 studies).[75] In sex-stratified analyses, higher physical activity level was inversely associated with PD risk in men but not in women (§1.3.1.2).[75] Determining whether physical activity plays a role in women is important for developing appropriate interventions.[155]

As previously discussed (§1.3.1), reverse causation represent a key issue for the interpretation of the results of studies on the relation between physical activity and PD. PD patients may reduce their physical activity during the prodromal phase as a consequence of prodromal symptoms. To address this issues, some previous cohort studies excluded cases over the first 4-10 years of the follow-up.[69, 83-86] While findings were generally consistent with an inverse association, only one study reported a significant association.[86] Until today, four cohorts had a follow-up up to 10 years,[83, 86-88] three had a follow-up between 10 to 15 years[69, 84] and only one cohort had a longer follow-up of more than 15 years.[85] However, in this last study analyses may lack statistical power due to small number of participants. Therefore, larger cohorts with longer follow-ups are needed to assess whether reverse causation contributes to the inverse association between physical activity and PD. In addition, none of the previous studies used repeated physical activity measures, whereas

analyses of physical activity trajectories prior to PD diagnosis would help understand the temporal relation between physical activity and PD.

Our aim was to examine the association between time-varying measures of physical activity and PD incidence in women from the E3N cohort study over 29 years of follow-up, while addressing the potential for reverse causation. We also used a nested case-control design to compare physical activity trajectories in PD patients prior to diagnosis and matched controls.

## **4.2. Methods**

### **4.2.1. Assessment of physical activity**

Six questionnaires included physical activity related questions (1990-Q1, 1993-Q3, 1997-Q5, 2002-Q7, 2005-Q8, 2014-Q11; Table 6). Physical activity related to different recreational and household activities was assessed over the follow-up in different ways and units, and 11 types of activities were included in at least one questionnaire. MET values were attributed to each activity based on a compendium[156] and expert opinion[157] (Table 6). A MET is defined as the ratio of the work metabolic rate to a standard resting metabolic rate of 1.0, where 1 MET is considered a resting metabolic rate while sitting. For each activity, their respective METs were multiplied by their frequency and duration to obtain a physical activity score (MET-h/week).

The baseline questionnaire (1990-Q1) included six closed-ended questions on recreational physical activity: usual distance walked daily (<500, 500-2000, >2000 meters), average number of flights of stairs climbed daily (0, 1-4, >5), weekly average time spent in light household activities (0, 1-4, 5-13,  $\geq 14$  hours), weekly average time spent in heavy household activities (0, 1-4,  $\geq 5$  hours), weekly average time spent in moderate recreational activity (e.g., light gardening, sports of moderate intensity) (0, 1-4, 5-13,  $\geq 14$  hours), and weekly average time spent in vigorous recreational activities (e.g., vigorous sports) (0, 1-4,  $\geq 5$  hours).[157]

In subsequent questionnaires (1993-Q3, 1997-Q5, 2002-Q7, 2014-Q11), physical activity related questions were derived from the short questionnaire used in the European Prospective Investigation into Cancer and Nutrition (EPIC) study to assess physical activity.[158] Analyses of reproducibility and validity show that this questionnaire is more suitable for ranking subjects on overall physical activity than for estimating absolute energy expenditure.[159] A study on the validity of the questionnaire against accelerometers showed that there was fair agreement ( $P=0.29$ ) between the EPIC questionnaire and accelerometers measurements in ranking participants for total physical activity level.[160]

The questions assessed the duration (hours/week) that participants spent walking (including walking to work, shopping, and leisure time), cycling (including cycling to work, shopping, and leisure time), and engaging in sports during two typical weeks over the past year, one in summer and one in winter. These questionnaires used open-ended questions, allowing women to provide more detailed information on frequency and duration of each activity than at Q1. In addition, the number of hours practicing each activity was ascertained in the winter and summer, and durations were averaged over these two periods. The questionnaire from 2005-Q8 was similar to Q1 but asked open-ended rather than closed-ended questions.

There were five questions in common for Q3, Q5, and Q7; Q5 and Q7 included the same questions; Q8 had four matching questions with Q1; Q11 had four identical questions with Q5 and Q7 (Table 6).

#### **4.2.2. Nested case-control study**

In order to examine trajectories of physical activity prior to PD diagnosis in cases and matched controls, a nested case-control study was set-up. Each incident PD case was individually matched to 20 controls using incidence density sampling.[161, 162] To be selected as controls, women had to be alive, at risk of PD at the date of diagnosis of



the matched case (index date,  $T_0$ ), and have the same age at  $T_0$ .

Among 25,200 women (1,200 cases, 24,000 controls), we excluded 45 women (four PD cases) for whom physical activity was missing at all visits before  $T_0$ , and 80 controls matched to these four PD cases, leading to a final sample of 25,075 women (1,196 PD cases, 23,879 controls); 1,156 cases were matched to 20 controls, 39 cases to 19 controls, and 1 case to 18 controls.

**Table 6: Assessment of self-reported recreational and household physical activities in E3N questionnaires over the follow-up.**

<b>Activities</b>	<b>Unit</b>	<b>MET</b>	<b>Q1 (1990)</b>	<b>Q3 (1993)</b>	<b>Q5 (1997)</b>	<b>Q7 (2002)</b>	<b>Q8 (2005)</b>	<b>Q11 (2014)</b>	<b>Total number of measures</b>
Heavy cleaning	Hours per week	4	✓				✓		2
Light cleaning	Hours per week	3	✓				✓		2
Household (cooking, cleaning...)	Hours per week	3			✓	✓		✓	3
Stairs Climbing	Number of floors climbed per day	0.067	✓	✓					2
Walking 1	Meters walked per day	0.00075	✓						1
Walking 2	Hours per week	3		✓	✓	✓	✓	✓	5
Vigorous recreational activities	Hours per week	7.5	✓				✓		2
Moderate recreational activities	Hours per week	5	✓				✓		2
Sport	Hours per week	6		✓	✓	✓		✓	4
Cycling	Hours per week	6		✓	✓	✓		✓	4
Gardening and bricolage	Hours per week	4.25		✓	✓	✓		✓	4

MET=Metabolic equivalent of task values.

## 4.3. Statistical Analysis

Physical activity was assessed using different measurement tools during the follow-up. When measurement tools of a risk factor (here physical activity) change over the follow-up in a cohort study, standard longitudinal approaches (e.g., mixed models) cannot be used and appropriate methods are needed to deal with this difficulty and to generate a time-varying latent physical activity variable that represents the same quantity of interest for all the participants who had at least one physical activity assessment during the follow-up (N=98,766).[163, 164]

### 4.3.1. Linear mixed models

Linear mixed models (LMM) allow to analyse change in longitudinal continuous markers. LMM include fixed and random effects (intercept and slope). Fixed effects ( $\beta_0, \beta_1, \beta_2$ ) correspond to population-level effects for the intercept, covariate  $X$ , and time. Normally distributed random effects ( $b_{0i}, b_{1i}$ ) represent correlated subject-specific effects with variance-covariance matrix  $B$ . The error term  $\epsilon_{it}$  is normally distributed with variance  $\sigma_\epsilon^2$ . An example of a LMM for the dependent variable  $Y_{it}$  measured in participant  $i$  at time  $t$  is:

$$Y_{it} = \beta_0 + \beta_1 time_{it} + \beta_2 X_{it} + b_{0i} + b_{1i} time_{it} + \epsilon_{it} \quad (1)$$
$$\epsilon_{it} \sim \mathcal{N}(0, \sigma_\epsilon^2),$$
$$b_i = (b_{0i}, b_{1i})^T \sim \mathcal{N}\left(0, B = \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{bmatrix}\right)$$

### 4.3.2. Latent process mixed models

Latent process mixed models (LPMMs) allow the longitudinal analysis of a possibly non-Gaussian risk factor in the presence of a change in the measurement tools over the follow-up. These methods focus on the key quantity of interest, which is not directly observed (latent variable), rather than on the type of observations of this quantity. In this setting, observations obtained using different measurement tools can be analyzed

longitudinally, provided that the tools measure the same quantity of interest. LPMMs represent an extension of linear mixed models. They are more flexible in terms of assumptions and allow the study of a single non-Gaussian longitudinal marker obtained using different measurement tools. This method allows to separate the *structural model* that describes the quantity of interest (latent process) according to time, from the *measurement model* that links the quantity of interest to the observations.

The structural model defines the latent process  $\Lambda_i(t)$  similarly to a standard linear mixed model (equation 1), but without the error term:

$$\Lambda_i(t) = \beta_0 + \beta_1 time_{it} + \beta_2 X_{it} + b_{0i} + b_{1i} time_{it} \quad (2)$$

To account for different types of longitudinal measures, a flexible nonlinear measurement model uses an outcome-specific link function to model the relation between the latent process  $\Lambda_i(t)$  and the observed value of the outcome  $Y_{it}$  at measurement  $time_{it}$ :

$$Y_{it} = H(\tilde{Y}_{it}; \eta) = H(\Lambda_i(time_{it}) + \varepsilon_{it}; \eta) \quad (3)$$

where  $\varepsilon_{it}$  are independent Gaussian measurement errors,  $H$  is a parameterized link function, and  $\tilde{Y}_{it}$  characterizes the noisy latent process at  $time_{it}$ .

We used a transformation based on quadratic I-splines with  $m$  knots:

$$H^{-1}(Y_{ij}; \eta) = \eta_0 + \sum_{l=1}^{m+1} \eta_l^2 B_l^I(Y_{it}) \quad (4)$$

where  $(B_1^I, \dots, B_{m+1}^I)$  is the bases of I-splines.

### 4.3.3. Latent process mixed models for multivariate longitudinal markers

The *structural model* and the *measurement model* can be computed for multiple longitudinal markers, instead of a single one, that measure the same quantity of

interest. In this case, the underlying latent process  $\Lambda_i(t)$  defined above generates  $K$  longitudinal markers instead of only one (as in the previous section 4.3.2, equation 2). Here, the latent process represents the unmeasured common factor underlying the observations of different markers obtained through different tools from different subjects. The structural model for  $\Lambda_i(t)$  according to time and covariates is exactly the same as in the previous section, but the measurement model is extended to the multivariate setting in order to take into account the specific relationship between the underlying latent process and each longitudinal marker.

In the measurement model, the measurement error is accounted for through an intermediate variable  $\tilde{Y}_{kit}$ :

$$\tilde{Y}_{kit} = \Lambda_i(t_{kit}) + X_{Yi}(t_{kit})^T \gamma_k + b_{ki} + \varepsilon_{kit} \quad (5)$$

where  $X_{Yi}(t_{kit})^T \gamma_k$  are covariates with a marker-specific effect  $\gamma_k$  (called contrast),  $b_{ki}$  is a random intercept, and  $\varepsilon_{kit}$  represents random measurement error.

The relationship between each longitudinal marker and the underlying quantity of interest is modelled through the marker-specific link function  $H_k$ :

$$Y_{kit} = H_k(\tilde{Y}_{kit}; \eta_k) \quad (6)$$

where  $H_k$  can be defined using l-splines.

In this model, the intercept is constrained to be 0 and the variance of the random intercept is constrained to be equal to 1. Hence, a random intercept is required while no mean intercept is allowed in the structural model.

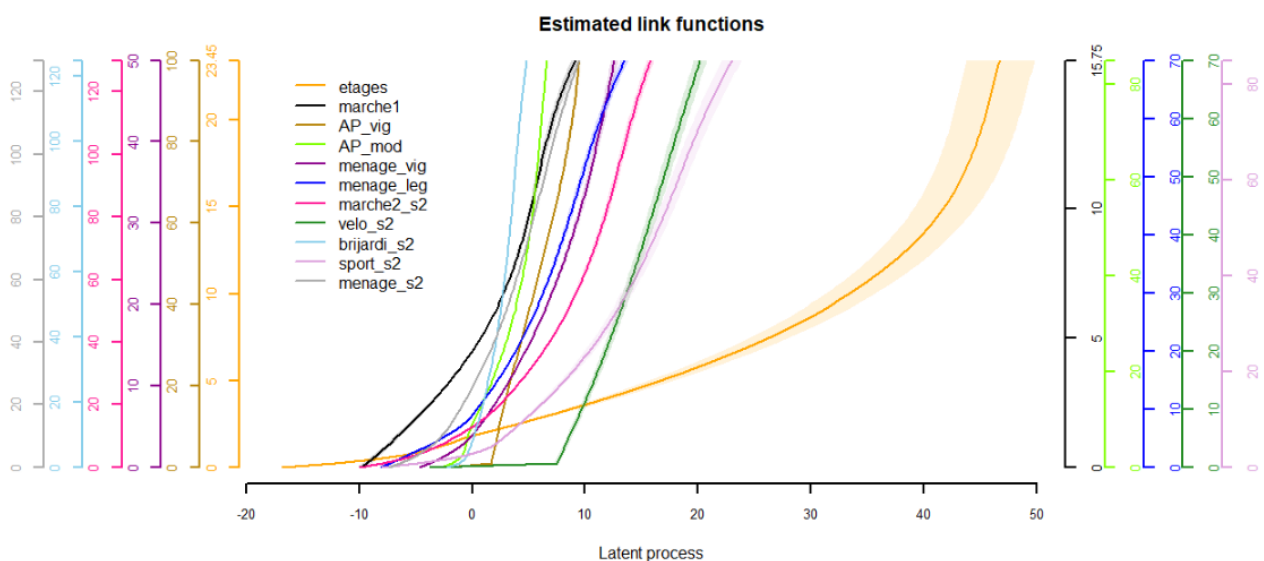
#### **4.3.4. Construction of the latent physical activity variable in E3N**

Our aim was to define a latent physical variable for all the participants who had at least one measure of physical activity throughout the follow-up.

In order to link repeated individual physical activity measures (in MET-h/week) obtained for different activities from different questionnaires to their common underlying latent process and to correct the departure of each activity from a normal distribution, we constructed separate LPMMs for each activity. We selected the best activity-specific parameterized link function (among quadratic I-splines link functions with 2-, 3-, 4-, and 5-knots) according to the Akaike information criterion (AIC) in activity-specific models (Figure 3). We then included the 11 activities (with the shapes of activity-specific link functions previously determined) in the same LPMM for multivariate longitudinal outcomes, in order to model the latent physical activity trajectory according to age. Age was centered at its mean at Q1 (49 years) and change over age was approximated by natural cubic splines with two internal knots (lower AIC). The within-participant correlation was captured by independent random intercept and slopes on the three functions of age.

We finally obtained subject-specific latent physical activity predictions for each subject at each visit based on the best fitting LPMM, and these predictions were used in subsequent analyses described below (analyses of trajectories, Cox models).

**Figure 3: Relationship between the 11 activities (METs-h/week) and the latent process.**



### **4.3.5. Trajectories of latent physical activity**

Trajectories of latent physical activity from the index time  $T_0$  to the beginning of the study (retrospective time scale) were examined in cases and controls over 29 years of follow-up using a linear mixed model.

The latent physical activity trajectory was modeled as a function of retrospective time and time squared, and was adjusted for PD status, age at  $T_0$ , and two-way interactions of time with PD status and age; it was further adjusted for parity, age at menarche, residence, and time-varying smoking and menopausal status. The within participant correlation was captured by correlated random intercept and slopes on time and time squared.

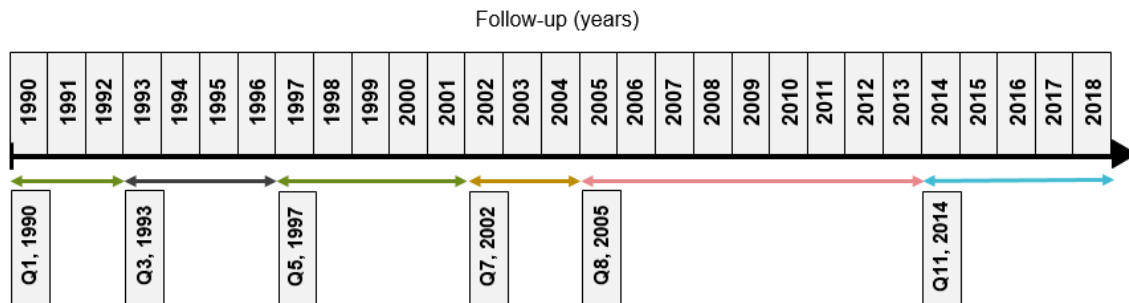
### **4.3.6. Survival Analysis**

We used time-varying Cox's proportional hazards models to examine the association between physical activity and the incidence of PD, and to estimate HR, their 95% CI, and two-tailed  $P$ -values ( $\alpha=0.05$ ). Participants were censored at the date of PD diagnosis, or at the end of follow-up (maximum date at the time of the last questionnaire or last drug reimbursement). As age is a strong determinant of PD, we used age as the time scale.[165]

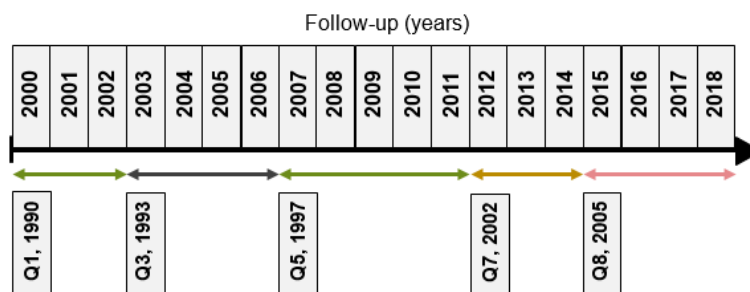
To address the potential for reverse causation, we included a lag of increasing duration (5, 10, 15, and 20 years) between time-varying variables (including latent physical activity) and PD incidence. Given results of the analyses of trajectories described above, we used a 10-year lag in our main analysis (Figure 4): latent physical activity was lagged by 10 years and we started the follow-up in 2000 (i.e., 10 years after the baseline assessment), so that participants who developed PD before 2000 (prevalent cases) were excluded. Women were followed from 2000 until PD diagnosis or end of follow-up (maximum of the date of the last questionnaire and last drug reimbursement). The same approach was used for other lags.

**Figure 4: Time-varying physical activity measures with no lag (A) and with 10-year lag (B).**

**A. Physical activity measures**



**B. Physical activity lagged by 10-years**



Panel A shows the actual physical activity measures from each questionnaire. Panel B shows the physical activity measures after lagging. Therefore, given the lag, PD incidence between 2000 and 2003 is a function of physical activity level assessed between 1990 and 1993 (from Q1); PD incidence between 2003 and 2006 is a function of physical activity level assessed between 1993 and 1996 (from Q3); PD incidence between 2007 and 2011 is a function of physical activity level assessed between 1997 and 2001 (from Q5); PD incidence between 2012 and 2014 is a function of physical activity level assessed between 2002 and 2004 (from Q7); PD incidence between 2016 and 2018 is a function of physical activity level assessed between 2005 and 2008 (from Q8).

Latent physical activity was included as time-varying quartiles in the models, and linear trends in HRs were tested through ordinal variables defined by the median of each quartile. We used restricted cubic splines to test for departures from linearity. Analyses were adjusted for baseline parity, age at menarche, and residence, and time-varying smoking and menopausal status. Missing values were coded as specific categories to retain the same number of participants in all analyses.

We performed several sets of sensitivity analyses:

- We constructed an alternative latent physical activity variable by excluding the baseline physical activity questions (1990-Q1) that were less precise than subsequent



measures.

- We constructed an alternative latent physical activity variable by excluding one activity (climbing stairs) that showed a different relation with the latent process compared to other activities (Figure 3).
- In order to examine whether our findings may have been confounded by diet, we performed analyses based on physical activity assessed at Q5 (1997) and adjusted for dietary exposures assessed earlier (1993-Q3; adherence to prudent diet, caffeine and lactose intake). Analyses with a lag longer than 10 years were not possible due to an insufficient number of cases.
- We performed analyses stratified by median age (72.8 years) in order to examine whether associations were similar for PD patients who developed PD before or after that age.

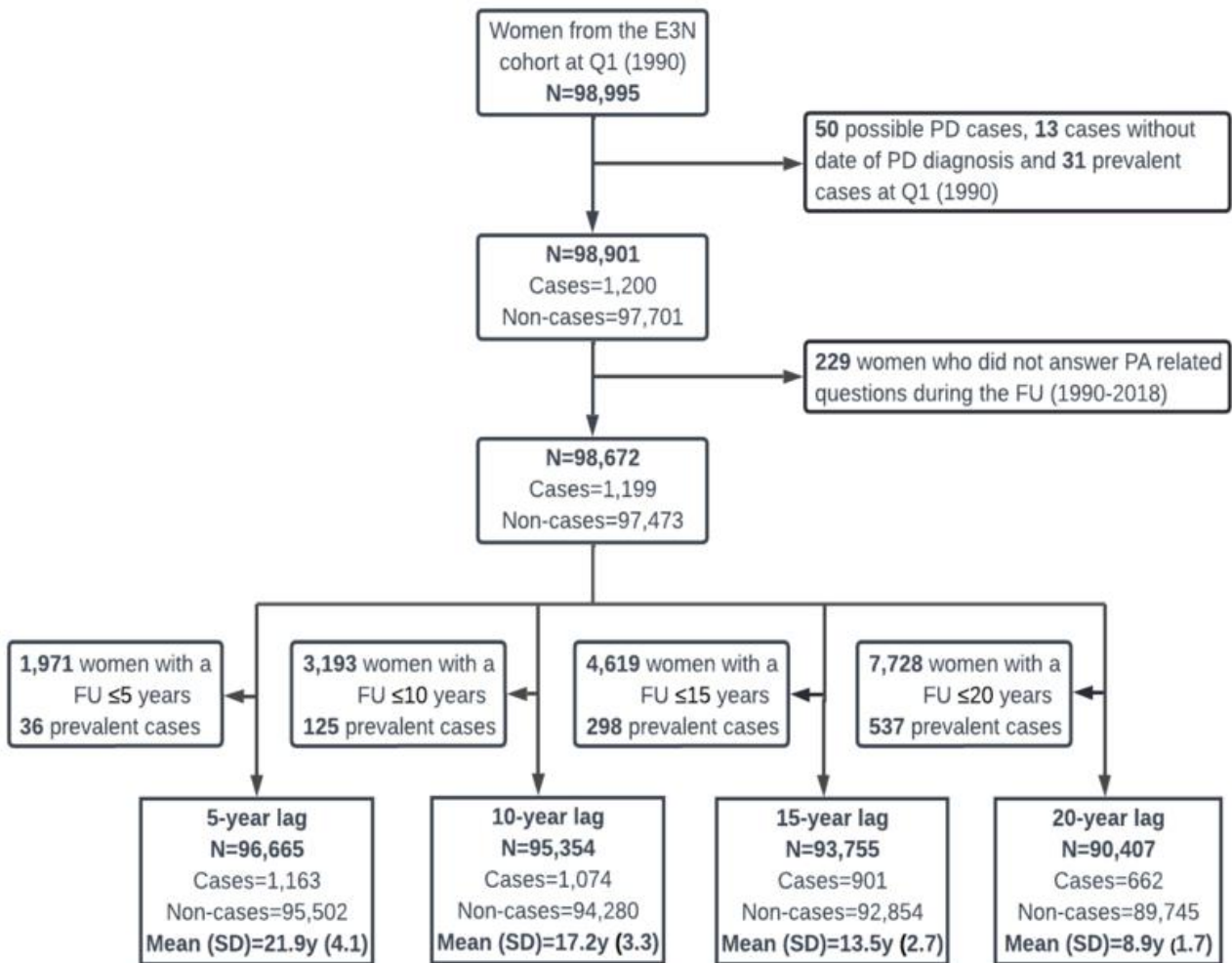
## **4.4. Results**

Figure 5 shows a flowchart for inclusion of participants into the study. We excluded 50 possible PD cases, 13 cases without a diagnosis date, 31 prevalent cases at Q1 and 229 women who did not answer physical activity questions at any questionnaire. The number of participants available for survival analyses decreased with increasing lags. For instance, for our main survival analysis (10-year lag), we further excluded 3,193 women whose follow-up ended before 2000 and 125 prevalent PD patients, leaving 95,354 women followed for 19 years (mean=17.2, SD=3.3) of whom 1,074 developed PD. Compared to previous studies, our study included the largest number of PD patients and had the longest follow-up (Table 2).

Table 7 describes baseline (1990-Q1) participant's characteristics. Mean age was 49.3 years (SD=6.6) and mean physical activity level was 45.3 (SD=30.1) METs-h/week. Women with later age of menarche,  $\geq 3$  children, and who lived in rural areas had higher physical activity levels than their counterparts. Women with incident PD were older, less frequently smokers, more frequently postmenopausal, and had later age at menarche

and more often  $\geq 3$  children than those who remained PD-free (Table 8).

**Figure 5: Flow chart for inclusion into the study on the relation between physical activity and Parkinson’s disease incidence in survival analyses.**



**Table 7: Participant's characteristics at baseline (1990-Q1).**

Baseline characteristics, n (%)	Total 96,665	Latent physical activity (LPA) at baseline <sup>a</sup>			
		Quartile 1 24,166 (25.0)	Quartile 2 24,166 (25.0)	Quartile 3 24,166 (25.0)	Quartile 4 24,167 (25.0)
<b>PA (METs-h/week), M (SD)<sup>b</sup></b>	45.3 (30.1)	27.0 (15.7)	37.3 (18.7)	46.0 (21.7)	70.8 (38.7)
<b>Age (y), M (SD)</b>	49.3 (6.6)	49.9 (6.5)	48.5 (6.4)	48.5 (6.4)	50.5 (7.0)
<45	32,749 (33.8)	6,934 (28.7)	9,350 (38.7)	9,436 (39.0)	7,029 (29.1)
[45-50[	23,530 (24.3)	6,180 (25.6)	6,046 (25.0)	5,993 (24.8)	5,311 (22.0)
[50-55[	18,999 (19.7)	5,347 (22.1)	4,405 (18.2)	4,332 (17.9)	4,915 (20.3)
[55-60[	12,826 (13.3)	3,539 (14.6)	2,645 (10.9)	2,684 (11.1)	3,958 (16.4)
≥60	8,561 (8.9)	2,166 (9.0)	1,720 (7.2)	1,721 (7.2)	2,954 (12.2)
<b>Education</b>					
<High school	12,815 (13.8)	3,703 (16.0)	2,970 (12.8)	2,807 (12.1)	3,335 (14.4)
≥High school	79,814 (86.2)	19,373 (84.0)	20,210 (87.2)	20,362 (87.9)	19,869 (85.6)
<i>Missing</i>	4,036	1,090	986	997	963
<b>Smoking</b>					
Never	52,014 (53.9)	12,822 (53.1)	12,580 (52.2)	12,850 (53.2)	13,762 (57.0)
Ex	29,911 (31.0)	7,298 (30.2)	7,659 (31.8)	7,707 (32.0)	7,247 (30.0)
Current	14,564 (15.1)	4,012 (16.7)	3,852 (16.0)	3,565 (14.8)	3,135 (13.0)
<i>Missing</i>	176	34	75	44	23
<b>Age at menarche (y), M (SD)</b>	12.8 (1.4)	12.8 (1.4)	12.8 (1.4)	12.8 (1.4)	12.9 (1.4)
≤11	19,761 (20.9)	5,030 (21.4)	4,963 (21.0)	4,942 (21.0)	4,826 (20.4)
[12-13]	47,656 (50.5)	11,942 (50.7)	12,118 (51.2)	11,892 (50.4)	11,704 (49.5)
≥14	27,002 (28.6)	6,564 (27.9)	6,573 (27.8)	6,734 (28.6)	7,131 (30.1)
<i>Missing</i>	2,246	630	512	598	506
<b>Menopausal status</b>					
Premenopausal	53,531 (57.5)	12,620 (53.7)	14,166 (61.7)	14,568 (62.9)	12,177 (51.6)
Natural menopause	32,520 (34.9)	8,841 (37.6)	7,119 (31.1)	7,058 (30.5)	9,502 (40.2)
Artificial menopause	6,410 (6.9)	1,810 (7.7)	1,436 (6.3)	1,382 (6.0)	1,782 (7.5)
Unknown type of menopause	694 (0.7)	217 (1.0)	200 (0.9)	129 (0.6)	148 (0.7)
<i>Missing</i>	3,510	678	1,245	1,029	558
<b>Parity</b>					
Nulliparous	11,586 (12.1)	3,705 (15.5)	2,865 (11.9)	2,549 (10.6)	2,467 (10.3)
One child	15,571 (16.2)	4,405 (18.4)	4,067 (17.0)	3,701 (15.4)	3,398 (14.2)
Two children	40,724 (42.5)	9,632 (40.3)	10,617 (44.2)	10,722 (44.7)	9,753 (40.6)
≥3 children	28,011 (29.2)	6,168 (25.8)	6,449 (26.9)	7,030 (29.3)	8,364 (34.9)
<i>Missing</i>	773	256	168	164	185
<b>Place of residence</b>					
Rural	13,089 (14.8)	2,138 (9.7)	2,844 (12.9)	3,525 (16.0)	4,582 (20.7)
Urban	75,327 (85.2)	19,939 (90.3)	19,233 (87.1)	18,570 (84.0)	17,585 (79.3)
<i>Missing</i>	8,249	2,089	2,089	2,071	2,000
<b>BMI (kg/m<sup>2</sup>), M (SD)</b>	22.6 (3.2)	22.9 (3.5)	22.5 (3.1)	22.4 (3.0)	22.6 (3.0)
<18.5	3,995 (4.2)	1,100 (4.7)	1,006 (4.3)	1,011 (4.3)	878 (3.7)
[18.5-25.0[	73,801 (78.3)	17,537 (74.7)	18,648 (79.0)	18,832 (79.7)	18,784 (79.4)
[25.0-30.0[	13,658 (14.5)	3,767 (16.1)	3,247 (13.8)	3,221 (13.6)	3,423 (14.5)
≥30.0	2,860 (3.0)	1,046 (4.5)	693 (2.9)	555 (2.4)	566 (2.4)
<i>Missing</i>	2,351	716	572	547	516

M, mean; SD: Standard deviation; BMI, body mass index; PA, physical activity.

Baseline characteristics are shown for participants available for survival analyses based on a 5y-lag.

<sup>a</sup> PA was not available at baseline for 1,169 women (1.2%); therefore we used the first LPA value available over the follow-up for these women (95% at Q3-1993 or Q5-1997).

<sup>b</sup> Total PA assessed at the baseline questionnaire (1990-Q1).

**Table 8: Participant's characteristics at baseline (1990-Q1) according to Parkinson's disease status at the end of the follow-up.**

Characteristics, No. (%)	Parkinson's Disease	
	No	Yes
<b>Age (y), M (SD)</b>	95,502 (98.8)	1,163 (1.2)
<45	49.3 (6.6)	52.6 (6.5)
[45-50[	32,564 (34.0)	185 (15.9)
[50-55[	23,298 (24.4)	232 (19.9)
[55-60[	18,686 (19.6)	313 (26.9)
≥60	12,577 (13.2)	249 (21.4)
<b>Education</b>	8,377 (8.8)	184 (15.9)
<High school	12,665 (13.8)	150 (13.5)
≥High school	78,852 (86.2)	962 (86.5)
	<i>Missing</i>	3,985
<b>Smoking</b>		51
Never	51,138 (54.0)	699 (60.6)
Ex	29,363 (31.0)	329 (28.5)
Current	14,260 (15.0)	125 (10.9)
	<i>Missing</i>	741
<b>Age at menarche (y), M (SD)</b>		10
≤11	12.8 (1.4)	12.9 (1.5)
[12-13]	19,512 (20.9)	249 (22.0)
≥14	47,140 (50.5)	516 (45.6)
	<i>Missing</i>	26,636 (28.6)
<b>Menopausal status</b>		32
Premenopausal	53,382 (58.0)	424 (37.5)
Natural menopause	31,727 (34.5)	551 (48.7)
Artificial menopause	6,251 (6.8)	137 (12.1)
Unknown type of menopause	664 (0.7)	19 (1.7)
	<i>Missing</i>	3,478
<b>Parity</b>		32
Nulliparous	11,451 (12.1)	135 (11.8)
One child	15,420 (16.3)	151 (13.1)
Two children	40,257 (42.5)	467 (40.6)
≥3 children	27,614 (29.1)	397 (34.5)
	<i>Missing</i>	760
<b>Place of residence</b>		13
Rural	12,959 (14.8)	130 (12.2)
Urban	74,392 (85.2)	935 (87.8)
	<i>Missing</i>	8,151
<b>BMI (kg/m<sup>2</sup>), M (SD)</b>		98
<18.5	22.6 (3.2)	22.9 (3.0)
[18.5-25.0[	3,956 (4.2)	39 (3.4)
[25.0-30.0[	72,918 (78.3)	883 (77.5)
≥30.0	13,466 (14.5)	192 (16.8)
	<i>Missing</i>	2,834 (3.0)
		26 (2.3)
		23

M, mean; SD: Standard deviation; BMI, body mass index.

#### **4.4.1. Trajectories of physical activity preceding Parkinson's disease**

The mean (SD) age of cases and controls at the index date was 71.9 years (SD=7.8). Figure 6 shows the latent physical activity trajectories in cases and controls for the most common profile of covariates and three different ages at the index date; Annex 2 presents estimates from the corresponding model showing a significant interaction between PD status and time ( $P=0.003$ ).

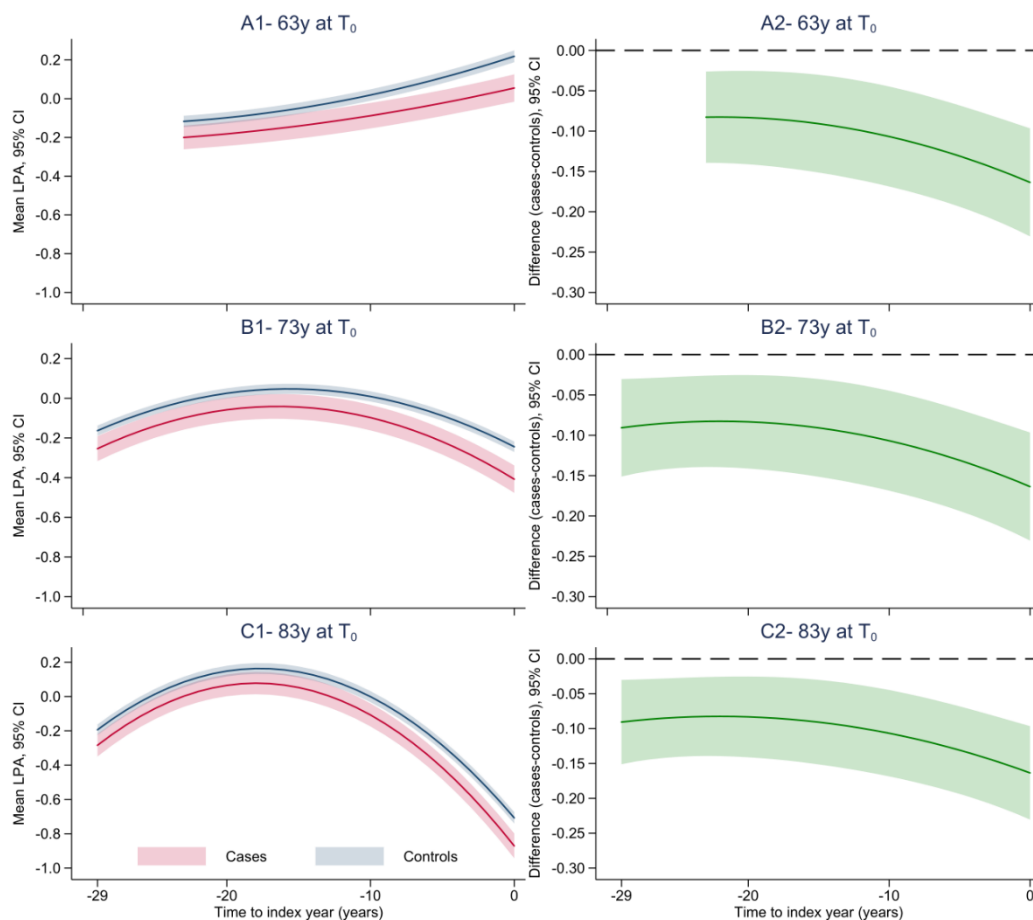
After an initial increase of latent physical activity in cases and controls, the latent physical activity decreased with a steeper decline in cases than in controls. Latent physical activity was significantly lower in cases than in controls throughout the follow-up; it was significantly lower in cases than controls 29 years before the index date and the difference started to increase approximately 10 years before the index date; therefore, the case-control difference was larger at the index date than at the beginning of the study. Based on these findings, we used a 10-year lag for our main survival analysis.

#### **4.4.2. Physical activity and Parkinson's disease incidence**

Table 9 shows the association between time-varying latent physical activity and PD incidence. In our main analysis (10-year lag), the hazard of PD decreased with increasing latent physical activity ( $P$ -trend=0.001), with 25% lower incidence in those in the highest quartile compared to the lowest (HR=0.75, 95% CI=0.63-0.89). Analyses based on splines showed no departure from linearity ( $P=0.25$ ; Figure 7). Using longer lags yielded similar associations (Table 9); the inverse association was borderline significant for the 20-year lag ( $P$ -trend=0.06), likely due to a smaller number of PD cases.

Sensitivity analyses adjusted for adherence to prudent diet, and caffeine and lactose intake (Table 10) or excluding baseline physical activity assessments (Table 11) yielded similar results; analyses excluding one activity (climbing stairs) from the definition of the latent physical activity variable also yielded consistent findings (data not shown). Analyses stratified by median age showed similar associations in both age groups (Table 12).

**Figure 6: Trajectories of the latent physical activity (LPA) in cases with Parkinson’s disease and matched controls up to 29 years before the index date.**



Figures A1, B1, and C1 show the trajectories (95% confidence interval, CI) of mean latent physical activity (LPA) in 1,196 PD cases and 23,879 matched controls based on a linear mixed model with a quadratic function of retrospective time; the model’s coefficients are shown in Annex 2.

Figures A2, B2, and C2 show the differences (95% CI) between the mean trajectories in PD cases and controls. Differences whose CI do not include 0 (horizontal dashed line) are statistically significant.

We used a retrospective time scale, with  $T_0$  (time=0) representing the year of PD diagnosis in cases and the index date in controls. The model was adjusted for PD status, age at  $T_0$ , and two-way interactions of time with PD status and age at  $T_0$ . It was further adjusted for baseline parity, place of residence, age at menarche, and time-varying smoking and menopausal status.

Given the significant interaction between age at  $T_0$  and time, trajectories were plotted for three different ages at  $T_0$  (63 years, 73 years, 83 years) and the most common profile of E3N participants (never smokers, age at menarche at 12-13 years, natural menopause, two children, and living in urban areas).

**Table 9: Association of time-varying physical activity with Parkinson's disease incidence.**

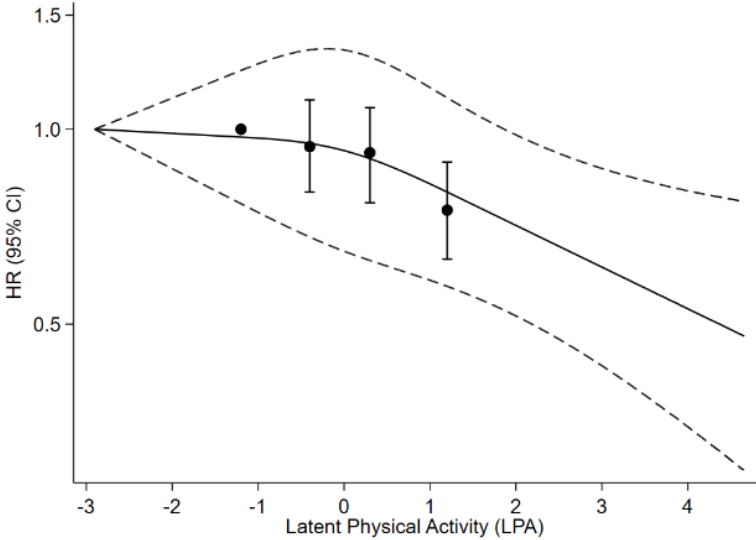
Latent physical activity	Cases (n)	IR	Age-adjusted HR (95% CI)	P- value	Multivariable HR (95% CI) <sup>a</sup>	P- value
<b>5-year lag (FU 1995-2018)</b>	<b>1,163</b>					
Quartile 1	316	0.60	1.00 (Ref.)	-	1.00 (Ref.)	-
Quartile 2	304	0.60	1.01 (0.86-1.18)	0.95	0.99 (0.85-1.16)	0.94
Quartile 3	289	0.56	0.96 (0.81-1.12)	0.58	0.94 (0.80-1.11)	0.49
Quartile 4	254	0.45	0.76 (0.64-0.89)	<0.001	0.75 (0.63-0.88)	<0.001
			<i>P</i> -linear trend	<0.001	<i>P</i> -linear trend	<0.001
<b>10-year lag (FU 2000-2018)</b>	<b>1,074</b>					
Quartile 1	286	0.73	1.00 (Ref.)	-	1.00 (Ref.)	-
Quartile 2	274	0.69	0.95 (0.80-1.12)	0.57	0.94 (0.80-1.11)	0.50
Quartile 3	268	0.67	0.93 (0.78-1.10)	0.37	0.92 (0.77-1.08)	0.30
Quartile 4	246	0.55	0.76 (0.64-0.90)	0.002	0.75 (0.63-0.89)	0.001
			<i>P</i> -linear trend	0.002	<i>P</i> -linear trend	0.001
<b>15-year lag (FU 2005-2018)</b>	<b>901</b>					
Quartile 1	237	0.80	1.00 (Ref.)	-	1.00 (Ref.)	-
Quartile 2	229	0.74	0.93 (0.78-1.12)	0.44	0.92 (0.77-1.10)	0.37
Quartile 3	222	0.72	0.90 (0.75-1.08)	0.27	0.89 (0.74-1.07)	0.20
Quartile 4	213	0.63	0.79 (0.66-0.95)	0.01	0.78 (0.64-0.94)	0.008
			<i>P</i> -linear trend	0.01	<i>P</i> -linear trend	0.008
<b>20-year lag (FU 2010-2018)</b>	<b>662</b>					
Quartile 1	166	0.88	1.00 (Ref.)	-	1.00 (Ref.)	-
Quartile 2	178	0.85	0.97 (0.78-1.20)	0.78	0.96 (0.78-1.19)	0.71
Quartile 3	167	0.81	0.92 (0.74-1.14)	0.44	0.90 (0.73-1.12)	0.36
Quartile 4	151	0.73	0.83 (0.67-1.04)	0.11	0.82 (0.65-1.02)	0.08
			<i>P</i> -linear trend	0.09	<i>P</i> -linear trend	0.06

FU, follow-up; IR, age-standardized incidence rate of Parkinson's disease per 1,000 person-years.

Hazard ratios (HR) and 95% confidence intervals (CI) calculated using Cox proportional hazards models for time-dependent variables with age as the time scale.

<sup>a</sup> Models are adjusted for baseline place of residence (rural/urban), age at menarche ( $\leq 11/12-13/\geq 14$  years), parity (nulliparous/one child/two children/ $\geq$ three children), and time-varying smoking (never/ex/current) and menopausal status (premenopausal/natural menopause/artificial menopause/unknown type of menopause).

**Figure 7: Hazard ratios of Parkinson’s disease in relation to latent physical activity (10-year lag) modelled with restricted cubic splines.**



Hazard ratios (HR) and 95% confidence intervals (CI) calculated using Cox proportional hazards models with age as the time scale, and adjusted for baseline place of residence (rural/urban), age at menarche ( $\leq 11/12-13/\geq 14$  years), parity (nulliparous/one child/two children/ $\geq$ three children), and time-varying smoking (never/ex/current) and menopausal status (premenopausal/natural menopause/artificial menopause/unknown type of menopause).

Latent physical activity was modelled as a restricted cubic spline and 3 knots provided the best fit (lower AIC values). There was no significant departure from linearity ( $P=0.25$ ). The solid line represents the HR of PD for continuous Latent physical activity modelled with splines; 95% CIs are shown as dashed lines.

The dots correspond to HRs for quartiles of latent physical activity compared to the reference quartile (Table 9) together with their 95% CI (vertical bars); HRs are plotted at the median of each quartile.



**Table 10: Association of physical activity (Q5-1997) with Parkinson’s disease incidence (N=81,777): analyses adjusted for diet (Q3-1993).**

PA (METs-h/week) in 1997 (Q5)	Cases (n)	IR	Age-adjusted HR (95% CI)	P- value	Multivariable Model 1 HR (95% CI) <sup>a</sup>	P- value	Multivariable Model 2 HR (95% CI) <sup>b</sup>	P- value
<b>5-year lag (FU 2002-2018)</b>	<b>881</b>							
Quartile 1	235	0.79	1.00 (Ref.)	-	1.00 (Ref.)	-	1.00 (Ref.)	-
Quartile 2	207	0.68	0.85 (0.70-1.02)	0.09	0.85 (0.70-1.02)	0.08	0.84 (0.70-1.01)	0.07
Quartile 3	225	0.69	0.86 (0.72-1.03)	0.10	0.85 (0.70-1.02)	0.07	0.84 (0.70-1.01)	0.06
Quartile 4	214	0.61	0.77 (0.64-0.92)	0.005	0.75 (0.62-0.90)	0.002	0.74 (0.62-0.90)	0.002
			<i>P</i> -linear trend	0.01	<i>P</i> -linear trend	0.005	<i>P</i> -linear trend	0.004
<b>10-year lag (FU 2007-2018)</b>	<b>685</b>							
Quartile 1	176	0.87	1.00 (Ref.)	-	1.00 (Ref.)	-	1.00 (Ref.)	-
Quartile 2	160	0.77	0.87 (0.71-1.08)	0.22	0.87 (0.70-1.08)	0.20	0.86 (0.70-1.07)	0.18
Quartile 3	182	0.83	0.94 (0.77-1.16)	0.57	0.93 (0.75-1.14)	0.47	0.92 (0.74-1.13)	0.41
Quartile 4	167	0.72	0.82 (0.66-1.01)	0.07	0.80 (0.65-0.99)	0.04	0.79 (0.64-0.98)	0.04
			<i>P</i> -linear trend	0.11	<i>P</i> -linear trend	0.07	<i>P</i> -linear trend	0.06

PA, physical activity; FU, follow-up; IR, age-standardized incidence rate of Parkinson’s disease per 1,000 person-years.

Hazard ratios (HR) and 95% confidence intervals (CI) calculated using Cox proportional hazards models with age as the time scale.

Lags longer than 10y were not possible to implement due to an insufficient number of cases.

<sup>a</sup> Multivariable Model 1 is adjusted for baseline place of residence (rural/urban), age at menarche ( $\leq 11/12-13/\geq 14$  years), parity (nulliparous/one child/two children/ $\geq$ three children), smoking (never/ex/current), and menopausal status (premenopausal/natural menopause/artificial menopause/unknown type of menopause).

<sup>b</sup> Multivariable Model 2 is further adjusted for caffeine intake (mg, quartiles), adherence to prudent diet (quartiles), and lactose intake (g, quartiles) assessed in 1993 (Q3).

**Table 11: Association of physical activity with Parkinson’s disease incidence: exclusion of physical activity assessed at baseline (Q1).**

Latent physical activity	Cases (n)	IR	Age-adjusted HR (95% CI)	P-value	Multivariable HR (95% CI) <sup>a</sup>	P-value
<b>5-year lag (FU 1998-2018)</b>						
	<b>1,015</b>					
Quartile 1	297	0.74	1.00 (Ref.)	-	1.00 (Ref.)	-
Quartile 2	248	0.66	0.90 (0.76-1.06)	0.21	0.89 (0.75-1.05)	0.18
Quartile 3	246	0.63	0.87 (0.73-1.03)	0.11	0.86 (0.73-1.02)	0.08
Quartile 4	224	0.50	0.69 (0.58-0.82)	<0.001	0.68 (0.57-0.81)	<0.001
			<i>P</i> -linear trend	<0.001	<i>P</i> -linear trend	<0.001
<b>10-year lag (FU 2003-2018)</b>						
	<b>841</b>					
Quartile 1	283	0.80	1.00 (Ref.)	-	1.00 (Ref.)	-
Quartile 2	281	0.79	0.99 (0.82-1.20)	0.92	0.98 (0.81-1.18)	0.83
Quartile 3	260	0.73	0.93 (0.77-1.12)	0.46	0.92 (0.76-1.11)	0.37
Quartile 4	250	0.61	0.77 (0.63-0.93)	0.007	0.76 (0.62-0.92)	0.004
			<i>P</i> -linear trend	0.006	<i>P</i> -linear trend	0.003
<b>15-year lag (FU 2008-2018)</b>						
	<b>652</b>					
Quartile 1	156	0.81	1.00 (Ref.)	-	1.00 (Ref.)	-
Quartile 2	179	0.94	1.16 (0.93-1.44)	0.18	1.15 (0.92-1.42)	0.21
Quartile 3	165	0.83	1.02 (0.82-1.27)	0.85	1.00 (0.81-1.25)	0.97
Quartile 4	152	0.68	0.83 (0.67-1.04)	0.11	0.82 (0.65-1.02)	0.08
			<i>P</i> -linear trend	0.06	<i>P</i> -linear trend	0.04

FU, follow-up; IR, age-standardized incidence rate of Parkinson’s disease per 1,000 person-years. Analyses are based on PA assessed at Q3 (1993) or later.

Hazard ratios (HR) and 95% confidence intervals (CI) calculated using Cox proportional hazards models with age as the time scale. Lags longer than 15 years were not possible to implement due to an insufficient number of cases.

<sup>a</sup> Multivariable models are adjusted for baseline place of residence (rural/urban), age at menarche ( $\leq 11/12-13/\geq 14$  years) and parity (nulliparous/one child/two children/ $\geq$ three children), and time-varying smoking (never/ex/current), and menopausal status (premenopausal/natural menopause/artificial menopause/unknown type of menopause).

**Table 12: Association of physical activity with Parkinson’s disease incidence: analyses stratified by median age.**

Latent physical activity	Age $\leq$ 72.8 years			Age $>$ 72.8 years		
	Cases (n)	Multivariable HR (95% CI)	P-value	Cases (n)	Multivariable HR (95% CI)	P-value
<b>10-year lag (FU 2000-2018)</b>						
	<b>497</b>			<b>577</b>		
Quartile 1	125	1.00 (Ref.)	-	161	1.00 (Ref.)	-
Quartile 2	139	1.01 (0.79-1.29)	0.92	135	0.89 (0.70-1.12)	0.31
Quartile 3	132	0.97 (0.76-1.24)	0.80	136	0.87 (0.69-1.10)	0.25
Quartile 4	101	0.78 (0.60-1.01)	0.06	145	0.73 (0.58-0.92)	0.01
		<i>P</i> -linear trend	0.06		<i>P</i> -linear trend	0.01

FU, follow-up. Analyses were stratified in two age groups based on the median age at diagnosis in PD cases (72.8 years). Hazard ratios (HR) and 95% confidence intervals (CI) calculated using Cox proportional hazards models with age as the time scale.

Multivariable models are adjusted for baseline place of residence (rural/urban), age at menarche ( $\leq 11/12-13/\geq 14$  years), parity (nulliparous/one child/two children/ $\geq$ three children), and time-varying smoking (never/ex/current) and menopausal status (premenopausal/natural menopause/artificial menopause/unknown type of menopause).

## 4.5. Discussion

In this cohort study of ~100,000 French women with six physical activity measures over 29 years of follow-up, increasing physical activity was associated with reduced PD incidence, while taking into account the potential for reverse causation. Analyses of physical activity trajectories showed that PD cases had lower physical activity levels than controls 29 years before the index date, and that case-control differences in physical activity level increased ~10 years prior to diagnosis in agreement with the hypothesis that prodromal PD leads to a reduction in physical activity.

Previous cohort studies on the relation between physical activity and PD yielded inconsistent findings (Table 2). One meta-analysis (eight studies, 2,192 PD patients) with a median follow-up period of 12 years (range=6.1-22.0) showed that physical activity was associated with lower PD incidence in analyses of men and women combined (HR=0.79, 95% CI=0.68-0.91).[75] This association was statistically significant in men (HR=0.90, 95% CI=0.85-0.95) and weaker and not significant in women (HR=0.95, 95% CI=0.87-1.04). However, only four studies examined women only (604 PD patients),[69, 83, 84, 86] and did not adjust for characteristics of reproductive life associated with PD.[166] In addition, no previous study used repeated physical activity measures to take into account changes in physical activity over the follow-up. Five studies performed sensitivity analysis using a lag between the physical activity assessment and PD incidence (4-year lag,  $n=3$ ; 8-year lag,  $n=1$ ; 10-year lag,  $n=1$ ), and only one with a 4-year lag showed a significant inverse association, overall and in sex-stratified analyses (198 women with PD).[86] One study showed that moderate-to-vigorous physical activity at ages 35-39 years was associated with lower PD risk in men and women,[86] while another found a significant inverse association between higher physical activity level in early adulthood and PD in men but not in women.[69]

The main difference between previous studies and ours is that we identified a considerably larger number of women with incident PD over a longer follow-up

allowing us to perform analyses with longer lags while retaining a sufficient number of PD patients. Physical activity trajectories in controls were consistent with decreasing physical activity levels in the elderly.[167, 168] This decline was steeper in cases, with a case-control difference that started to increase ~10 years before PD diagnosis, emphasizing the importance of performing lagged analyses to estimate associations between physical activity and PD not biased by reverse causation. Analyses with lags  $\geq 10$  years confirmed an inverse association between physical activity and PD incidence that was significant for a 15-year lag and borderline significant for a 20-year lag, likely due to the smaller number of cases. Therefore, our findings suggest that reverse causation is unlikely to explain the inverse association between physical activity and PD.

Converging evidence from studies in PD patients, including observational studies[34, 169] and randomized controlled trials,[27, 71, 72] suggest that physical activity improves PD motor and non-motor symptoms. In PD patients, aerobic exercise stabilizes disease progression in the cortico-striatal sensorimotor network and enhances cognitive performance.[32] Our results extend these findings and suggest that physical activity may help to prevent or delay PD onset, possibly by slowing PD pathological processes, in agreement with one study that showed a reduced prevalence of PD prodromal symptoms in individuals more physically active in midlife.[170]

These findings have triggered interest in elucidating the mechanisms that explain beneficial effects of physical activity for PD. Exercise induces recovery of motor function and neuroprotection of dopaminergic neurons in animal models of PD, regulates dopaminergic and glutamatergic transmission, mobilizes neurotrophic factors (BDNF, GDNF), modulates neuro-inflammatory mechanisms, attenuates mitochondrial dysfunction and oxidative stress, and enhances brain plasticity.[77-80] In humans, physical activity has been associated with brain structural and functional changes until late adulthood.[81] In postmenopausal women, higher fitness levels were associated with higher antioxidant enzyme activity and lower levels of oxidative

stress.[82]

The main strengths of our study are its large size and long follow-up which allowed us to perform lagged analyses in order to address the potential for reverse causation. We used repeated physical activity measures rather than a single measure, and a method specifically designed to allow longitudinal analyses when measurement tools change over the follow-up.[163] Our approach to ascertain PD patients yielded incidence rates comparable to those in women from Western Europe, in favor of its validity.[134] Finally, few studies specifically examined the relation between physical activity and PD in women, possibly because PD is more frequent in men than women, while our study focused on this understudied population,[132] and analyses were adjusted for the characteristics of reproductive life.

The main limitation of our study is that we used self-reported physical activity rather than objective measures (e.g., accelerometer) that are considered more valid, although they do not capture all types of physical activity (underestimating, for example, cycling and carrying a load). They are however difficult and costly to implement on a large scale, and only capture physical activity over a few days. Measurement error is inevitable for self-reported physical activity, but is reduced through the estimation of the latent process by the LPMM[164] and is likely to be non-differential and lead to underestimated associations. Recent studies showed that both self-reported and objectively measured physical activity were associated with outcomes such as all-cause and cardiovascular mortality or self-rated health.[171, 172] Second, there was an insufficient number of questions for vigorous and moderate physical activity to generate separate latent variables. Third, our analyses based on time-varying physical activity were not adjusted for diet as it was not recorded at baseline; we performed sensitivity analyses adjusted for dietary characteristics associated with PD, showing that diet was not a strong confounder of the association between physical activity and PD. Finally, E3N participants are mostly educated and health-conscious teachers who are not representative of the general population.

However, it is generally considered that representativeness is not essential for estimating associations, and associations in occupational cohorts are not necessarily different compared to those estimated in the general population.[173, 174]

In conclusion, our findings reinforce the evidence in favor of the health benefits of physical activity and provide stronger evidence than previous studies in favor of an inverse association between physical activity and PD in women not explained by reverse causation. These results are important for planning interventions for PD prevention,[26] and warrant further studies to understand which type and level of physical activity are beneficial.



## **5. Objective 2: Association between adiposity and Parkinson's disease**





## 5.1. Introduction

The relation between BMI and the risk of PD is complex and previous studies provided inconsistent results. A meta-analysis of 10 cohort studies showed no association between continuous BMI and PD, both overall (HR per 5 kg/m<sup>2</sup>=1.00, 95% CI=0.89-1.12; heterogeneity I<sup>2</sup>=64.5%, P=0.003), in men (6 studies; HR=1.03, 95% CI=0.90-1.18) and women (3 studies; HR=1.04, 95% CI=0.83-1.30).[91] In another meta-analysis, overweight (eight studies) or obesity (four studies) were not associated with PD, while underweight (three studies) was associated with increased PD risk.[92] Differences in duration of follow-up across studies partly explained between-study heterogeneity. Fewer studies on other abdominal adiposity measures (WC; WHR; WHtR) are available and also yielded inconsistent findings.[95-99]

Weight loss is common in PD patients who have lower average weight than controls, with a difference that increases with disease duration.[89] Moreover, one cohort study showed that weight loss begins 2-4 years before diagnosis,[68] and a study that used dual-energy x-ray absorptiometry showed that PD patients began to lose total/fat mass ~6-7 years before diagnosis.[175] Therefore, the potential for reverse causation needs to be considered for studies on the relation between BMI and the risk of PD. As previously discussed (§1.3.2), PD prodromal symptoms (e.g., constipation, fatigue) may lead to changes in diet, physical activity, or other behaviours leading to changes in weight before diagnosis. Therefore, cohort studies with a long follow-up are needed to examine the relation between BMI and PD to overcome the influence of reverse causation. In addition, all previous cohort studies were based on a single BMI measure. Additional studies with repeated BMI measures are needed to examine BMI trajectories prior to diagnosis.

Our aim was to examine the association of time-varying measures of BMI, WC, WHR, and WHtR with PD incidence in women from the French E3N cohort study, by using lagged analyses to address reverse causation, and to compare BMI trajectories

over 29 years prior to diagnosis in patients and matched controls.

## **5.2. Methods**

### **5.2.1. Assessment of body mass index and abdominal adiposity**

Weight was self-reported at all visits (11 measures) and height at Q1, Q4 (1994) and Q6-Q11 (2000-2014). Participants were asked to measure their anthropometric characteristics with a measuring tape, without shoes and in underwear. BMI was computed as weight divided by height squared ( $\text{kg}/\text{m}^2$ ) and categorized into a four-level variable (World Health Organization [WHO] recommendations): underweight,  $<18.5$ ; normal weight used as the reference category,  $[18.5-25.0[$ ; overweight,  $[25.0-30.0[$ ; obese,  $\geq 30.0 \text{ kg}/\text{m}^2$ ].[176] Height was standardized by using the most frequent value.

WC was self-measured at the narrowest torso circumference (cm) and hip circumference (HC) at the widest circumference over the buttocks (cm). Both were assessed at Q4 and between Q7-Q11 (2002-2014). We computed WHR as WC divided by HC, and WHtR as WC divided by height (all in cm). WC was categorized as  $\leq 80$ ,  $]80-88]$ , and  $>88$  cm; for WHR, cutoffs were:  $<0.85$  (reference),  $\geq 0.85$ ; for WHtR, cutoffs were:  $<0.5$ ,  $[0.5-0.6[$ ,  $\geq 0.6$ . For WC and WHtR, we used the middle category as the reference to examine quadratic relations.[177]

The validity of self-reported anthropometric measures was previously assessed in E3N.[178] Briefly, 150 participants completed a self-administered questionnaire the day before a clinical assessment by a skilled technician. Correlation coefficients were high for weight (0.94), height (0.89), BMI (0.92), hip circumference (0.90), but lower for waist circumference (0.79).

## **5.2.2. Nested case-control study**

In order to examine BMI trajectories prior to diagnosis in PD patients and matched controls, we set-up a case-control nested within E3N. Each incident PD patient was individually matched to 20 controls using incidence density sampling.[161] To be selected as controls, participants had to be alive and at risk of PD at the date of diagnosis of matched cases (index date,  $T_0$ ), and have the same age at  $T_0$ .

Among 25,200 women (1,200 PD patients, 24,000 controls), we excluded 50 women (four patients) with missing BMI at all visits before  $T_0$  and additional 78 controls matched to these four PD patients, leading to a final sample of 1,196 patients and 23,876 controls; 1,153 patients were matched to 20 controls, 42 patients to 19 controls, and one patient to 18 controls.

## **5.3. Statistical Analysis**

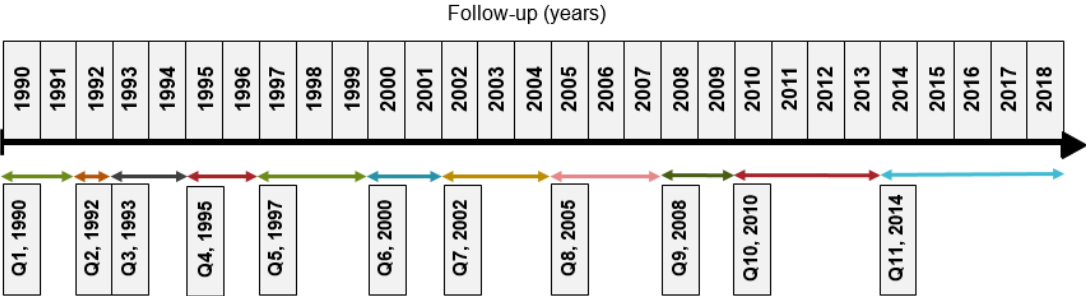
### **5.3.1. Survival Analysis**

We used Cox proportional hazards regression for time-varying variables with age as the time scale to estimate HRs, 95% CIs, and two-tailed  $P$ -values ( $\alpha=0.05$ ). To address the potential for reverse causation, time-varying exposures (adiposity measures, covariates) were lagged by five years in our main analysis (Figure 8). Therefore, the follow-up started five years after the baseline assessment and participants who developed PD over the first five years of follow-up were excluded (prevalent PD). Given the 5-year lag, women were followed since 1995 (for BMI) and 2000 (for WC, WHR, WHtR) until PD diagnosis or end of follow-up (maximum of last available questionnaire and last drug reimbursement). We also used a longer lag of 10-years for all exposures and 20-years for BMI (not possible for WC and WHR that were not measured at Q1). In sensitivity analyses, and to allow comparisons with other studies, we used the baseline exposures only (i.e., exposures were not time-varying).

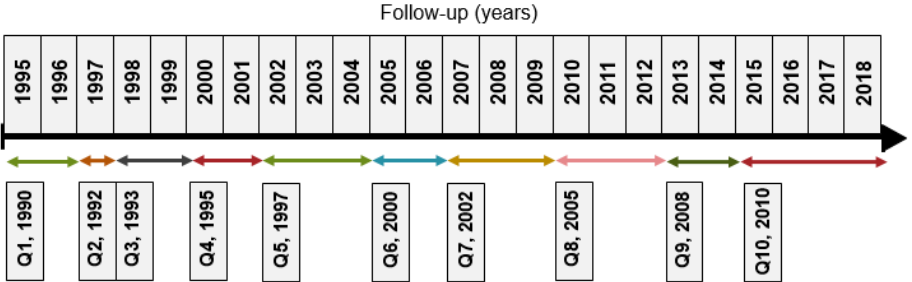
To test for linear trends in HRs, we included in the model ordinal variables defined by the median of adiposity measures in each category. We also tested quadratic trends by including the ordinal variables as both linear and quadratic effects; we performed a test of their combined effect if the quadratic term was statistically significant (i.e., in case of departure from linearity). We also modelled BMI using restricted cubic splines (3 knots) in sensitivity analyses.

**Figure 8: Time-varying BMI measures with no lag (A) and with 5-year lag (B).**

**A. BMI measures**



**B. BMI lagged by 5-years**



Main analyses were adjusted for baseline parity, age at menarche, and place of residence, and time-varying smoking, menopausal status, and physical activity; analyses for WC were also adjusted for height. Missing values were coded as specific categories to retain the same number of participants in all analyses. In sensitivity analysis using baseline BMI, we used the same covariates as above measured at baseline. For models adjusted for diet assessed at Q3, we started the analyses at Q5 (1997), so that diet was assessed before BMI.

Obesity increases the risk of hypercholesterolemia, cardiovascular disease, and diabetes. Statins (used for hypercholesterolemia and cardiovascular disease) are associated with reduced PD incidence,[179] while diabetes might increase PD incidence.[114] Hence, hypercholesterolemia, cardiovascular disease, and diabetes may be mediators of the association between BMI and PD. We performed mediation analyses adjusted for time-varying hypercholesterolemia, cardiovascular disease, and diabetes, to examine whether they attenuated the total effect of BMI on PD.

### **5.3.2. BMI trajectories in PD patients and controls**

BMI trajectories over 29 years of follow-up were examined within the nested case-control study using a retrospective timescale going back from the index year ( $T_0$ ) to the beginning of the study.

We modelled BMI (dependent variable) as a categorical four-level variable (reference,  $[18.5-25[ \text{ kg/m}^2$ ) using a generalized estimating equations (GEE) multinomial logistic regression model that allows to take into account repeated observations for a given subject. This model allowed us to compare trajectories of the frequencies of the four BMI categories in PD patients and controls over time, and to compute annual differences between the two groups.

We started with a full model that included the following terms: intercept, PD status, time (in years, divided by 10), time squared (to allow for non-linear BMI changes over time), two-way interactions between time and PD status ( $\text{time} \times \text{PD}$ ,  $\text{time}^2 \times \text{PD}$ ), age at  $T_0$  (centered at 73 years), age squared (to allow for non-linear BMI changes with age at  $T_0$ ), two-way interactions of age with time ( $\text{age} \times \text{time}$ ,  $\text{age}^2 \times \text{time}$ ,  $\text{age} \times \text{time}^2$ ,  $\text{age}^2 \times \text{time}^2$ ) and PD status ( $\text{age} \times \text{PD}$ ,  $\text{age}^2 \times \text{PD}$ ), as well as all three-way interactions ( $\text{age} \times \text{time} \times \text{PD}$ ,  $\text{age}^2 \times \text{time} \times \text{PD}$ ,  $\text{age} \times \text{time}^2 \times \text{PD}$ ,  $\text{age}^2 \times \text{time}^2 \times \text{PD}$ ); the model was then simplified progressively by excluding non-significant terms ( $P > 0.05$ ). In sensitivity analyses, we repeated these analyses using restricted cubic splines. Analyses were

adjusted for baseline parity, age at menarche and place of residence, and time-varying smoking, physical activity, and menopausal status.

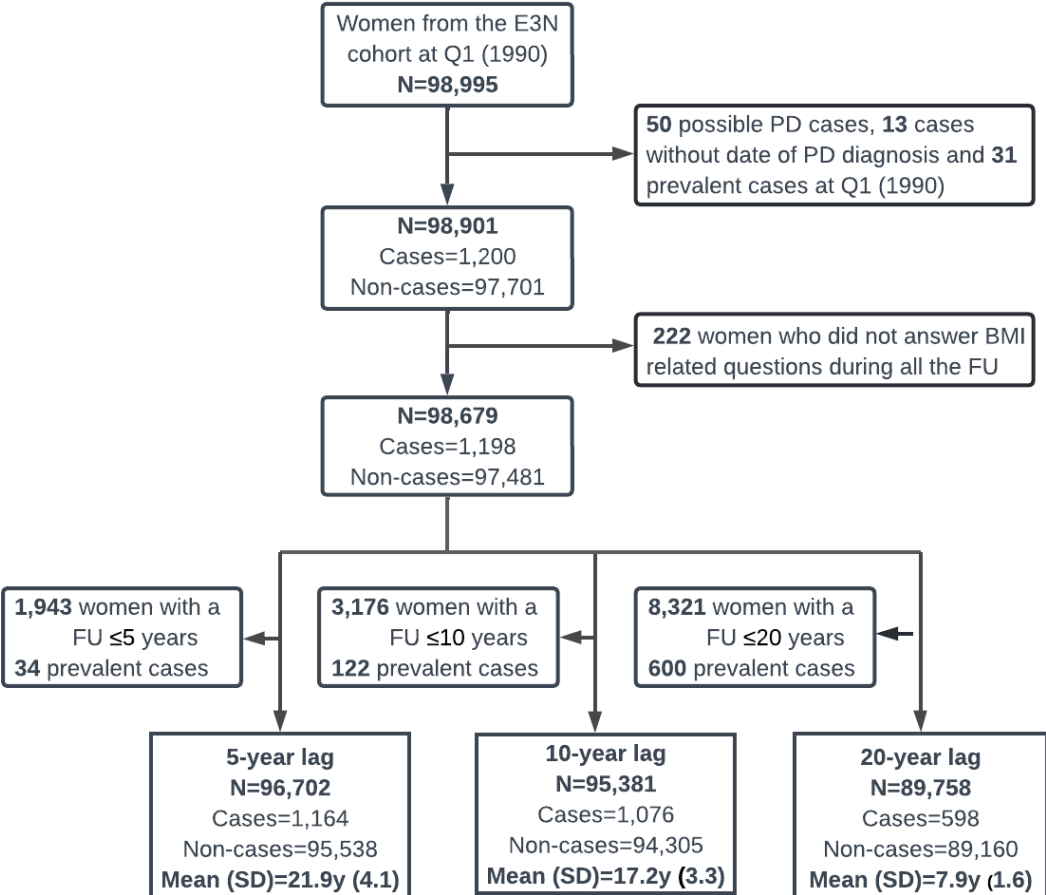
## 5.4. Results

Figure 9 shows a flowchart for inclusion into the study. We excluded 50 possible PD cases, 13 cases without a diagnosis date, 31 prevalent cases at Q1, and 222 women who did not provide information on weight and/or height at any visit. For our main analysis (5-year lag), we excluded 1,943 women not followed for at least 5 years and 34 additional prevalent PD cases, leaving 96,702 women followed during 24 years (mean=22.0, SD=4.0), of whom 1,164 developed PD. Compared to previous studies, ours is the third largest in terms of number of PD patients after two overlapping studies based on Korean healthcare electronic databases,[98, 99] and the largest study that performed analyses in women.

Participant's characteristics, overall and according to BMI categories, are presented in Table 13. The mean age of the women was 49.3 years (SD=6.6). At baseline (1990-Q1), mean BMI was 22.6 kg/m<sup>2</sup> (SD=3.2); 77.9% of the women had normal BMI, 4.2% were in the underweight category, 14.7% in the overweight category, and 3.2% were obese. BMI increased as women grew older; after 24 years of follow-up (2014-Q11), mean BMI was 24.0 kg/m<sup>2</sup> (SD=4.1; underweight, 4.7%; normal BMI, 61.3%; overweight, 25.7%; obesity, 8.3%).

Women who were less educated, ex-smokers, with early age of menarche, more children, and living in rural areas had higher BMI than their counterparts. Women with incident PD were older, less frequently smokers and in premenopausal status, practiced lower levels of physical activity, had later age of menarche, more children, and lived slightly more often in urban areas at baseline than those who did not develop PD; there was no difference in education level (Table 14). There was a strong age-adjusted correlation of BMI with WC (0.78) and WHtR (0.79), and weaker with WHR (0.30).

**Figure 9: Flow chart for inclusion into the study (5-year, 10-year, and 20-year lag).**





**Table 13: Participants' characteristics at baseline (1990-Q1) overall and according to body mass index.**

Characteristics, n (%)	Total 96,702	Body mass index (kg/m <sup>2</sup> )			
		<18.5	[18.5-25.0[	[25.0-30.0[	≥30.0
<b>BMI (kg/m<sup>2</sup>), M (SD)</b>	22.6 (3.2)	17.7 (0.7)	21.7 (1.6)	26.8 (1.3)	33.0 (3.1)
<b>Age (years), M (SD)</b>	49.3 (6.6)	47.5 (6.4)	49.0 (6.5)	51.5 (6.9)	51.2 (6.8)
<45	32,757 (33.8)	1,919 (47.0)	26,930 (35.7)	3,182 (22.4)	726 (23.6)
[45:50[	23,536 (24.3)	971 (23.8)	18,755 (24.9)	3,100 (21.9)	710 (23.1)
[50:55[	19,009 (19.7)	560 (13.7)	14,631 (19.4)	3,151 (22.2)	667 (21.7)
[55:60[	12,833 (13.3)	374 (9.2)	9,205 (12.2)	2,706 (19.1)	548 (17.8)
≥60	8,567 (8.9)	256 (6.3)	5,853 (7.8)	2,036 (14.4)	422 (13.8)
<b>Education</b>					
<High school	12,818 (13.8)	355 (9.1)	8,905 (12.3)	2,790 (20.7)	768 (26.3)
≥High school	79,841 (86.2)	3,549 (90.9)	63,447 (87.7)	10,697 (79.3)	2,148 (73.7)
<i>Missing</i>	4,043	176	3,022	688	157
<b>Physical activity</b>					
Q1	21,412 (22.2)	1,012 (24.8)	15,857 (21.1)	3,509 (24.8)	1,034 (33.9)
Q2	26,900 (27.9)	1,143 (28.0)	21,151 (28.1)	3,766 (26.6)	840 (27.5)
Q3	26,464 (27.4)	1,121 (27.5)	21,032 (27.9)	3,654 (25.9)	657 (21.5)
Q4	21,765 (22.5)	800 (19.7)	17,237 (22.9)	3,206 (22.7)	522 (17.1)
<i>Missing</i>	161	4	97	40	20
<b>Smoking</b>					
Never	52,015 (53.9)	2,070 (50.8)	40,333 (53.6)	7,959 (56.3)	1,653 (53.9)
Ex	29,916 (31.0)	1,181 (29.0)	23,348 (31.0)	4,385 (31.0)	1,002 (32.7)
Current	14,574 (15.1)	821 (20.2)	11,543 (15.4)	1,799 (12.7)	411 (13.4)
<i>Missing</i>	197	8	150	32	7
<b>Age at menarche (years), M (SD)</b>	12.8 (1.4)	13.2 (1.5)	12.8 (1.4)	12.6 (1.4)	12.3 (1.5)
≤11	19,765 (20.9)	587 (14.8)	14,398 (19.5)	3,760 (27.3)	1,020 (34.4)
[12-13]	47,660 (50.5)	1,930 (48.4)	37,635 (51.1)	6,742 (48.8)	1,353 (45.6)
≥14	26,989 (28.6)	1,468 (36.8)	21,635 (29.4)	3,294 (23.9)	592 (20.0)
<i>Missing</i>	2,288	95	1,706	379	108
<b>Menopausal status</b>					
Premenopausal	53,527 (57.4)	2,618 (66.9)	43,699 (60.1)	5,946 (43.5)	1,264 (42.9)
Natural menopause	32,531 (34.9)	1,092 (27.9)	24,094 (33.2)	6,100 (44.7)	1,245 (42.3)
Artificial menopause	6,415 (6.9)	177 (4.5)	4,418 (6.1)	1,430 (10.5)	390 (13.2)
Unknown type	719 (0.8)	25 (0.7)	466 (0.6)	183 (1.3)	45 (1.6)
<i>Missing</i>	3,510	168	2,697	516	129
<b>Parity</b>					
Nulliparous	11,625 (12.1)	710 (17.6)	8,837 (11.8)	1,667 (11.9)	411 (13.5)
One child	15,576 (16.2)	811 (20.0)	12,342 (16.5)	1,972 (14.0)	451 (14.9)
Two children	40,726 (42.5)	1,649 (40.7)	32,727 (43.8)	5,332 (38.0)	1,018 (33.5)
≥3 children	28,000 (29.2)	878 (21.7)	20,892 (27.9)	5,074 (36.1)	1,156 (38.1)
<i>Missing</i>	775	32	576	130	37
<b>Place of residence</b>					
Rural	13,088 (14.8)	481 (12.9)	9,979 (14.5)	2,144 (16.5)	484 (17.0)
Urban	75,322 (85.2)	3,255 (87.1)	58,893 (85.5)	10,813 (83.5)	2,361 (83.0)
<i>Missing</i>	8,292	344	6,502	1,218	228
<b>WC (cm), M (SD)<sup>a</sup></b>	76.2 (9.1)	65.5 (5.0)	74.3 (6.6)	86.4 (8.1)	99.7 (10.4)
<b>WHR, M (SD)<sup>a</sup></b>	0.79 (0.06)	0.77 (0.05)	0.78 (0.05)	0.81 (0.06)	0.84 (0.06)
<b>WHtR, M (SD)<sup>a</sup></b>	0.47 (0.06)	0.41 (0.03)	0.46 (0.04)	0.54 (0.05)	0.62 (0.06)

M, mean; SD, Standard deviation; BMI, body mass index; WC, waist circumference; WHR, waist-hip ratio; WHtR, waist-height ratio.

BMI was missing at baseline for 561 (0.6%) women, but was available over the follow-up; we therefore used the first BMI measure available over the follow-up for these women (80% assessed at Q3-1993 to Q5-1997).

<sup>a</sup>In 1994.

**Table 14: Participant's characteristics at baseline (1990) according to Parkinson's disease status at the end of the follow-up.**

Characteristics, <i>n</i> (%)	Parkinson's Disease	
	No	Yes
	95,538 (98.8)	1,164 (1.2)
<b>Age (years), M (SD)</b>	49.3 (6.6)	52.6 (6.5)
	<45	185 (15.9)
	[45:50[	233 (20.0)
	[50:55[	312 (26.8)
	[55:60[	250 (21.5)
	≥60	184 (15.8)
<b>Education</b> (MV=3,992/51)	<High school	150 (13.5)
	≥High school	963 (86.5)
<b>Physical activity</b> (MV=159/2)	Q1	293 (25.2)
	Q2	325 (28.0)
	Q3	298 (25.6)
	Q4	246 (21.2)
<b>Smoking</b> (MV=195/2)	Never	702 (60.4)
	Ex	334 (28.7)
	Current	126 (10.9)
<b>Age at menarche (years), mean (SD)</b> (MV=2,256/32)	≤11	249 (22.0)
	[12-13]	517 (45.7)
	≥14	366 (32.3)
<b>Menopausal status</b> (MV=3,478/32)	Premenopausal	421 (37.2)
	Natural menopause	555 (49.0)
	Artificial menopause	138 (12.2)
	Unknown type	18 (1.6)
<b>Parity</b> (MV=762/13)	Nulliparous	135 (11.7)
	One child	152 (13.2)
	Two children	467 (40.6)
	≥3 children	397 (34.5)
<b>Place of residence</b> (MV=8,194/98)	Rural	130 (12.2)
	Urban	936 (87.8)
<b>BMI (kg/m<sup>2</sup>), M (SD)</b> (MV=557/4)	22.6 (3.2)	22.9 (3.0)
	<18.5	41 (3.5)
	[18.5-25]	895 (77.2)
	[25-30]	197 (17.0)
	≥30	27 (2.3)
<b>WC (cm), M (SD)<sup>a</sup></b> (MV=30,117/319)	76.2 (9.1)	77.2 (8.5)
	≤80	595 (70.4)
	]80-88]	175 (20.7)
	≥88	75 (8.9)
<b>WHR, M (SD)<sup>a</sup></b> (MV=30,332/322)	0.79 (0.1)	0.79 (0.1)
	<0.85	722 (85.7)
	≥0.85	120 (14.3)
<b>WHtR, M (SD)<sup>a</sup></b> (MV=30,117/319)	0.47 (0.1)	0.48 (0.1)
	<0.5	583 (69.0)
	[0.5-0.6[	239 (28.3)
	≥0.6	23 (2.7)
	<i>Missing</i>	319

M, mean; MV, missing values in participants without and with PD; SD: Standard deviation; BMI, body mass index; WC, waist circumference; WHR, waist-hip ratio; WHtR, waist-height ratio; Q, quartile.

<sup>a</sup>In 1994.

### 5.4.1. BMI and Parkinson's disease incidence

Table 15 shows the association between time-varying BMI and PD incidence. In our main analysis (5-year lag), PD incidence was 24% lower in the obese group compared to the normal BMI group (HR=0.76, 95% CI=0.59-0.98). Women in the overweight group also had a decreased hazard of PD that was not statistically significant (HR=0.89, 95% CI=0.78-1.03), and there was no association for the underweight group (HR=0.99, 95% CI=0.71-1.37). PD incidence declined with increasing BMI ( $P$ -linear=0.016).

PD incidence remained lower in obese women compared to women with normal BMI in analyses lagged by 10 and 20 years (Table 15). In analyses lagged by 20 years, only a quadratic trend was present ( $P$ -quadratic=0.03), due to a lower, but not statistically significant, risk in underweight women. Overweight women had lower PD risk than women with normal BMI using a 10-year lag, but not with a 20-year lag. When we modelled BMI with restricted cubic splines, compared to a BMI of 22 kg/m<sup>2</sup> (reference), PD incidence decreased progressively as BMI increased (Figure 10).

Sensitivity analyses based on baseline BMI (1990-Q1) showed a similar pattern of association for obesity without significant trends (Table 16). Adjustment for diet (Q3-1993) did not attenuate the association between obesity (Q5-1997) and PD incidence (Table 17).

The inverse association between obesity and PD was not modified in analyses adjusted for potential mediators (hypercholesterolemia, cardiovascular disease, diabetes; Table 18).

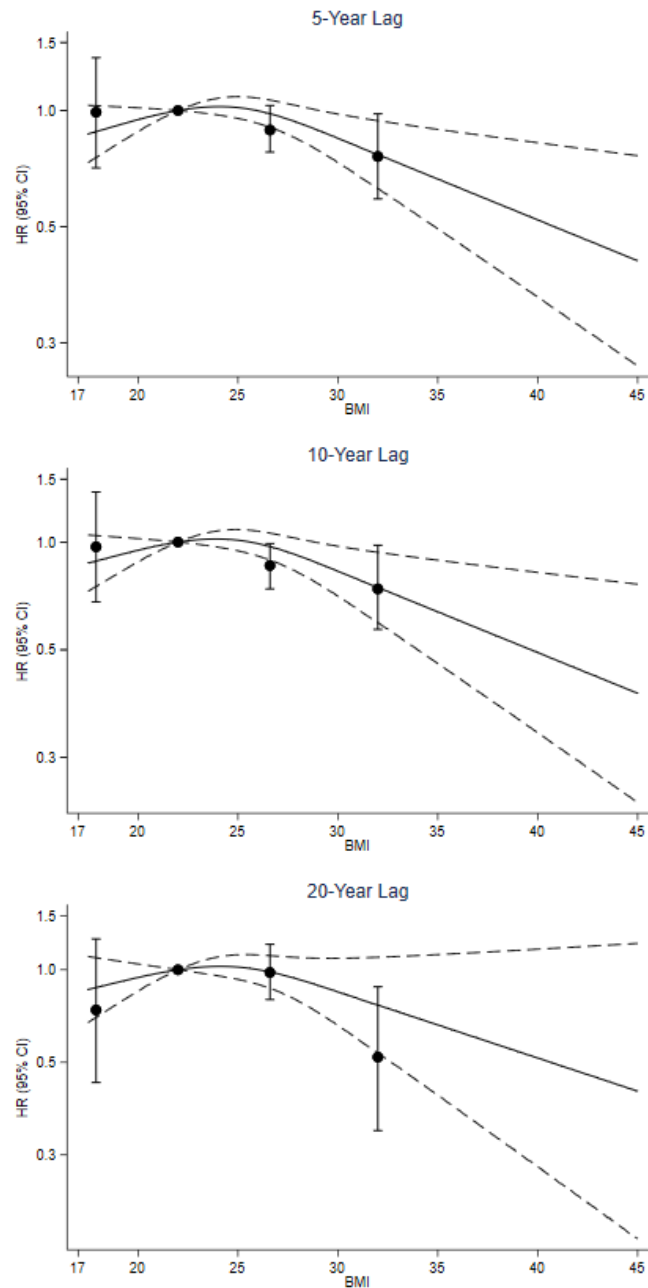
**Table 15: Association of time-varying body mass index with Parkinson's disease incidence over the follow-up.**

BMI (kg/m <sup>2</sup> )	Cases (n)	IR	HR (95% CI) <sup>a</sup>	P- value	P-trend	
					Linear	Quadratic
<b>5-year lag</b>						
<b>(FU 1995-2018, 96,702 participants) 1,164</b>						
<18.5	38	0.56	0.99 (0.71-1.37)	0.952		
[18.5-25[	779	0.57	1.00 (Ref.)	-		
[25-30[	278	0.53	0.89 (0.78-1.03)	0.114		
≥30	69	0.46	0.76 (0.59-0.98)	0.032	0.016	--
<b>10-year lag</b>						
<b>(FU 2000-2018, 95,381 participants) 1,076</b>						
<18.5	32	0.66	0.97 (0.68-1.38)	0.848		
[18.5-25[	756	0.69	1.00 (Ref.)	-		
[25-30[	233	0.61	0.86 (0.74-0.99)	0.041		
≥30	55	0.57	0.74 (0.57-0.98)	0.036	0.009	--
<b>20-year lag</b>						
<b>(FU 2010-2018, 89,758 participants) 598</b>						
<18.5	14	0.66	0.74 (0.43-1.26)	0.262		
[18.5-25[	453	0.85	1.00 (Ref.)	-		
[25-30[	117	0.87	0.98 (0.80-1.21)	0.886		
≥30	14	0.46	0.52 (0.30-0.88)	0.016	0.156	0.030

BMI, body mass index; FU, follow-up; IR, age-standardized incidence rate of Parkinson's disease per 1,000 person-years. Hazard ratios (HR) and 95% confidence intervals (CI) calculated using Cox proportional hazards models for time-varying variables with age as the time scale.

<sup>a</sup> Models are adjusted for baseline place of residence (rural/urban), age at menarche ( $\leq 11/12-13/\geq 14$  years) and parity (nulliparous/one child/two children/ $\geq$ three children), and time-varying smoking (never/ex/current), menopausal status (no/natural/artificial/unknown type), and physical activity (quartiles).

**Figure 10: Hazard ratios of Parkinson’s disease in relation to time-varying body mass index modelled with restricted cubic splines.**



Hazard ratios (HR) and 95% confidence intervals (CI) calculated using Cox proportional hazards models with age as the time scale, and adjusted for baseline place of residence (rural/urban), age at menarche ( $\leq 11/12-13/\geq 14$  years) and parity (nulliparous/one child/two children/ $\geq$ three children), and time-varying smoking (never/ex/current), menopausal status (no/natural/artificial/unknown type), and physical activity (quartiles). The figures are shown for increasing lags between BMI assessment and PD incidence.

Body mass index (BMI) was modelled as a restricted cubic spline and 3 knots provided the best fit (lower AIC values). The mean BMI (22 kg/m<sup>2</sup>) in the normal weight group ([18.5-25.0[ kg/m<sup>2</sup>) represents the reference value. The solid line represents the HR of PD for continuous BMI modelled with splines; 95% CIs are shown as dashed lines.

The dots correspond to HRs for BMI categories compared to the reference category ([18.5-25[ kg/m<sup>2</sup>; Table 15) together with their 95% CI (vertical bars); HRs are plotted at the median of each category.

**Table 16: Association of body mass index (BMI) assessed at baseline in 1990 (Q1) with Parkinson's disease incidence over the follow-up.**

BMI (kg/m <sup>2</sup> )	Cases			P-value	P-trend	
	(n)	IR	HR (95% CI) <sup>a</sup>		Linear	Quadratic
<b>5-year lag</b>						
<b>(FU 1995-2018, 96,141 participants) 1,160</b>						
<18.5	41	0.56	1.00 (0.73-1.37)	0.989		
[18.5-25[	895	0.55	1.00 (Ref.)	-		
[25-30[	197	0.56	0.98 (0.84-1.15)	0.804		
≥30	27	0.40	0.66 (0.44-0.97)	0.032	0.112	--
<b>10-year lag</b>						
<b>(FU 2000-2018, 94,842 participants) 1,074</b>						
<18.5	38	0.65	1.00 (0.72-1.38)	0.989		
[18.5-25[	830	0.66	1.00 (Ref.)	-		
[25-30[	182	0.68	0.99 (0.85-1.17)	0.946		
≥30	24	0.48	0.64 (0.43-0.97)	0.035	0.150	--
<b>20-year lag</b>						
<b>(FU 2010-2018, 89,389 participants) 598</b>						
<18.5	23	0.89	1.08 (0.71-1.64)	0.718		
[18.5-25[	459	0.83	1.00 (Ref.)	-		
[25-30[	105	0.97	1.12 (0.91-1.39)	0.292		
≥30	11	0.52	0.59 (0.32-1.07)	0.084	0.574	--

BMI, body mass index; IR, age-standardized incidence rate of Parkinson's disease per 1,000 person-years; FU, follow-up. Hazard ratios (HR) and 95% confidence intervals (CI) calculated using Cox proportional hazards models with age as the time scale.

<sup>a</sup> Models are adjusted for baseline place of residence (rural/urban), age at menarche (≤11/12-13/≥14 years) and parity (nulliparous/one child/two children/≥three children), smoking (never/ex/current), menopausal status (no/natural/artificial/unknown type), and physical activity (quartiles).

**Table 17: Association of body mass index (BMI) measured in 1997 (Q5) with Parkinson's disease incidence over the follow-up: analyses adjusted for diet measured in 1993 (Q3).**

BMI (kg/m <sup>2</sup> )	Cases (n)	HR (95% CI) <sup>a</sup>	P- value	HR (95% CI) <sup>b</sup>	P- value
<b>5-year lag (FU 2002-2018, 82,696 participants)</b>					
<18.5	26	1.08 (0.73-1.60)	0.693	1.09 (0.74-1.62)	0.660
[18.5-25[	628	1.00 (Ref.)	-	1.00 (Ref.)	-
[25-30[	203	0.92 (0.78-1.08)	0.291	0.92 (0.78-1.08)	0.307
≥30	39	0.71 (0.51-0.98)	0.040	0.71 (0.52-0.99)	0.043
		<i>P</i> -linear trend	0.028		0.029
		<i>P</i> -quadratic trend	--		--
<b>10-year lag (FU 2007-2018, 81,151 participants)</b>					
<18.5	18	0.95 (0.59-1.52)	0.816	0.95 (0.60-1.53)	0.842
[18.5-25[	502	1.00 (Ref.)	-	1.00 (Ref.)	-
[25-30[	167	0.97 (0.81-1.15)	0.708	0.97 (0.81-1.16)	0.718
≥30	30	0.71 (0.49-1.03)	0.068	0.71 (0.49-1.03)	0.071
		<i>P</i> -linear trend	0.148		0.152
		<i>P</i> -quadratic trend	--		--
<b>15-year lag (FU 2012-2018, 77,172 participants)</b>					
<18.5	11	0.91 (0.50-1.67)	0.770	0.92 (0.50-1.68)	0.792
[18.5-25[	320	1.00 (Ref.)	-	1.00 (Ref.)	-
[25-30[	97	0.92 (0.73-1.16)	0.465	0.91 (0.72-1.15)	0.432
≥30	15	0.58 (0.35-0.98)	0.043	0.58 (0.34-0.97)	0.040
		<i>P</i> -linear trend	0.074		0.063
		<i>P</i> -quadratic trend	--		--

BMI, body mass index; IR, age-standardized incidence rate of Parkinson's disease per 1,000 person-years; FU, follow-up.

Hazard ratios (HR) and 95% confidence intervals (CI) calculated using Cox proportional hazards models with age as the time scale.

<sup>a</sup> Model is adjusted for baseline place of residence (rural/urban), age at menarche (≤11/12-13/≥14 years), parity (nulliparous/one child/two children/≥three children), smoking (never/ex/current), menopausal status (no/natural/artificial/unknown type), physical activity (quartiles), lactose and caffeine intake (quartiles), and adherence to prudent diet (quartiles).

<sup>b</sup> Model is adjusted for baseline place of residence (rural/urban), age at menarche (≤11/12-13/≥14 years), parity (nulliparous/one child/two children/≥three children), smoking (never/ex/current), menopausal status (no/natural/artificial/unknown type), physical activity (quartiles), and total caloric intake (quartiles).

**Table 18: Association of time-varying body mass index with Parkinson's disease incidence over the follow-up: adjustment for potential mediators (hypercholesterolemia, cardiovascular diseases, and diabetes).**

BMI (kg/m <sup>2</sup> )	Model 1 <sup>a</sup>		Model 2 <sup>b</sup>		Model 3 <sup>c</sup>		Model 4 <sup>d</sup>		Model 5 <sup>e</sup>	
	HR (95% CI)	P- value	HR (95% CI)	P- value	HR (95% CI)	P- value	HR (95% CI)	P- value	HR (95% CI)	P- value
<b>5-year lag (FU 1995-2018, 96,702 participants)</b>										
<18.5	0.99 (0.71-1.37)	0.952	0.99 (0.71-1.37)	0.952	0.99 (0.71-1.37)	0.953	0.99 (0.72-1.37)	0.957	0.99 (0.72-1.37)	0.958
[18.5-25[	1.00 (Ref.)	-	1.00 (Ref.)	-	1.00 (Ref.)	-	1.00 (Ref.)	-	1.00 (Ref.)	-
[25-30[	0.89 (0.78-1.03)	0.114	0.89 (0.78-1.03)	0.114	0.90 (0.78-1.03)	0.120	0.89 (0.78-1.02)	0.104	0.89 (0.78-1.03)	0.109
≥30	0.76 (0.59-0.98)	0.032	0.76 (0.59-0.98)	0.032	0.77 (0.60-0.98)	0.035	0.75 (0.59-0.97)	0.026	0.76 (0.59-0.97)	0.029
P-linear trend		0.016		0.016		0.018		0.013		0.014
P-quadratic trend		--		--		--		--		--
<b>10-year lag (FU 2000-2018, 95,381 participants)</b>										
<18.5	0.97 (0.68-1.38)	0.848	0.97 (0.68-1.38)	0.848	0.97 (0.68-1.38)	0.847	0.97 (0.68-1.38)	0.846	0.97 (0.68-1.38)	0.844
[18.5-25[	1.00 (Ref.)	-	1.00 (Ref.)	-	1.00 (Ref.)	-	1.00 (Ref.)	-	1.00 (Ref.)	-
[25-30[	0.86 (0.74-0.99)	0.041	0.86 (0.74-0.99)	0.040	0.86 (0.74-1.00)	0.044	0.86 (0.74-1.00)	0.044	0.86 (0.74-1.00)	0.045
≥30	0.74 (0.57-0.98)	0.036	0.74 (0.56-0.98)	0.036	0.75 (0.57-0.99)	0.041	0.75 (0.57-0.99)	0.043	0.75 (0.57-1.00)	0.046
P-linear trend		0.009		0.009		0.011		0.011		0.012
P-quadratic trend		--		--		--		--		--
<b>20-year lag (FU 2010-2018, 89,758 participants)</b>										
<18.5	0.74 (0.43-1.26)	0.262	0.74 (0.43-1.25)	0.258	0.74 (0.43-1.26)	0.261	0.74 (0.43-1.26)	0.263	0.73 (0.43-1.25)	0.257
[18.5-25[	1.00 (Ref.)	-	1.00 (Ref.)	-	1.00 (Ref.)	-	1.00 (Ref.)	-	1.00 (Ref.)	-
[25-30[	0.98 (0.80-1.21)	0.886	0.98 (0.80-1.20)	0.829	0.99 (0.80-1.21)	0.905	0.98 (0.80-1.21)	0.871	0.98 (0.80-1.20)	0.835
≥30	0.52 (0.30-0.88)	0.016	0.51 (0.30-0.87)	0.014	0.52 (0.31-0.89)	0.017	0.51 (0.30-0.88)	0.015	0.51 (0.30-0.88)	0.015
P-linear trend		0.156		0.136		0.169		0.151		0.144
P-quadratic trend		0.030		0.027		0.033		0.029		0.028

BMI, body mass index; FU, follow-up; IR, age-standardized incidence rate of Parkinson's disease per 1,000 person-years.

Hazard ratios (HR) and 95% confidence intervals (CI) calculated using Cox proportional hazards models for time-varying variables with age as the time scale.

<sup>a</sup> Model 1 is adjusted for baseline place of residence (rural/urban), age at menarche (≤11/12-13/≥14 years) and parity (nulliparous/one child/two children/≥three children), and time-varying smoking (never/ex/current), menopausal status (no/natural/artificial/unknown type), and physical activity (quartiles).

<sup>b</sup> Model 2 = Model 1 + time-varying hypercholesterolemia (Yes/No).

<sup>c</sup> Model 3 = Model 1 + time-varying cardiovascular diseases (Yes/No).

<sup>d</sup> Model 4 = Model 1 + time-varying diabetes (Yes/No).

<sup>e</sup> Model 5 = Model 1 + time-varying hypercholesterolemia (Yes/No), cardiovascular diseases (Yes/No), and diabetes (Yes/No)



## **5.4.2. Abdominal adiposity and Parkinson's disease incidence**

After adjustment for height, women with larger WC ( $\geq 88$  cm) had a decreased, not statistically significant, hazard of PD (HR=0.84, 95% CI=0.70-1.03) compared to those in the reference group ([80-88] cm; Table 19). Women in the lower WC group ( $\leq 80$  cm) also had a lower hazard of PD (HR=0.88, 95% CI=0.76-1.04) that was not statistically significant. Linear and quadratic trends were not statistically significant. In the 10-year lagged analysis, the inverse association became significant for those in the highest WC group (HR=0.74, 95% CI=0.59-0.94), with a significant quadratic trend. Similar results were observed for WHtR. There was no association for WHR in none of the analyses.

Sensitivity analyses using baseline adiposity measures showed a similar pattern, except that both lower and higher WC were significantly associated with lower PD incidence compared to the reference category, with a significant quadratic trend in analyses lagged by 5 and 10 years (Table 20). Further adjustment for diet led to similar conclusions.

**Table 19: Association of time-varying adiposity measures (waist circumference, waist-hip ratio, and waist-height ratio) with Parkinson’s disease incidence over the follow-up.**

Lag	Adiposity measures	Cases (n)	IR	HR (95% CI) <sup>a</sup>	P-value	P-trend	
						Linear	Quadratic
<b>5-year lag (FU 2000-2018, 86,077 participants)</b>	<b>WC, cm</b>	<b>901</b>					
	≤80	484	0.66	0.88 (0.76-1.04)	0.127		
	]80-88]	237	0.78	1.00 (Ref.)	-		
	≥88	180	0.66	0.84 (0.70-1.03)	0.089	0.879	-
	<b>WHR</b>	<b>893</b>					
	<0.85	657	0.68	1.00 (Ref.)	.		
	≥0.85	236	0.70	1.02 (0.88-1.19)	0.760		
	<b>WHtR</b>	<b>901</b>					
	<0.5	474	0.67	0.96 (0.84-1.11)	0.596		
[0.5-0.6[	363	0.72	1.00 (Ref.)	.			
≥0.6	64	0.63	0.85 (0.65-1.10)	0.219	0.633	-	
<b>10-year lag (FU 2005-2018, 83,517 participants)</b>	<b>WC, cm</b>	<b>766</b>					
	≤80	462	0.77	0.89 (0.75-1.06)	0.202		
	]80-88]	190	0.91	1.00 (Ref.)	-		
	≥88	114	0.68	0.74 (0.59-0.94)	0.012	0.251	0.043
	<b>WHR</b>	<b>759</b>					
	<0.85	583	0.77	1.00 (Ref.)	-		
	≥0.85	176	0.84	1.07 (0.90-1.27)	0.449		
	<b>WHtR</b>	<b>766</b>					
	<0.5	442	0.77	0.92 (0.79-1.07)	0.287		
[0.5-0.6[	290	0.86	1.00 (Ref.)	-			
≥0.6	34	0.59	0.63 (0.44-0.90)	0.011	0.373	0.036	

IR, age-standardized incidence rate of Parkinson’s disease per 1,000 person-years; WC, waist circumference; WHR, waist-hip ratio; WHtR, waist-height ratio; FU, follow-up.

Hazard ratios (HR) and 95% confidence intervals (CI) calculated using Cox proportional hazards models for time-varying variables with age as the time scale.

<sup>a</sup>Models are adjusted for baseline place of residence (rural/urban), age at menarche (≤11/12-13/≥14 years) and parity (nulliparous/one child/two children/≥three children), and time-varying smoking (never/ex/current), menopausal status (no/natural/artificial/unknown type) and physical activity (quartiles).

**Table 20: Association of adiposity measures (waist circumference, waist-hip ratio, waist-height ratio) measured in 1995 (Q4) with Parkinson’s disease incidence over the follow-up.**

Lag	Adiposity measures	Cases (n)	IR	HR (95% CI) <sup>a</sup>	P-value	P-trend	
						Linear	Quadratic
<b>5-year lag</b>	<b>WC, cm</b>	<b>781</b>					
<b>(FU 2000-2018, 67,855 participants)</b>	≤80	549	0.67	0.82 (0.69-0.98)	0.027		
	]80-88]	166	0.86	1.00 (Ref.)	-		
	≥88	66	0.56	0.65 (0.49-0.86)	0.003	0.690	0.008
	<b>WHR</b>	<b>778</b>					
	<0.85	667	0.68	1.00 (Ref.)	-		
	≥0.85	111	0.76	1.10 (0.90-1.35)	0.350		
	<b>WHtR</b>	<b>781</b>					
	<0.5	536	0.67	0.92 (0.78-1.07)	0.279		
	]0.5-0.6[	225	0.76	1.00 (Ref.)	-		
≥0.6	20	0.54	0.64 (0.40-1.01)	0.056	0.793	-	
<b>10-year lag</b>	<b>WC, cm</b>	<b>679</b>					
<b>(FU 2005-2018, 66,833 participants)</b>	≤80	480	0.75	0.81 (0.67-0.98)	0.033		
	]80-88]	143	0.98	1.00 (Ref.)	-		
	≥88	56	0.64	0.65 (0.48-0.88)	0.006	0.777	0.015
	<b>WHR</b>	<b>676</b>					
	<0.85	580	0.76	1.00 (Ref.)	-		
	≥0.85	96	0.87	1.12 (0.90-1.40)	0.296		
	<b>WHtR</b>	<b>679</b>					
	<0.5	468	0.76	0.90 (0.76-1.07)	0.220		
	]0.5-0.6[	195	0.87	1.00 (Ref.)	-		
≥0.6	16	0.59	0.60 (0.36-1.00)	0.051	0.867	-	

IR, age-standardized incidence rate of Parkinson’s disease per 1,000 person-years; WC, waist circumference; WHR, waist-hip ratio; WHtR, waist-height ratio; FU, follow-up.

Hazard ratios (HR) and 95% confidence intervals (CI) calculated using Cox proportional hazards models with age as the time scale.

<sup>a</sup> Models are adjusted for baseline place of residence (rural/urban), age at menarche (≤11/12-13/≥14 years), parity (nulliparous/one child/two children/≥three children), smoking (never/ex/current), menopausal status (no/natural/artificial/unknown type), and physical activity (quartiles).

### 5.4.3. Trajectories of BMI in PD patients and controls

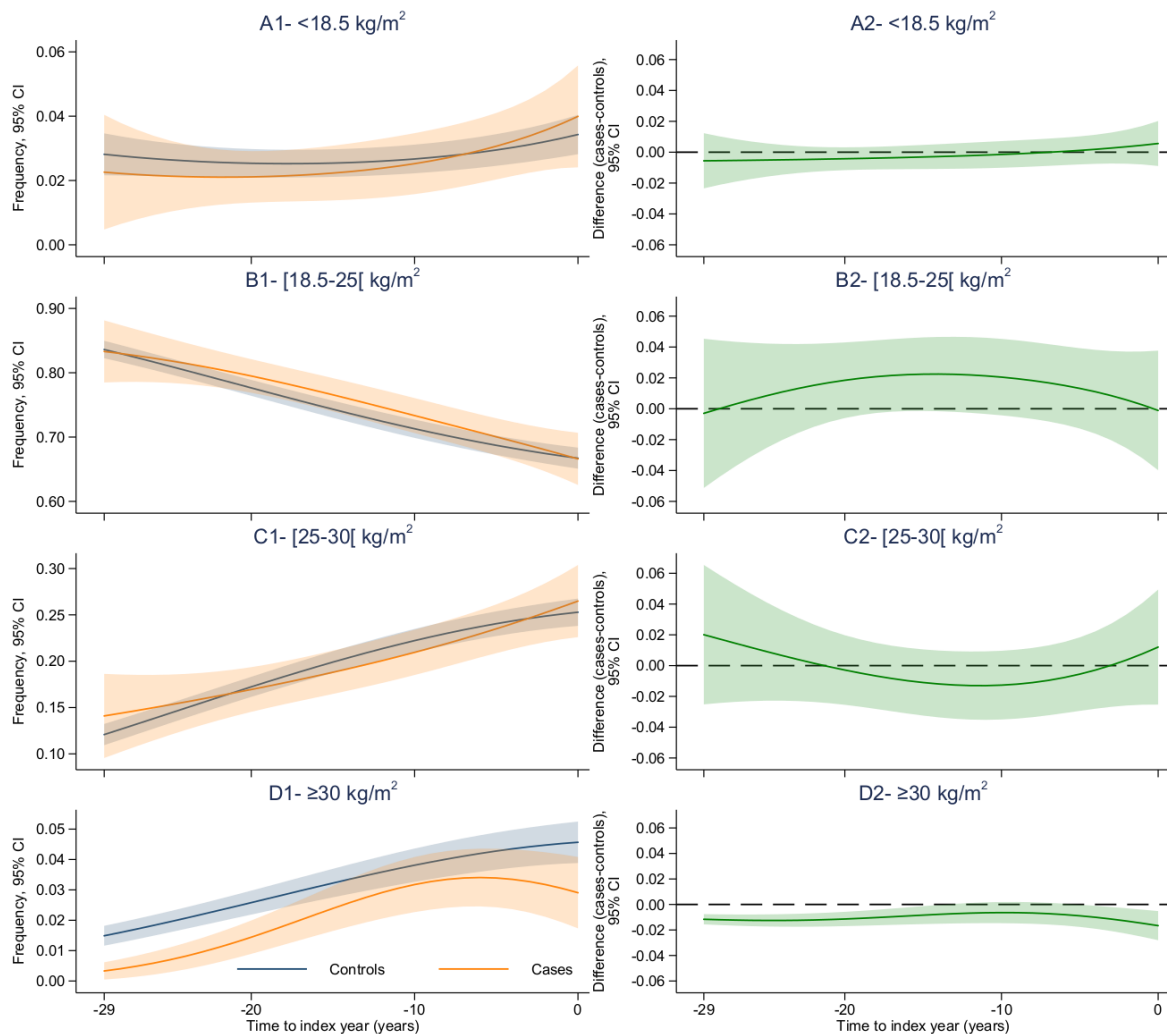
Figure 11 shows the trajectories of the frequencies (panels A1-D1) of the four BMI categories in PD patients and controls and corresponding case-control differences (panels A2-D2) using a retrospective time scale over 29 years of follow-up based on a multinomial GEE logistic model; regression coefficients are shown in Annex 3.

The frequency of obesity steadily increased among controls over the follow-up, while it started to decrease five to 10 years prior to diagnosis in patients ( $P$ -interaction=0.019; Figure 11-D1). Obesity was significantly less frequent in patients than in controls every year, even 29-years before the index date ( $T_0$ ; Figure 11-D2). For instance, in women with 73 years old at the index date, the prevalence of obesity was of 0.29% in patients and 1.31% in controls 29 years before the index date (difference=-1.01, 95% CI=-0.016,-0.008,  $P<0.001$ ) and it was 2.54% in patients and 4.00% in controls at the index date (difference=-1.46, 95% CI=-0.028,-0.005,  $P=0.004$ ) (Figure 11-D1 and D2).

There were no statistical differences between patients and controls for the trajectories of other BMI groups (A-Underweight, B-Normal weight, C-Overweight); the frequencies of each group followed similar trajectories in cases and controls and differences were not statistically significant as confidence intervals included 0 at all times.

Sensitivity analyses using restricted cubic splines to model retrospective time yielded similar results (Annex 4).

**Figure 11: Trajectories of the frequency of the four categories of body mass index in cases and controls up to 29 years before the index date.**



The four panels on the left (A1-D1) show the frequencies (solid lines) and 95% confidence intervals (CI, shaded areas) of the four BMI categories in cases (orange) and controls (blue) according to retrospective time before the index date ( $T_0$ , time=0).

The four panels on the right (A2-D2) show the corresponding case-control differences of the frequencies of the four BMI groups according to retrospective time, together with their 95% CI. CIs that do not include 0 correspond to statistically significant differences between cases and controls.

Frequencies were predicted by a multinomial GEE logistic model, using a retrospective time scale, with  $T_0$  representing the age at PD diagnosis in cases and age at the index date in matched controls. The model includes a quadratic function of retrospective time, case-control status, and the interactions of case-control status with time. The model was adjusted for baseline age at menarche, parity, and place of residence, and time-varying smoking, menopausal status, and physical activity. The model's coefficients are shown in Annex 3.

Trajectories were plotted for the most common profile of E3N participants: 73 years of age at the index date ( $T_0$ ), never smokers, age at menarche at 12-13 years, natural menopause, 2 children, living in urban areas, third quartile (Q3) of physical activity.

## 5.5. Discussion

In the E3N cohort study of ~100,000 women with repeated assessments of anthropometric measures over 29 years of follow-up, obesity was associated with lower PD incidence, even when assessed 20 years before PD diagnosis. There was a similar pattern for higher WC and WHtR, with significant associations in analyses lagged by 10 years. Analyses of BMI trajectories showed that obesity was less frequent in PD patients than in controls 29 years before the index date, and that BMI decreased in PD patients 5-10 years prior to PD diagnosis.

Previous cohorts that examined the association between midlife BMI and PD yielded conflicting findings (Table 3). None of the studies used time-varying BMI, and, therefore, did not take into account the influence of BMI changes over the follow-up. Our analyses using baseline BMI yielded an association estimate of the same size as in the Nurses' Health Study (202 PD patients; 22 years of follow-up) that reported a HR of 0.7 (95% CI=0.4-1.2) for women with obesity.[96] As part of the NeuroEPIC4PD study (378 male PD patients, 356 female PD patients; 20 years of follow-up), analyses using a 5-year lag yielded an HR of 0.71 (95% CI=0.46-1.08) in men, similar to our findings, while the HR was of 1.00 (95% CI=0.67-1.50) in women.[95] Another study based on healthcare databases from Korea (9-year follow-up) also reported that obesity was associated with lower PD incidence in analyses in men and women combined (HR 5-year lag=0.80, 95% CI=0.74-0.86).[99] Of two other studies that reported results of analyses in women, a smaller study from Finland reported an increased risk of PD for those overweight/obese compared with those with a BMI <23 kg/m<sup>2</sup>,[103] and the other reported no association between obesity and PD.[97] Given the low prevalence of obesity, very large studies with a long follow-up are needed to examine its association with PD, and some smaller studies may have been underpowered.

Reverse causation represents a major threat to study exposures associated with PD because exposures may be affected by motor and non-motor symptoms during the

prodromal phase of PD. The preclinical phase of PD has been suggested to be as long as 20 years or longer in some patients, and loss of initiative, reward processing disruption, constipation, changes in eating behaviour, and olfactory dysfunction may be present.[102, 180] Weight loss is common in PD patients,[89] and previous studies showed that weight loss starts before PD diagnosis.[68, 175] We also showed that the frequency of obesity decreased 5-10 years prior to diagnosis. Some previous studies, but not all, addressed this issue by excluding participants who developed PD during the first years of the follow-up; they generally used lags  $\leq 5$  years,[95-97, 99, 103-105] and only two studies used longer lags.[85, 106] We were able to address this issue thanks to the long follow-up of the cohort and because the majority of patients developed PD after 10 years of follow-up, due to the relatively young age at cohort inception.[134] Hence, we were able to perform lagged analyses while retaining a sufficient number of PD patients using longer lags than previous studies, and confirmed our main finding in analyses lagged by 20 years.[102]

Prospective studies on the association between abdominal adiposity and PD incidence yielded conflicting findings. While most studies found no association between WC and PD risk,[95-98] higher WC was associated with higher PD risk in two studies from South Korea based on the same healthcare databases with follow-ups comprised between 6-9 years.[99, 100] In women, abdominal adiposity is strongly influenced by menopause, and no study adjusted for menopausal status in women. WHR is less correlated with body fat than other adiposity measures and BMI/WC outweighed WHR to identify components of metabolic syndrome, which could explain the lack of association for WHR.[181]

The association between obesity and lower incidence of PD is consistent with studies that used MR to examine the association between BMI or adiposity traits and PD and showed that PD risk decreased with increasing BMI. Given that a set of assumptions are met, MR uses genetic instruments associated with an exposure to estimate exposure-disease associations that are not biased by confounding or reverse

causation. Two two-sample MR studies were consistent in showing that increasing BMI was associated with lower PD risk, in favour of a protective causal association.[93, 94] Another MR study reported inverse associations between nine adiposity traits and PD.[101] One limitation of two-sample MR, however, is that it cannot examine non-linear relations. In addition, MR results can be biased for time-varying exposures,[182] in particular if associations between genetic instruments and exposures change with age, as it has been shown for BMI.[182, 183] We aimed at triangulating the evidence on the association between BMI and PD, by comparing results from different approaches that have different and unrelated key sources of potential bias.[184]

Obesity is associated with insulin resistance and compensative hyperinsulinemia.[107] Insulin crosses the blood-brain-barrier and there is evidence to suggest that it influences a multitude of neuroprotective pathways in the brain including the promotion of neuronal survival and dopaminergic transmission.[102, 108, 109] In addition, the Insulin/IGF1 signalling pathway contributes to the control of neuronal excitability, and growing evidence suggests that its dysfunction contributes to the progressive loss of neurons in PD.[185] Although increasing circulating insulin levels represents a plausible explanation, the temporal relation between peripheral and brain insulin resistance remains unclear, and additional research is needed to better understand the inverse association between obesity and lower incidence of PD.[102] In addition, increasing evidence supports a role of microbiota in PD etiology,[110] and obesity is also characterized by changes in microbiota.[111] Additional studies are needed to examine whether obesity related changes in microbiota could play a role in PD. Finally, our analyses showing that adjustment for hypercholesterolemia, cardiovascular disease, and diabetes did not change the association between obesity and PD suggest that these characteristics are unlikely to be strong mediators of this association.

Our findings suggest different roles of obesity in PD and dementia. Obesity has been associated with worse cognitive outcomes through neuro-inflammation and



altered metabolism of free fatty acids.[186] Additional studies are needed to elucidate the mechanisms underlying different associations in PD and dementia.

The main strengths of the study are its long follow-up and large size, which allowed us to perform lagged analyses in order to address the potential for reverse causation while retaining a large number of patients. Ours is the largest study in women to have examined the association between BMI and other adiposity traits in relation to PD incidence. Our approach to ascertain PD patients yielded incidence rates that are comparable to those in women of Western Europe;[134] this is in favour of the validity of our approach and suggests that the selected nature of the cohort did not lead to lower PD rates, likely because PD onset was long after inclusion in the cohort. BMI was measured several times over the follow-up which allowed us to use time-varying Cox models to take into account changes in BMI over the follow-up, and to compare BMI trajectories in patients and controls.

Our study also has limitations. First, E3N participants are mostly educated and health-conscious teachers who are not representative of the general population and have a lower frequency of obesity. Due to the relatively low frequency of obesity, analyses according to severity of obesity were not possible. However, it is generally considered that representativeness is not essential for estimating associations, and associations in occupational cohorts are not necessarily different compared to those estimated in the general population.[173, 174] Only women were included, which hampers generalizability to men; however, women represent an understudied population in PD research, and additional studies are necessary in women.[132] Second, anthropometric characteristics were self-reported which may have led to exposure misclassification; however, it is likely to be non-differential with respect to PD status and bias associations towards the null. Self-reported WC had the lowest correlation with clinical assessments by a skilled technician, compared with all the other measures, although it remained relatively high (correlation=0.79). Typically, WC is underreported by ~1-3 cm, with larger circumference being associated with larger

under-reporting,[187] which may lead to underestimate the association between WC and PD. Third, WC and HC were first assessed at Q4 (1995) and analysis with lags longer than 15 years were not possible. Last, because BMI and WC/WHtR were highly correlated, we were unable to examine their individual and joint effects due to collinearity. Alternative approaches would require using more sophisticated techniques that accurately measure body fat (e.g., dual energy X-ray absorptiometry).

In conclusion, obesity was associated with lower PD incidence in women, even when BMI was measured long before diagnosis, in agreement with recent MR studies. These findings warrant further investigations to understand the underlying mechanisms. Our analyses underscore the importance of lagged analyses due to changes in BMI over prodromal PD.



## **6. Objective 3: Association between diabetes and Parkinson's disease**



## 6.1. Introduction

Several cohort studies have examined the relationship between diabetes and PD, with conflicting results. A meta-analysis of nine prospective studies showed that diabetes was associated with 29% increased PD incidence in men and women combined (RR=1.29, 95% CI=1.15-1.45), although with marked heterogeneity ( $I^2=93.9\%$ ,  $P$ -heterogeneity $\leq 0.001$ ).[115] This positive association was also found in women (RR=1.50, 95% CI=1.07-2.11) and men separately (RR=1.40, 95% CI=1.17-1.67).[188] Heterogeneity may be partly explained by differences in study design and in analyses, including duration of follow-up, PD ascertainment, diabetes assessment, and adjustment for confounders. Of the nine studies, five were prospective cohort studies with in person assessments, while four relied on healthcare databases. All prospective cohort studies performed some type of validation of PD and diabetes diagnoses, except one in which PD diagnoses were based on drug claims only and not validated.[125] In studies based on healthcare databases, PD and diabetes diagnoses were based on different algorithms and were not validated. Of the prospective studies, follow-up was longer than 15 years in three studies;[70, 125, 126] of the studies based on healthcare databases, three had a follow-up of 8-12 years,[119, 121, 127] and one did not report the date of the end of follow-up.[129] Adjustment for confounders tended to be more complete in prospective cohort studies than in those based on healthcare databases.

Alternatively, a meta-analysis of 14 case-control studies showed that diabetes was associated with 49% lower PD risk (OR=0.75, 95% CI=0.58-0.98), again with marked heterogeneity ( $I^2=75\%$ ,  $P$ -heterogeneity= $<0.001$ ).[116]

As discussed in the Introduction (§1.3), these findings need to be interpreted with caution due to a number of potential biases, including surveillance bias, reverse causation, residual confounding, diagnostic misclassification, survival bias, and inadequate definition of the date of PD onset in healthcare databases. Further studies with a long follow-up that examine the role of incident diabetes, duration of diabetes,

and use of medical care are needed to address these issues.

Our aim was to examine the association of time-varying diabetes and PD incidence in women from the French E3N cohort study. We performed analyses based on incident diabetes to address the potential for survival bias, and analyses with a lag and according to diabetes duration to tackle the potential for reverse causation. We used a nested case-control design to compare trajectories of diabetes prevalence prior to PD in patients and matched controls. We also performed a meta-analysis of previous cohort studies in order to examine potential causes of heterogeneity.

## **6.2. Methods**

### **6.2.1. Ascertainment of diabetes**

The ascertainment of women with diabetes is described in Figure 12. Several previous studies have examined in E3N the relation of body silhouettes,[189, 190] wine consumption,[191] or diet,[192-197] with diabetes. Potential diabetic patients were identified using two data sources: self-reports in the E3N questionnaires up to 2003; MGEN drug claims databases (ATC drug codes A10A, A10B, and A10X except Benfluorex that is no longer marketed in France) between 01/01/2004 until 31/12/2016.

Until 2004, women who self-reported diabetes at least once (use of anti-diabetic drugs, diet for diabetes, hospitalization for diabetes) were considered as potential cases of diabetes (N=4,289). Among them, 2,315 women had  $\geq 1$  reimbursement for anti-diabetic medications in 2004 or later in drug claims databases, and were therefore considered as having diabetes. Among 1,974 women without reimbursement of anti-diabetic drugs, those alive with an accurate address (N=1,735) were sent a diabetes-specific questionnaire to obtain further information (diagnosis date, symptoms, fasting or random glucose and glycated hemoglobin [HbA1c] concentrations at diagnosis and at the date of the questionnaire, current therapy). Among 1,480 women who completed the questionnaire, 342 diabetic cases were confirmed and met at least one of the

following criteria:

- Use of anti-diabetic drugs;
- Fasting glucose  $\geq 7.0$  mmol/l or random glucose  $\geq 11.1$  mmol/l concentrations at diagnosis (WHO recommendations);[198]
- Fasting glucose  $\geq 7.0$  mmol/l and/or HbA1c  $\geq 7\%$  concentrations at the time of the questionnaire.[199, 200]

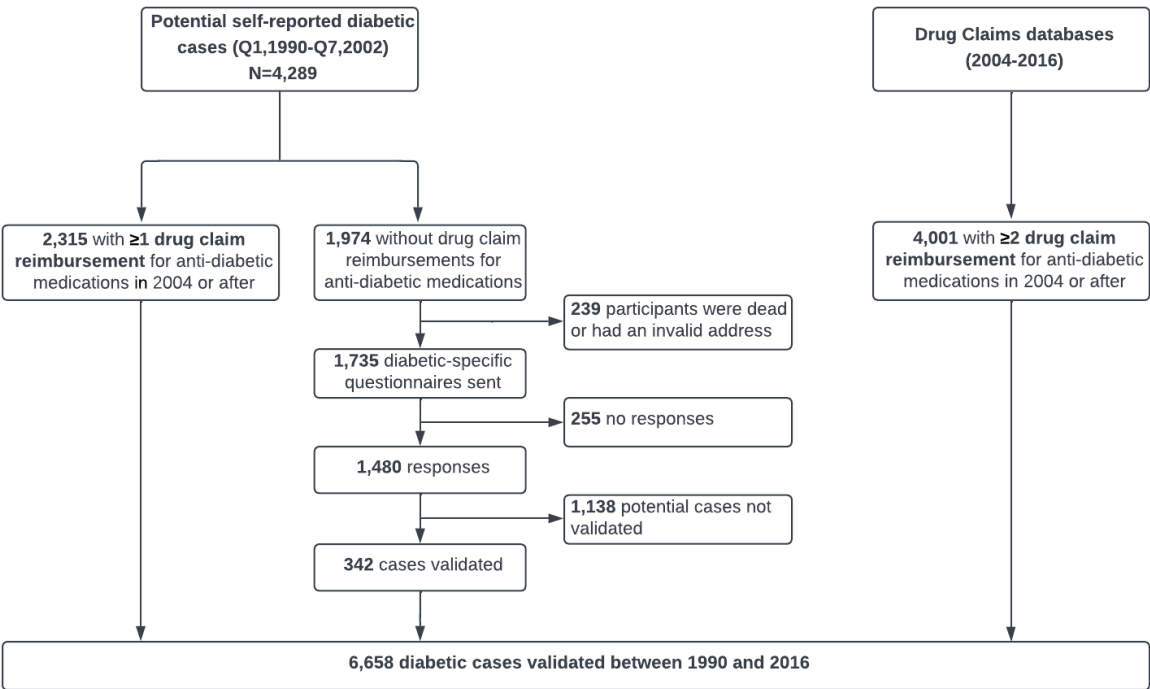
A total of 2,657 self-reported diabetes cases were therefore validated before 2004 (Figure 12). The date of diagnosis was defined as the self-reported date in E3N questionnaires. If no date was available, we attributed January 1<sup>st</sup> of the year of diagnosis reported in the diabetic-specific questionnaire. Potential cases that were not validated through diabetic-specific questionnaires due to death/invalid address ( $n=239$ ) or no response ( $n=255$ ) were considered in our main analysis as non-diabetic participants and were excluded in sensitivity analysis.

A second set of potential cases of diabetes were identified through MGEN drug claims databases in 2004 or later. A total of 4,001 women had  $\geq 2$  drug reimbursements within a year period for anti-diabetic drugs, and were considered as having diabetes. The date of the first reimbursement was used as the date of diabetes diagnosis.

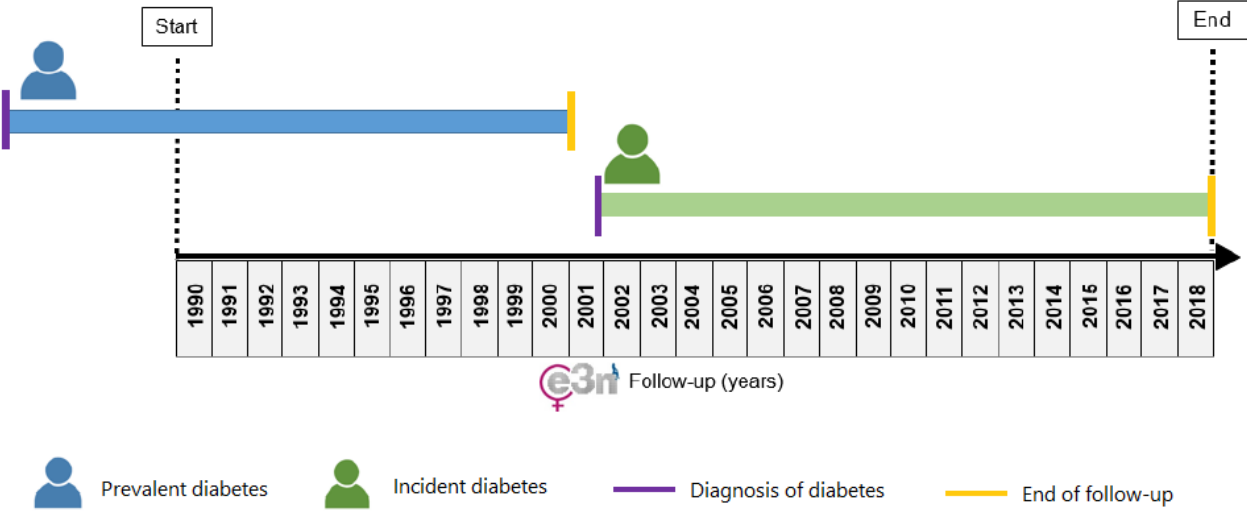
In total, we identified 6,658 diabetic women between 1990 and 2016. Participants with diabetes diagnosed before the start of the follow-up were considered as prevalent cases and those who were diagnosed during the follow-up were considered as incident cases (Figure 13). Although the validation process did not allow to distinguish type 1 diabetes (T1DM) from T2DM, the age range of our study population implies that incident cases of diabetes mostly have T2DM.



**Figure 12: Ascertainment of diabetes in the E3N cohort study.**



**Figure 13: Prevalent cases and incident cases of diabetes.**



The two horizontal lines in blue and green represent the duration of diabetes in two patients, one with incident and one with prevalent diabetes. The first participant (in blue) was diagnosed with diabetes before entering the study and was followed until 2000. The second participant (in green) was diagnosed with diabetes in 2001 and was censored in 2018.

## **6.2.2. Nested case-control study**

In order to examine trajectories of diabetes prevalence prior to diagnosis in PD patients and matched controls, we set-up a case-control nested within E3N. Each incident PD patient was individually matched to 20 controls using incidence density sampling.[161] To be selected as controls, participants had to be alive and at risk of PD at the date of diagnosis of matched cases (index date,  $T_0$ ), and have the same age at  $T_0$ .

The final sample included 25,200 women (1,200 PD patients, 24,000 controls).

## **6.3. Statistical Analysis**

### **6.3.1. Survival Analysis**

We used Cox proportional hazards regression with age as the time scale to estimate HRs, 95% CIs, and two-tailed  $P$ -values ( $\alpha=0.05$ ). For our main analyses based on a 29-years follow-up, we did not include a lag between exposure and PD incidence; for these analyses, we examined the association of time-varying diabetes status, duration of diabetes, and age at onset of diabetes with the incidence of PD. We also performed additional analyses for diabetes with and without cardiovascular disease (coronary heart disease, stroke) as a proxy for diabetes severity.

We then examined the association between time-varying diabetes status and PD incidence while including lags of 5, 10 and 15 years. In analyses with a 5-year lag, the follow-up started five years after the baseline assessment and participants who developed PD over the first five years of follow-up were excluded. Given the 5-year lag, women were followed since 1995 until PD diagnosis or end of follow-up (latest available questionnaire and drug reimbursement). A similar approach was used for longer lags.

In order to examine the role of incident diabetes, we repeated these analyses

after excluding women with prevalent diabetes (diagnosed before the beginning of the follow-up in 1990).

Our main analyses were adjusted for baseline parity, age at menarche, and place of residence, and time-varying smoking, menopausal status, physical activity, BMI, hypertension, and hypercholesterolemia. Missing values were coded as specific categories to retain the same number of participants in all analyses.

Since MGEN drug claims databases were available starting in 2004, we also performed analyses after that date with further adjustment for the number of healthcare visits per month. We were also able to adjust these analyses for use of lipophilic statins as previous studies showed an inverse association between statins use and PD,[179] and statins can be used for primary prevention of cardiovascular disease in persons at risk (e.g., dyslipidemia, hypertension, smoking, diabetes).[201]

### **6.3.2. Trajectories**

Trajectories of diabetes prevalence over 29 years of follow-up were examined within a nested case-control study with a retrospective time scale. We modelled time-varying diabetes status (binary dependent variable, Yes vs No) using a GEE logistic regression model that allows us to take into account repeated observations for a given subject. We started by including the following covariates in the model: PD status, time (in years, divided by 10), time squared (to allow for non-linear changes over time), age at  $T_0$  (centered at 73 years), age squared (to allow for non-linear changes with age), and two-way interactions between PD status, time, and age; we then performed a backwards stepwise selection of variables ( $P < 0.05$ ). The final model included the following terms: intercept, PD status, time,  $time^2$ , two-way interactions between time and PD status ( $time \times PD$ ,  $time^2 \times PD$ ), age,  $age^2$ . The model was adjusted for potential confounders of the association between diabetes and PD: baseline parity, age at menarche and place of residence, and time-varying smoking, physical activity, menopausal status, BMI,

hypertension and hypercholesterolemia. Sensitivity analyses were conducted using restricted cubic splines.

Based on this model, we compared the frequency of diabetes before in cases and controls before the index date, and estimated the annual difference between the two groups.

### **6.3.3. Meta-analysis**

We used a random effects model to compute pooled relative risks (RR, 95% CIs) for the association between diabetes and PD risk. Heterogeneity between studies was evaluated using the  $I^2$  statistic that measures how much of the heterogeneity is due to between study variation rather than by chance. A formal test of heterogeneity was performed using the Q test.

We identified all cohort studies on the relation between diabetes and PD based on previous meta-analyses[114, 115, 188, 202-205] and a PubMed search, and updated previous pooled estimates by including our own findings. We performed separate analyses for prospective cohort studies with in-person assessments and those based on healthcare databases. We also examined the influence of studies that included a lag between diabetes and PD incidence or that examined the role of diabetes duration.

### **6.3.4. Statistical power**

We computed the statistical power of our survival analyses to detect an association for an exposure with 7% frequency and a type 1 error of 5% for different effect sizes.

The minimum detectable HR with a statistical power of 80% was of 1.36; the statistical power to detect HRs of 1.1, 1.2, 1.3, 1.4, and 1.5 was of 14%, 38%, 66%, 86%, and 95% respectively.

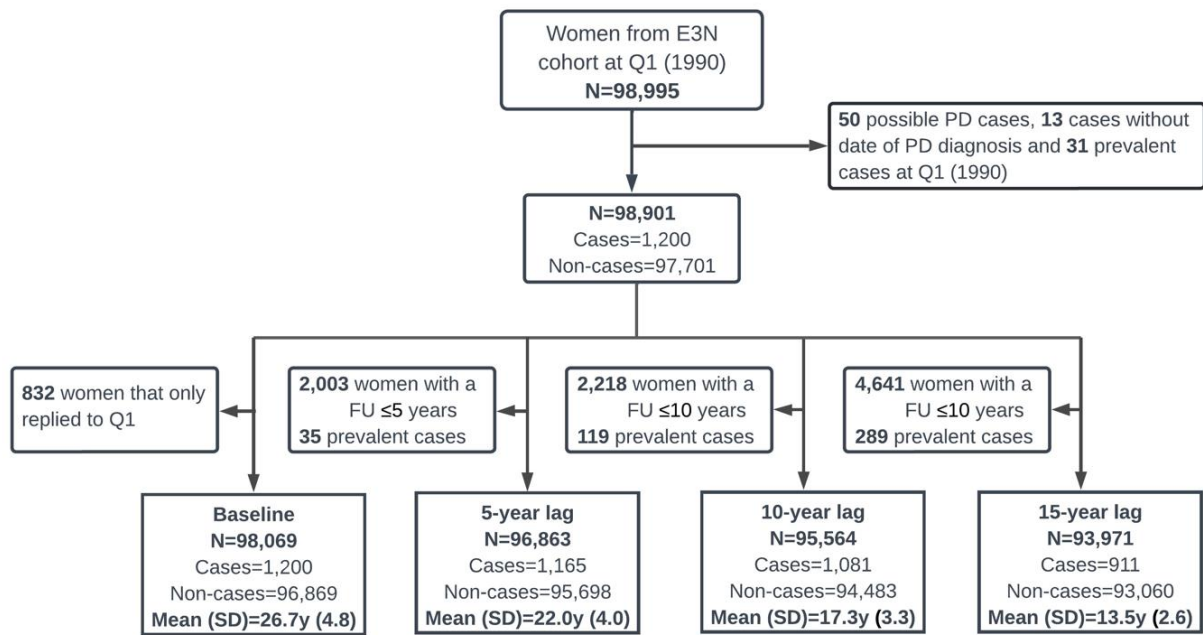
## 6.4. Results

Figure 14 describes the inclusion of participants into the study. We excluded 50 possible PD cases, 13 cases without a diagnosis date, and 31 prevalent cases at Q1. In our main analyses, 839 women were excluded because they were not followed after Q1 (1990), leaving 98,069 women followed over 29 years (mean follow-up=26.7, SD=4.8) of whom 1,200 women developed PD. Compared to previous studies in women, ours is the third largest in terms of PD patients, after one study based on the Spanish information system for Research in Primary Care (5,975 women with PD)[131] and another based on the National Health Insurance claim data of Taiwan (1,324 women with PD).[127]

Table 21 presents participants baseline (1990) characteristics, overall and according to diabetes status at the end of follow-up. The mean age at Q1 was 49.4 years (SD=6.7) and 0.75% ( $n=735$ ) of the participants were diabetic at that time. The frequency of diabetes increased over the follow-up, and ~7% ( $n=6,654$ ) of the participants had diabetes at the end of the follow-up, when they were 76.1 years (SD=7.3) old on average.

Women with diabetes at the end of the follow-up were older at baseline, less physically active and educated, and more frequently menopausal; they had higher BMI, a more frequent history of hypercholesterolemia and hypertension, an earlier age at menarche, and more children (Table 21). Women with incident PD were older, less frequently smokers, practiced lower levels of physical activity, had later age at menarche, more children, and were more frequently menopausal and hypertensive (Table 22).

**Figure 14: Flow-chart for inclusion into the study.**



**Table 21: Participants' characteristics at baseline (1990-Q1) overall and according to diabetes at the end of the follow-up.**

Characteristics, <i>n</i> (%)	Total 98,069	Diabetes	
		No 91,415 (93.2)	Yes 6,654 (6.8)
<b>Age (y), mean (SD)</b>	49.4 (6.7)	49.3 (6.6)	50.9 (6.7)
<45	33,072 (33.7)	31,451 (34.4)	1,621 (24.4)
[45:50[	23,797 (24.3)	22,213 (24.3)	1,584 (23.8)
[50:55[	19,307 (19.7)	17,869 (19.5)	1,438 (21.6)
[55:60[	13,078 (13.3)	11,862 (13.0)	1,216 (18.3)
≥60	8,815 (9.0)	8,020 (8.8)	795 (11.9)
<b>Education</b>			
<High school	13,091 (13.9)	11,803 (13.5)	1,288 (20.3)
≥High school	80,809 (86.1)	75,762 (86.5)	5,047 (79.7)
<i>Missing</i>	4,169	3,850	319
<b>BMI (kg/m<sup>2</sup>), mean (SD)</b>	22.6 (3.2)	22.4 (2.9)	25.8 (4.4)
<18	4,073 (4.3)	4,021 (4.5)	52 (0.9)
[18-25[	74,699 (78.1)	71,556 (80.2)	3,143 (48.9)
[25-30[	13,904 (14.5)	11,719 (13.1)	2,185 (34.0)
≥30	2,927 (3.1)	1,885 (2.2)	1,042 (16.2)
<i>Missing</i>	2,466	2,234	232
<b>Physical activity</b>			
Q1	24,170 (25.0)	22,209 (24.7)	1,961 (30.0)
Q2	24,170 (25.0)	22,571 (25.0)	1,599 (24.4)
Q3	24,170 (25.0)	22,635 (25.1)	1,535 (23.5)
Q4	24,170 (25.0)	22,725 (25.2)	1,445 (22.1)
<i>Missing</i>	1,389	1,275	114
<b>Smoking</b>			
Never	52,623 (54.1)	49,042 (54.1)	3,581 (54.4)
Ex	29,978 (30.8)	28,043 (30.9)	1,935 (29.4)
Current	14,612 (15.1)	13,542 (15.0)	1,070 (16.2)
<i>Missing</i>	856	788	68
<b>Age at menarche (y), mean (SD)</b>	12.8 (1.4)	12.8 (1.4)	12.6 (1.5)
≤11	20,046 (21.0)	18,367 (20.6)	1,679 (26.0)
[12-13]	48,284 (50.4)	45,166 (50.6)	3,118 (48.3)
≥14	27,347 (28.6)	25,686 (28.8)	1,661 (25.7)
<i>Missing</i>	2,392	2,196	196
<b>Menopausal status</b>			
Premenopausal	53,999 (57.3)	50,991 (58.1)	3,008 (46.9)
Natural menopause	32,865 (34.9)	30,236 (34.4)	2,629 (41.0)
Artificial menopause	6,550 (7.0)	5,859 (6.7)	691 (10.8)
Unknown type	775 (0.8)	690 (0.8)	85 (1.3)
<i>Missing</i>	3,880	3,639	241
<b>Parity</b>			
Nulliparous	11,872 (12.2)	11,065 (12.3)	807 (12.2)
One child	15,815 (16.3)	14,800 (16.3)	1,015 (15.4)
Two children	41,179 (42.3)	38,671 (42.6)	2,508 (38.1)
≥3 children	28,406 (29.2)	26,149 (28.8)	2,257 (34.3)
<i>Missing</i>	797	730	67
<b>Place of residence</b>			
Rural	13,281 (14.8)	12,352 (14.8)	929 (15.2)
Urban	76,342 (85.2)	71,150 (85.2)	5,192 (84.8)
<i>Missing</i>	8,446	7,913	533
<b>Hypercholesterolemia</b>			
No	86,084 (87.8)	81,301 (88.9)	4,783 (71.9)
Yes	11,985 (12.2)	10,114 (11.1)	1,871 (28.1)
<b>Hypertension</b>			
No	86,562 (88.3)	81,293 (88.9)	5,269 (79.2)
Yes	11,507 (11.7)	10,122 (11.1)	1,385 (20.8)

SD, Standard deviation; BMI, body mass index; Q, quartile; y, years.

**Table 22: Participants' characteristics at baseline (1990) according to Parkinson's disease status at the end of the follow-up.**

Characteristics, <i>n</i> (%)	Parkinson's Disease	
	No	Yes
<b>Age (y), mean (SD)</b>	96,869 (98.8)	1,200 (1.2)
<45	32,887 (33.9)	185 (15.4)
[45:50[	23,558 (24.3)	239 (19.9)
[50:55[	18,983 (19.6)	324 (27.0)
[55:60[	12,821 (13.2)	257 (21.4)
≥60	8,620 (9.0)	195 (16.3)
<b>Education</b>		
<High school	12,934 (13.9)	157 (13.7)
≥High school	79,821 (86.1)	988 (86.3)
<i>Missing</i>	4,114	55
<b>BMI (kg/m<sup>2</sup>), mean (SD)</b>	22.6 (3.2)	22.9 (3.0)
<18.5	4,033 (4.3)	40 (3.4)
[18.5-25]	73,793 (78.1)	906 (77.3)
[25-30]	13,706 (14.5)	198 (16.9)
≥30	2,899 (3.1)	28 (2.4)
<i>Missing</i>	2,438	28
<b>Physical activity</b>		
Q1	23,832 (25.0)	338 (28.8)
Q2	23,886 (25.0)	284 (24.2)
Q3	23,903 (25.0)	267 (22.9)
Q4	23,887 (25.0)	283 (24.1)
<i>Missing</i>	1,361	28
<b>Smoking</b>		
Never	51,897 (54.0)	726 (61.1)
Ex	29,643 (30.9)	335 (28.2)
Current	14,484 (15.1)	128 (10.7)
<i>Missing</i>	845	11
<b>Age at menarche (y), mean (SD)</b>	12.8 (1.4)	12.9 (1.5)
≤11	19,791 (21.0)	255 (21.9)
[12-13]	47,752 (50.5)	532 (45.6)
≥14	26,968 (28.5)	379 (32.5)
<i>Missing</i>	2,358	34
<b>Menopausal status</b>		
Premenopausal	53,566 (57.6)	433 (37.1)
Natural menopause	32,292 (34.7)	573 (49.0)
Artificial menopause	6,408 (6.9)	142 (12.2)
Unknown type	755 (0.8)	20 (1.7)
<i>Missing</i>	3,848	32
<b>Parity</b>		
Nulliparous	11,732 (12.2)	140 (11.8)
One child	15,660 (16.3)	155 (13.1)
Two children	40,694 (42.4)	485 (40.8)
≥ 3 children	27,999 (29.1)	407 (34.3)
<i>Missing</i>	784	13
<b>Place of residence</b>		
Rural	13,144 (14.8)	137 (12.5)
Urban	75,383 (85.2)	959 (87.5)
<i>Missing</i>	8,342	104
<b>Hypercholesterolemia</b>		
No	85,045 (87.8)	1,039 (86.6)
Yes	11,824 (12.2)	161 (13.4)
<b>Hypertension</b>		
No	85,540 (88.3)	1,022 (85.2)
Yes	11,329 (11.7)	178 (14.8)

SD, Standard deviation; BMI, body mass index; Q, quartile; y, years.



### **6.4.1. Diabetes and Parkinson's disease incidence**

Sixty-three diabetic women developed PD. Table 23 shows that diabetes (prevalent and incident) was not associated with PD incidence (HR=1.04, 95% CI=0.80-1.35) in analyses without a lag; including lags of increasing duration did not change the results (Table 24). Duration of diabetes and age at diabetes did not play a role (Table 23). Neither diabetes without and with cardiovascular disease were associated with PD incidence ( $P$ -interaction=0.800).

Analyses restricted to the 2004-2018 period when the drug claims databases became available yielded similar findings and adjustment for the number of medical visits or statins use did not change the results (Table 23).

Similarly, in analyses based on incident diabetes identified during the follow-up (1990-2018), diabetes was not associated with PD incidence (HR=0.98, 95% CI=0.73-1.32, Table 25) and lagged analyses yielded similar results (Table 26). Duration of diabetes or age at diabetes onset did not play a role (Table 25).

In sensitivity analyses, we excluded a small number of potential cases of diabetes that were not validated through diabetic-specific questionnaires due to death/invalid address ( $n=239$ ) or no response ( $n=255$ ) and reached similar conclusions (data not shown).

### **6.4.2. Trajectories of diabetes in PD patients and controls**

Figure 15 shows the trajectories of the frequency of diabetes in PD patients and controls (panel A) and case-control differences (panel B), using a retrospective time scale over 29 years of follow-up based on a GEE logistic regression model (Annex 5). There was no difference in the trajectories of diabetes in PD patients and controls. At the index date, the prevalence of diabetes was 1.76% in patients and 1.49% in controls

( $P=0.4531$ ) with the most common profile of E3N participants. Sensitivity analyses using restricted cubic splines to model time yielded similar results (data not shown).

### **6.4.3. Meta-analysis**

The meta-analysis of the 14 studies, including eight studies based on healthcare databases and six prospective cohort studies, showed that diabetes was associated with higher PD risk overall (RR=1.35, 95% CI=1.33-1.37) with high heterogeneity ( $I^2=93.6\%$ ,  $P$ -heterogeneity<0.001). The association was slightly weaker in prospective studies (RR=1.26, 95% CI=1.13-1.40;  $I^2=60.0\%$ ,  $P$ -heterogeneity=0.029) than in those based on healthcare databases (RR=1.35, 95% CI=1.34-1.37;  $I^2=96.3\%$ ,  $P$ -heterogeneity<0.001), but the difference was not statistically significant ( $P=0.185$ , Figure 16 – Panel A). After excluding the prospective cohort study in which PD cases were identified based on drug claims,[125] the association in prospective studies decreased (RR=1.22, 95% CI=1.10-1.37) and the difference between the two types of studies was closer to significance ( $P=0.078$ , Figure 16 – Panel B). Only one prospective study found a positive and significant association between diabetes and PD.[120]

Three prospective cohort studies and one study based on healthcare databases performed lagged analyses or took into account duration of diabetes. In prospective studies, the association decreased by 7% (HR=1.19, 95% CI=1.04-1.36) and by 12% (RR=1.14, 95% CI=1.00-1.31) after excluding the prospective cohort study in which PD cases were identified based on drug claims.[125] The difference between prospective studies and those based on healthcare databases became statistically significant (Figure 17; Panel A,  $P=0.029$ ; Panel B,  $P=0.008$ ).

**Table 23: Association between time-varying diabetes (prevalent and incident cases) and PD incidence over the follow-up.**

		Cases (n)	Model 1		Model 2		Model 3		Model 4	
			HR (95% CI)	P-value	HR (95% CI)	P-value	HR (95% CI)	P-value	HR (95% CI)	P-value
<b>Baseline (FU 1990-2018)</b>										
<b>DM</b>	No	1,137	1.00 (Ref.)	-	1.00 (Ref.)	-				
	Yes	63	0.96 (0.74-1.23)	0.726	1.04 (0.80-1.35)	0.764				
<b>DM duration</b>	Per 5y	1,200	1.00 (0.92-1.08)	0.978	1.02 (0.94-1.10)	0.678				
	No DM	1,137	1.00 (Ref.)	-	1.00 (Ref.)	-				
	DM [0-5[y	17	0.91 (0.56-1.47)	0.701	0.99 (0.61-1.61)	0.981				
	DM ≥5y	46	0.97 (0.72-1.31)	0.858	1.06 (0.78-1.43)	0.709				
	No DM	1,137	1.00 (Ref.)	-	1.00 (Ref.)	-				
<b>DM onset</b>	DM [0-10[y	31	0.87 (0.61-1.25)	0.461	0.96 (0.67-1.37)	0.810				
	DM ≥10y	32	1.05 (0.74-1.49)	0.785	1.14 (0.80-1.63)	0.478				
	No DM	1,137	1.00 (Ref.)	-	1.00 (Ref.)	-				
	DM <65y	34	1.01 (0.72-1.42)	0.948	1.10 (0.78-1.56)	0.575				
	DM ≥65y	29	0.90 (0.62-1.30)	0.564	0.97 (0.67-1.42)	0.892				
<b>Baseline (FU 2004-2018)</b>										
<b>DM</b>	No	901	1.00 (Ref.)	-	1.00 (Ref.)	-	1.00 (Ref.)	-	1.00 (Ref.)	-
	Yes	56	0.94 (0.72-1.23)	0.641	1.04 (0.79-1.37)	0.798	0.99 (0.75-1.31)	0.968	1.05 (0.80-1.39)	0.711
<b>DM duration</b>	Per 5y	957	0.97 (0.89-1.06)	0.503	0.99 (0.91-1.08)	0.837	0.98 (0.90-1.07)	0.693	0.99 (0.91-1.09)	0.889
	No diabetes	901	1.00 (Ref.)	-	1.00 (Ref.)	-	1.00 (Ref.)	-	1.00 (Ref.)	-
	DM [0-5[y	16	0.96 (0.59-1.58)	0.878	1.06 (0.65-1.75)	0.804	1.09 (0.66-1.79)	0.746	1.09 (0.66-1.79)	0.748
	DM ≥5y	40	0.93 (0.68-1.28)	0.646	1.03 (0.74-1.42)	0.876	1.04 (0.75-1.45)	0.793	1.04 (0.75-1.44)	0.803
	No diabetes	901	1.00 (Ref.)	-	1.00 (Ref.)	-	1.00 (Ref.)	-	1.00 (Ref.)	-
	DM [0-10[y	30	0.94 (0.65-1.35)	0.733	1.04 (0.72-1.51)	0.823	0.99 (0.69-1.44)	0.974	1.06 (0.73-1.54)	0.745
	DM ≥10y	26	0.94 (0.63-1.38)	0.742	1.03 (0.69-1.53)	0.883	0.99 (0.67-1.48)	0.980	1.04 (0.70-1.55)	0.832
<b>DM onset</b>	No diabetes	901	1.00 (Ref.)	-	1.00 (Ref.)	-	1.00 (Ref.)	-	1.00 (Ref.)	-
	DM <65y	27	0.94 (0.64-1.38)	0.756	1.05 (0.71-1.54)	0.824	1.00 (0.68-1.48)	0.991	1.06 (0.72-1.57)	0.770
	DM ≥65y	29	0.93 (0.64-1.36)	0.721	1.03 (0.71-1.50)	0.879	0.99 (0.68-1.44)	0.946	1.05 (0.72-1.53)	0.803

Y, years; FU, follow-up. HR and 95% CI calculated using Cox proportional hazards models with age as the time scale.

Model 1= adjusted for age.

Model 2= adjusted for baseline place of residence (rural/urban), age at menarche (≤11/12-13/≥14 years), parity (nulliparous/one child/two children/≥three children), and for time-varying smoking (never/ex/current), menopausal status (no/natural/artificial/unknown type), body mass index (<18.5/[18.5-25]/[25-30]/≥30 kg/m<sup>2</sup>), physical activity (quartiles), hypercholesterolemia (Yes/No) and hypertension (Yes/No).

Model 3= Model 2 + time-varying number of medical visits (0/1/2/≥3 visits per month).

Model 4= Model 2 + time-varying cardiovascular diseases (Yes/No) and statins use (Yes/No).

**Table 24: Lagged analyses of the association between time-varying diabetes and PD incidence over the follow-up.**

Diabetes	Cases (n)	Model 1		Model 2	
		HR (95% CI)	P-value	HR (95% CI)	P-value
<b>5-year lag (FU 1995-2018)</b>					
No	1,119	1.00 (Ref.)	-	1.00 (Ref.)	-
Yes	46	0.98 (0.73-1.32)	0.920	1.03 (0.76-1.39)	0.855
<b>10-year lag (FU 2000-2018)</b>					
No	1,053	1.00 (Ref.)	-	1.00 (Ref.)	-
Yes	28	0.95 (0.65-1.39)	0.804	1.01 (0.69-1.48)	0.958
<b>15-year lag (FU 2005-2018)</b>					
No	895	1.00 (Ref.)	-	1.00 (Ref.)	-
Yes	16	0.89 (0.54-1.45)	0.634	0.92 (0.56-1.53)	0.756

FU, follow-up. HR and 95% CI calculated using Cox proportional hazards models with age as the time scale.

Model 1= adjusted for age.

Model 2= adjusted for baseline place of residence (rural/urban), age at menarche ( $\leq 11/12-13/\geq 14$  years), parity (nulliparous/one child/two children/ $\geq$ three children), and for time-varying smoking (never/ex/current), menopausal status (no/natural/artificial/unknown type), body mass index ( $< 18.5/[18.5-25/[25-30]/\geq 30$  kg/m<sup>2</sup>), physical activity (quartiles), hypercholesterolemia (Yes/No) and hypertension (Yes/No).

**Table 25: Association between incident diabetes (1990-2016) and PD incidence over the follow-up.**

		Cases (n)	Model 1		Model 2		Model 3		Model 4	
			HR (95% CI)	P-value	HR (95% CI)	P-value	HR (95% CI)	P-value	HR (95% CI)	P-value
<b>Baseline (FU 1990-2018)</b>										
<b>DM</b>	No	1,137	1.00 (Ref.)	-	1.00 (Ref.)	-	-	-	-	-
	Yes	49	0.90 (0.68-1.20)	0.477	0.98 (0.73-1.32)	0.911	-	-	-	-
<b>DM duration</b>	Per 5y	1,186	0.97 (0.84-1.11)	0.647	1.00 (0.87-1.15)	0.965	-	-	-	-
	No DM	1,137	1.00 (Ref.)	-	1.00 (Ref.)	-	-	-	-	-
	DM [0-5[y	17	0.91 (0.57-1.48)	0.715	0.99 (0.61-1.61)	0.984	-	-	-	-
	DM ≥5y	32	0.89 (0.63-1.27)	0.534	0.98 (0.68-1.40)	0.900	-	-	-	-
	No DM	1,137	1.00 (Ref.)	-	1.00 (Ref.)	-	-	-	-	-
<b>DM onset</b>	DM [0-10[y	31	0.88 (0.62-1.26)	0.502	0.96 (0.67-1.39)	0.845	-	-	-	-
	DM ≥10y	18	0.93 (0.58-1.49)	0.764	1.02 (0.64-1.63)	0.941	-	-	-	-
	No DM	1,137	1.00 (Ref.)	-	1.00 (Ref.)	-	-	-	-	-
	DM <65y	20	0.92 (0.59-1.42)	0.695	1.01 (0.65-1.58)	0.963	-	-	-	-
	DM ≥65y	29	0.89 (0.61-1.29)	0.544	0.97 (0.66-1.41)	0.853	-	-	-	-
<b>Baseline (FU 2004-2018)</b>										
<b>DM</b>	No	901	1.00 (Ref.)	-	1.00 (Ref.)	-	1.00 (Ref.)	-	1.00 (Ref.)	-
	Yes	48	0.94 (0.70-1.25)	0.665	1.04 (0.77-1.41)	0.780	1.00 (0.74-1.35)	0.986	1.06 (0.78-1.43)	0.708
<b>DM duration</b>	Per 5y	949	0.98 (0.85-1.12)	0.733	1.02 (0.89-1.17)	0.804	1.01 (0.88-1.16)	0.910	1.02 (0.89-1.17)	0.745
	No diabetes	901	1.00 (Ref.)	-	1.00 (Ref.)	-	1.00 (Ref.)	-	1.00 (Ref.)	-
	DM [0-5[y	16	0.96 (0.59-1.58)	0.876	1.06 (0.65-1.75)	0.805	1.00 (0.61-1.65)	0.999	1.08 (0.66-1.78)	0.760
	DM ≥5y	32	0.93 (0.65-1.32)	0.671	1.03 (0.72-1.48)	0.859	1.00 (0.70-1.44)	0.984	1.05 (0.73-1.50)	0.799
	No diabetes	901	1.00 (Ref.)	-	1.00 (Ref.)	-	1.00 (Ref.)	-	1.00 (Ref.)	-
<b>DM onset</b>	DM [0-10[y	30	0.94 (0.65-1.35)	0.729	1.04 (0.72-1.51)	0.827	0.99 (0.69-1.44)	0.966	1.06 (0.73-1.53)	0.765
	DM ≥10y	18	0.94 (0.59-1.50)	0.787	1.05 (0.65-1.68)	0.853	1.02 (0.64-1.64)	0.932	1.06 (0.66-1.70)	0.809
	No diabetes	901	1.00 (Ref.)	-	1.00 (Ref.)	-	1.00 (Ref.)	-	1.00 (Ref.)	-
	DM <65y	19	0.95 (0.60-1.49)	0.818	1.07 (0.68-1.70)	0.769	1.03 (0.65-1.64)	0.885	1.09 (0.68-1.73)	0.722
	DM ≥65y	29	0.93 (0.64-1.35)	0.707	1.03 (0.70-1.50)	0.895	0.98 (0.67-1.43)	0.928	1.04 (0.71-1.52)	0.836

DM, diabetes mellitus; Y, years; FU, follow-up. HR and 95% CI calculated using Cox proportional hazards models with age as the time scale.

Model 1= adjusted for age.

Model 2= adjusted for baseline place of residence (rural/urban), age at menarche ( $\leq 11/12-13/\geq 14$  years), parity (nulliparous/one child/two children/ $\geq$ three children), and for time-varying smoking (never/ex/current), menopausal status (no/natural/artificial/unknown type), body mass index ( $< 18.5/[18.5-25/[25-30]/\geq 30$  kg/m<sup>2</sup>), physical activity (quartiles), hypercholesterolemia (Yes/No) and hypertension (Yes/No).

Model 3= Model 2 + time-varying number of medical visits (0/1/2/ $\geq 3$  medical visits per month).

Model 4= Model 2 + time-varying cardiovascular diseases (Yes/No) and statins use (Yes/No).

**Table 26: Lagged analyses of the association between time-varying incident diabetes (1990-2016) and PD incidence over the follow-up.**

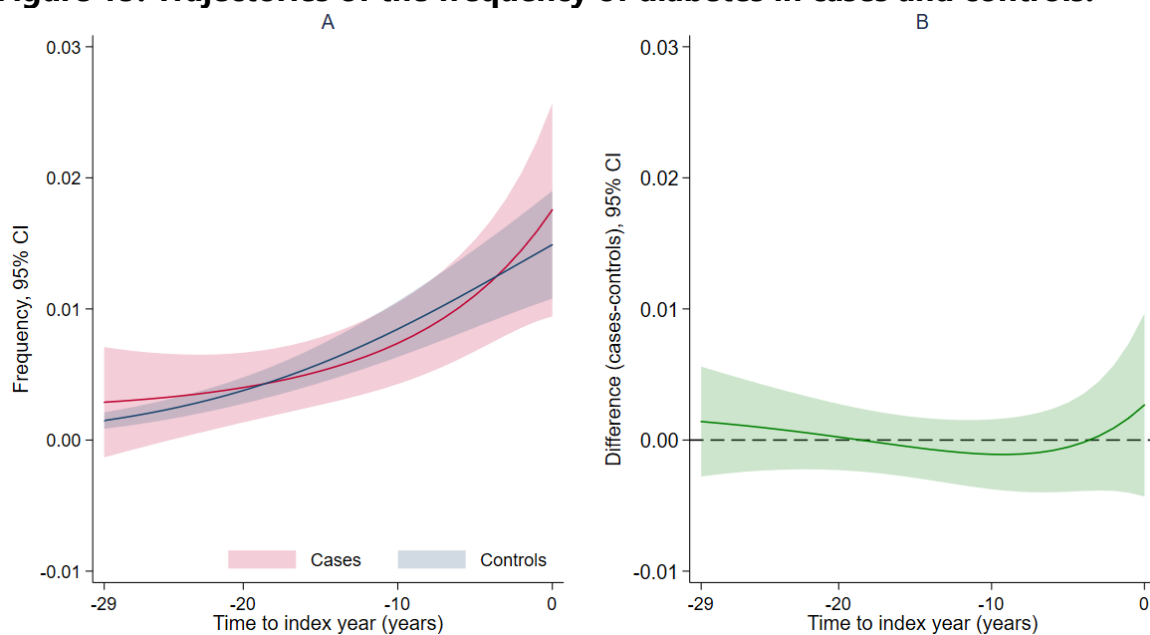
Diabetes	Cases (n)	HR (95% CI) <sup>a</sup>	P- value	HR (95% CI) <sup>b</sup>	P- value
<b>5-year lag (FU 1995-2018)</b>					
No	1,119	1.00 (Ref.)	-	1.00 (Ref.)	-
Yes	32	0.90 (0.63-1.27)	0.5418	0.94 (0.65-1.34)	0.7238
<b>10-year lag (FU 2000-2018)</b>					
No	1,053	1.00 (Ref.)	-	1.00 (Ref.)	-
Yes	18	0.93 (0.58-1.48)	0.7561	1.00 (0.62-1.60)	0.9906
<b>15-year lag (FU 2005-2018)</b>					
No	831	1.00 (Ref.)	-	1.00 (Ref.)	-
Yes	10	1.01 (0.54-1.89)	0.9677	1.07 (0.57-2.00)	0.8448

FU, follow-up. HR and 95% CI calculated using Cox proportional hazards models with age as the time scale.

<sup>a</sup> Model adjusted for age.

<sup>b</sup> Model adjusted for baseline place of residence (rural/urban), age at menarche ( $\leq 11/12-13/\geq 14$  years), parity (nulliparous/one child/two children/ $\geq$ three children), and for time-varying smoking (never/ex/current), menopausal status (no/natural/artificial/unknown type), body mass index ( $< 18.5/[18.5-25]/[25-30]/\geq 30$  kg/m<sup>2</sup>), physical activity (quartiles), hypercholesterolemia (Yes/No) and hypertension (Yes/No).

**Figure 15: Trajectories of the frequency of diabetes in cases and controls.**



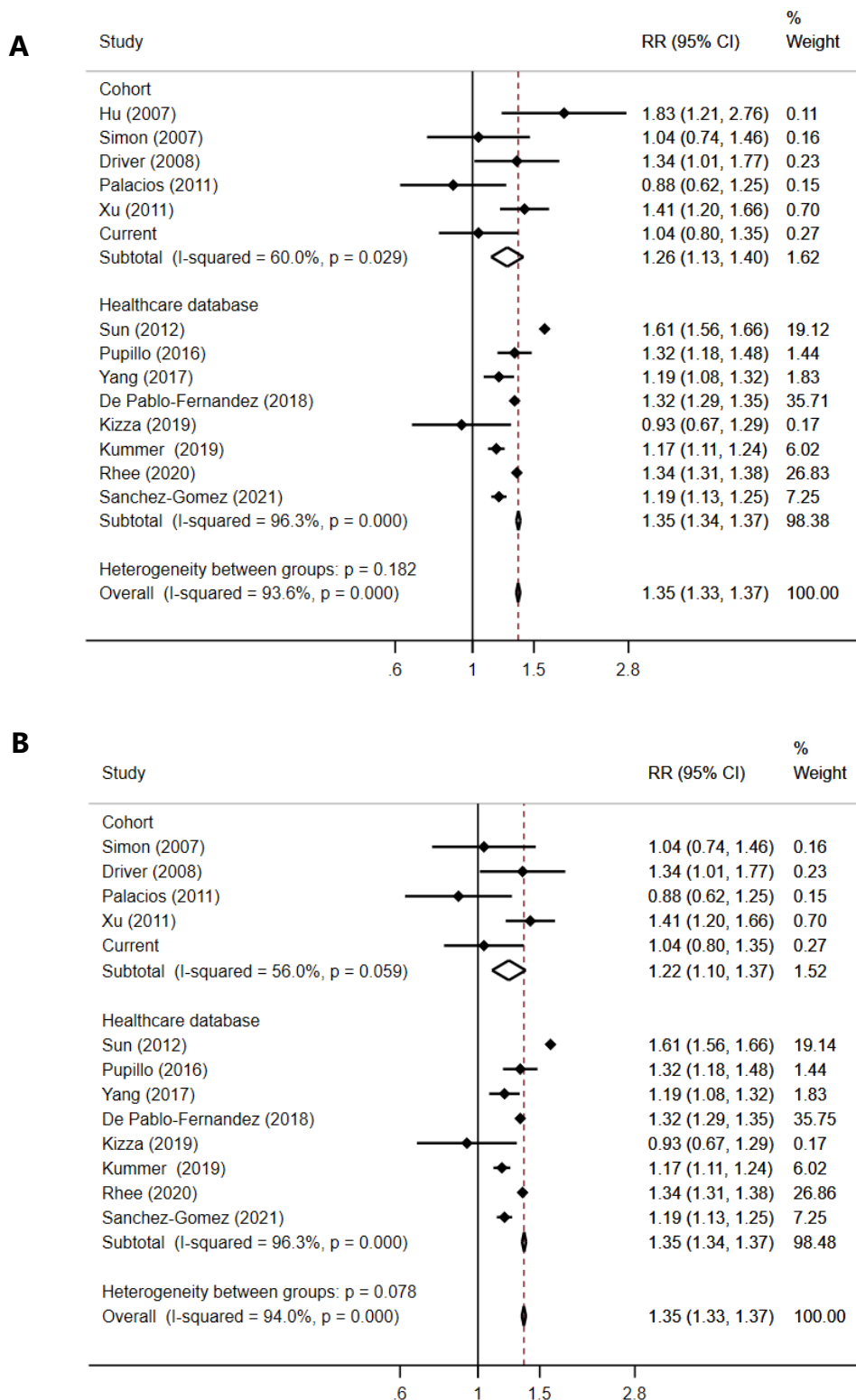
Panel A shows the frequencies (solid lines) and 95% CI (shaded areas) of diabetes (red) and controls (blue) according to a retrospective time before index date ( $T_0$ , time=0).

Panel B shows the corresponding case-control differences in frequencies of diabetes according to retrospective time, together with their 95% CI. Only those CI that do not include 0 are statistically significant.

Frequencies were predicted by a GEE logistic model using a retrospective time scale, with  $T_0$  representing the age at PD diagnosis in cases and age at index date in matched controls. The model includes a quadratic function of retrospective time, case-controls status, and the interactions of case-controls status with time. The model was adjusted for baseline age at menarche, parity, and place of residence, and time-varying smoking, menopausal status, and physical activity (Annex 5).

Trajectories were plotted for the most common profile of E3N participants: 73 years of age at index date ( $T_0$ ), never smokers, age at menarche at 12-13 years, natural menopause, 2 children, living in urban areas, third quartile (Q3) of physical activity, and normal BMI (BMI= $[18.5-25]$  kg/m<sup>2</sup>).

**Figure 16: Meta-analysis of the association between diabetes and Parkinson's disease.**

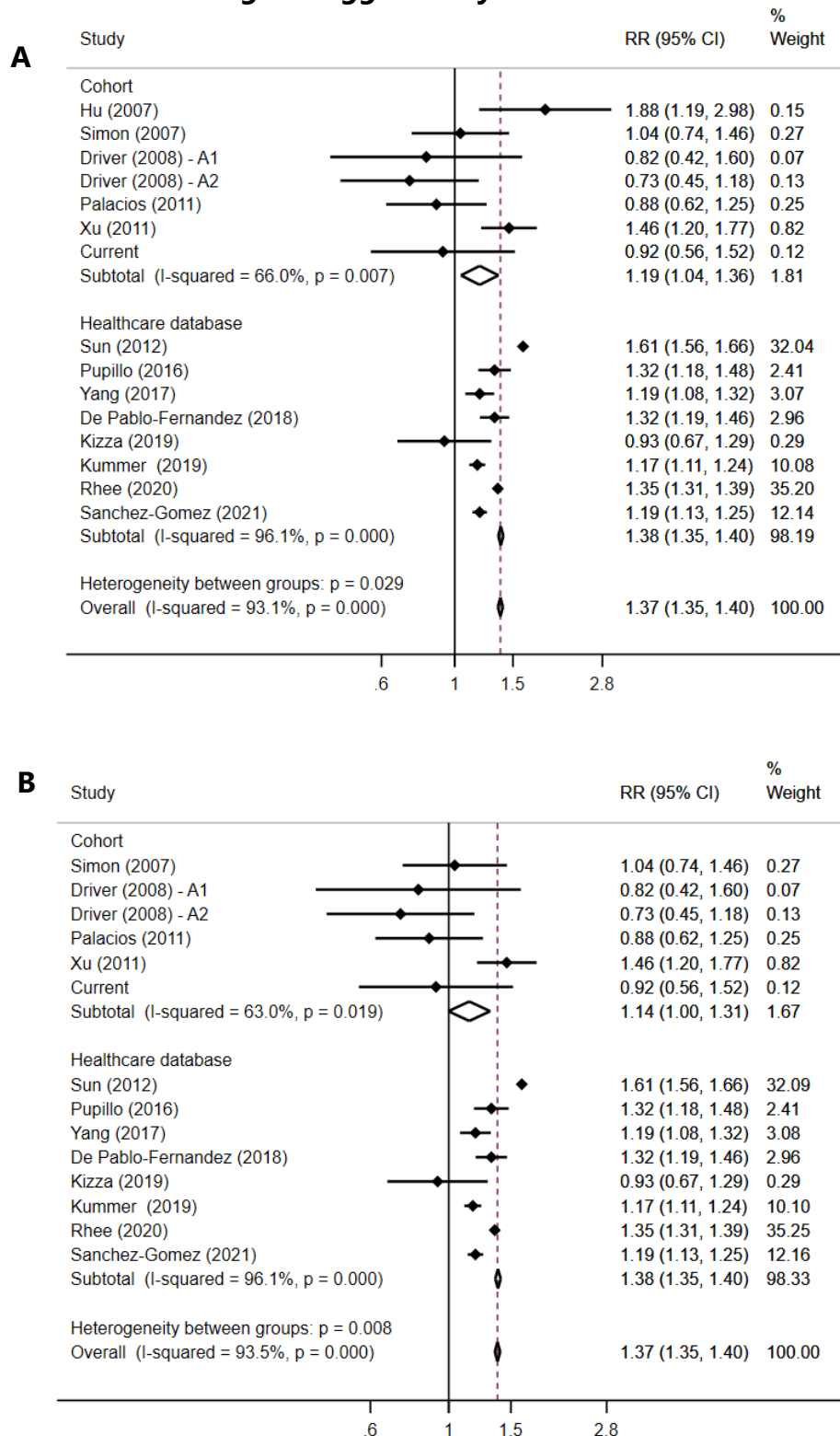


RR, relative risk; CI, Confidence interval.

Pooled RRs are indicated by the diamonds and 95% CI by the horizontal line.

Panel A – meta-analysis with 14 studies; Panel B – meta-analysis with 13 studies, after excluding one prospective cohort that identified PD cases based on drug claims.

**Figure 17: Meta-analysis of the association between diabetes and Parkinson's disease accounting for lagged analysis or duration of diabetes.**



RR, relative risk; CI, Confidence interval; A1= diabetes duration of 10-15 years; A2=diabetes duration of  $\geq 15$  years. Pooled RRs are indicated by the diamonds and 95% CI by the horizontal line.

Panel A – meta-analysis with 14 studies; Panel B – meta-analysis with 13 studies, after excluding one prospective cohort that identified PD cases based on drug claims.



## 6.5. Discussion

In the E3N cohort study of ~100,000 women, diabetes was not associated with PD incidence, even when assessed 5, 10, or 15 years before PD diagnosis. Analyses of diabetes trajectories showed that the frequency of diabetes was similar in PD patients and controls over the 29 years preceding the index date. Our meta-analysis suggests that prospective cohort studies and those based on healthcare databases yielded different results.

Previous cohorts examined the association between diabetes and PD, but results were controversial (Table 4). Our analyses based on time-varying diabetes (prevalent and incident) yielded an association estimate of the same size as in pooled analyses of the Nurses' Health Study (mean FU=22.9 years) and the Health Professionals Follow-up Study (mean FU=12.6 years) that reported a HR of 1.04 (95% CI=0.74-1.46; 530 patients).[126] Another analysis of data from the US Cancer Prevention study II Nutrition Cohort (564 PD patients, FU=13 years) found no association between diabetes and PD incidence in men (HR=0.87, 95% CI=0.57-1.33), women (HR=0.88, 95% CI=0.62-1.25), and men and women combined (HR=0.88, 95% CI=0.62-1.25).[97] A Finnish study (633 PD patients, mean FU=18 years) reported that diabetes increased PD incidence in men and women combined (HR=1.83, 95% CI=1.21-2.76) and in men (HR=1.78, 95% CI=1.01-3.12) and women alone (HR=1.91, 95% CI=1.04-3.52); however, PD diagnoses were not validated and relied on drug claims only in this study.[125] Another US cohort study examined this association and reported that diabetes at baseline was associated with higher PD incidence in men and women combined (1,565 PD patients, FU=11 years; HR=1.41, 95% CI=1.20-1.66); the association was only present among those who had  $\geq 10$  years diabetes duration before the baseline.[120] One study in men used time-varying diabetes in the analyses, and therefore took into account incident cases of diabetes during the cohort.[70] That study showed an association between diabetes and PD only among those with less than 10 years of diabetes duration, but not among those with longer diabetes duration; in addition,

diagnoses of diabetes tended to cluster around the time of PD diagnosis. The authors of this study hypothesized that their findings could be explained by detection bias from increased medical surveillance during the prodromal phase of the disease. We did not observe a similar pattern in E3N and analyses adjusted for the number of medical contacts did not change our conclusions.

Eight other studies based on healthcare databases examined the association between diabetes and PD, and seven reported increased PD incidence among diabetic participants.[118, 119, 121, 127-131] However, no validation of PD diagnoses was performed in any of these studies; this may lead to inadequately include patients with vascular parkinsonism in the analyses which could lead to bias analyses on the role of diabetes.[130, 206, 207] In addition, due to their short FU (all studies with a  $FU \leq 12$  years), reverse causation cannot be excluded. Two studies used a lag of  $\leq 5$  years that may be too short to fully eliminate the risk of reverse causation.[118, 129] The role of duration of diabetes was also examined in some studies. One study reported similar associations for shorter (1-4 years) and longer ( $\geq 5$  years) duration of diabetes,[121] while another reported a stronger association in those with longer ( $\geq 5$  years) duration of diabetes compared to those with shorter duration.[118] While some of these studies were able to adjust for confounders such as smoking, physical activity, or BMI, several did not, which may lead to residual confounding. Some of these studies adjusted for markers of socio-economic status, and only one[121] adjusted for markers of healthcare use that can have a strong impact on associations.[123] Finally, age at PD onset or diagnosis may be captured more imprecisely and overestimated in healthcare databases, and this may lead to some degree of misclassification in the delay between diabetes onset and PD diagnosis, and even in exposure status before PD diagnosis.

Our meta-analysis of cohort studies provides evidence in favour of prospective cohort studies and those based on healthcare databases showing different findings, with weaker or no association in prospective studies and stronger associations in studies based on healthcare databases. A combination of the methodological issues

described above may explain inconsistencies.

In addition, a meta-analysis of case-control studies reported an inverse association between diabetes and PD.[114] The reasons for inconsistent findings between case-control and cohort studies is unclear, but it has been argued that case-control studies may be more prone to recall and selection biases than prospective cohort studies.[114, 116]

Over the prodromal phase of PD, participants tend to experience changes in eating behavior and physical activity that can alter the risk of diabetes.[102] In addition, it has been suggested that PD neuronal loss alters glycaemic control,[70] and mitochondrial function,[124] both of which increase the risk of insulin resistance. Some previous studies addressed the risk of reverse causation,[97, 118, 120, 125, 129] but only lags of  $\leq 5$  years were used. Other studies addressed this issue by performing analyses of PD risk according to the duration of diabetes as discussed above.[70, 118, 120, 121] Our analyses showed no effect of including a lag or no role of disease duration.

Survival bias may be another potential issue for analyses on the relation between diabetes and PD. Diabetic patients have higher mortality than non-diabetic persons, and those who survived until the beginning of the study may be different from those who did not.[116, 117] One of the main differences between previous studies and ours is that we took into account incident diabetes over the long follow-up. In addition, we performed analyses that excluded prevalent diabetic patients showing similar results compared to our main analyses.

One MR study reported that genetically-predicted T2DM was associated with PD risk (OR=1.08, 95% CI= 1.02-1.14).[114] Although MR studies are considered to be less likely to be biased by residual confounding and reverse causation, they rely on several assumptions whose violations may yield spurious associations.[208] In particular, survival bias is an issue for MR studies of age-related diseases;[209] the available MR

study was based on PD cases who were significantly older than the controls, and the authors did not examine whether survival bias may have been an issue.

There is some evidence in the literature in favour of biological relationships between T2DM and PD. Both T2DM and PD are age-associated diseases, and insulin resistance, a hallmark of T2DM, may contribute to PD.[109, 117] Changes in insulin signaling pathway increase  $\alpha$ -Syn aggregation, mitochondrial dysfunction, and neuroinflammation.[117] A major reason for the interest in the relation between diabetes and PD has been the potential of drugs used to treat T2DM to be repurposed for PD. A double-blind, placebo-controlled trial showed that the glucagon-like peptide-1 (GLP-1) receptor agonist Exenatide may have an effect on reducing PD severity, in favour of a possible disease-modifying effect.[210] A recent observational cohort study of 100,288 T2DM patients of whom 329 developed PD over a median follow-up of 3.33 years showed that patients taking GLP-1 receptor agonists and dipeptidyl peptidase-4 (DPP4) inhibitors were at lower PD risk than patients taking other oral antidiabetic drugs.[211] This finding was not supported by a larger case-control study in diabetic patients (2,017 PD cases, 7,934 controls) showing no robust evidence on the association between specific diabetes medication classes at least three years before the index date and risk of PD.[212] Additional studies are underway to examine the benefits and safety of these drugs in PD.

Diabetes was ascertained through use of antidiabetic drugs. Although most diabetic patients are not perfectly stabilized over time, even when treated, it is possible that a good control of diabetes may have contributed to null findings. In addition, given that some antidiabetic drugs may have beneficial effects in PD as discussed above, it is also possible that a deleterious role of diabetes may have been mitigated by use of these drugs.[211] In E3N, the drug that was most frequently used by PD patients with diabetes was metformin; none of the PD patients used either GLP-1 receptor agonists or dipeptidyl peptidase-4 inhibitors.

The main strengths of this study include the long follow-up and its large size that allowed to address the potential for reverse causation. Our study is the third largest study in women, after two other studies based on healthcare databases. Diabetes status was updated over the follow-up through questionnaires and drug claim databases that allowed us to take into account changes in diabetic status over the follow-up, and to compare diabetes trajectories in PD patients and controls.

Our study has limitations. In addition to those already discussed in previous chapters regarding the generalizability of findings in E3N, the main limitation is that the number of patients with diabetes who developed PD was small, and our analyses may have been underpowered to detect a weak association ( $HR < 1.36$ ). In addition, the frequency of diabetes in E3N is lower than expected. The selected nature of the cohort, that included women who are health conscious with a lower prevalence of overweight and obesity, may be an explanation.[192] It also possible that some diabetic women may not have been detected, especially at early stages of diabetes when they were not treated. We were not able to distinguish T2DM from T1DM. However, given that the majority of cases of diabetes in adults is T2DM, most patients in our cohort are likely to have developed T2DM. Finally, due to the lack of information on diabetes complications, analyses according to disease severity were not possible, and we addressed this issue by using duration of diabetes as a proxy.

In conclusion, diabetes was not associated with PD incidence in women from the French E3N cohort study. Differences between prospective cohort studies and those based on healthcare databases may explain inconsistent findings and large between-study heterogeneity, and additional large and well-designed studies are needed to examine whether diabetes is associated with PD.

## **7. Main findings, Perspectives, and Conclusion**



## 7.1. Main findings

In this thesis, we showed that higher levels of physical activity and obesity were associated with reduced incidence of PD in women, while we found no association between diabetes and PD incidence.

Our analyses are based on ~100,000 women from the E3N study followed over 29 years. The large sample size and long follow-up allowed us to address the potential for reverse causation, while retaining a sufficient number of PD patients in the analysis. Reverse causation represents a major challenge for epidemiologic studies of neurodegenerative disorders, such as PD, given their long prodromal phase. In addition, our study helps to fill a gap in PD research, as women tend to be understudied due to the higher frequency of PD in men.

Our first study reinforces the evidence in favour of the health benefits of physical activity as a promising lifestyle disease-modifying intervention. We provide further evidence that time-varying physical activity is associated with a lower incidence of PD in women, and that this association is unlikely to be explained by reverse causation. Analyses of physical activity trajectories showed that PD patients had lower physical activity levels than controls over the whole follow-up. In addition, physical activity level decreased in cases approximately 10 years before diagnosis, in agreement with the hypothesis that prodromal PD may lead to a reduction in physical activity in the years preceding PD diagnosis. Our findings support the importance of addressing reverse causation in analyses on the relation between physical activity and PD, and suggest that analyses without an exposure lag or with short lags are likely to be biased by reverse causation.

Our second study provides evidence that higher BMI, in particular obesity ( $\text{BMI} \geq 30 \text{ kg/m}^2$ ), is associated with lower PD incidence in agreement with findings from the latest MR studies.[93, 94, 101] A similar pattern was found for higher WC and WHtR, with significant associations in analyses lagged by 10 years. Analyses of BMI trajectories



showed that obesity was less frequent in PD patients than controls even 29 years before index date, and that BMI started to decrease in PD patients in the 5-10 years before PD diagnosis, in agreement with the hypothesis that weight loss precedes PD diagnosis. Again, our analyses support the importance of appropriate analyses to address the potential for reverse causation, and may contribute to explain inconsistent findings from previous studies.

Our third study provides no evidence in favour of an association between diabetes and PD, and is not in agreement with results from recent meta-analyses of cohort studies[114, 115, 202] or one MR study.[114] However, the positive association between diabetes and PD in the literature appears to be mainly driven by studies based on healthcare databases, while prospective studies provide weaker or no evidence. A number of issues may explain inconsistent findings between the two types of studies, including surveillance bias, reverse causation, residual confounding, diagnostic misclassification, survival bias, or inadequate definition of the date of PD onset in healthcare databases.

## **7.2. Perspectives and conclusions**

Due to the lack of curative PD treatment and the increasing number of PD patients worldwide, there is an urgent need for potential lifestyle PD-modifying factors that may prevent or delay the onset of the disease. Reverse causation represents a major threat for epidemiologic studies of PD aiming at identifying potential preventative measures due to the long prodromal phase of the disease, and needs to be adequately addressed to provide valid conclusions.

At the present time, physical activity/exercise appears to be the therapeutic option with the best combination of feasibility and safety to be tested as a disease modifying intervention in patients with prodromal symptoms or at genetic risk of PD.[42] Exercise provides overall health benefits and is easily accessible at low cost, but

compliance represents a challenge. Although assessment of physical activity through questionnaires has extensively been used in epidemiologic studies for a wide range of outcomes, recent advances in technology provide more objective ways to measure physical activity, and additional studies using accelerometers would be useful. Further studies are also needed to better understand which types of activities may be more beneficial, and the biologic mechanisms underlying a beneficial role of physical activity/exercise in PD.

Our findings on the association between adiposity and PD do not have an immediate translation in terms of prevention, given the many other detrimental effects of obesity, but may provide clues about the mechanisms involved in PD, and additional studies are needed to examine this issue. Finally, additional studies that adequately address the issues described above are needed to examine the relation between diabetes and PD.

Over the past years, our knowledge of PD has significantly improved, in particular with the recognition of a long prodromal phase and the identification of persons at higher PD risk (prodromal symptoms, carriers of genetic mutations). Although prevention trials are starting to be considered in PD, many questions remain, and additional epidemiologic studies are needed to help identify interventions (e.g., lifestyle behaviours, drugs) that could be tested, alone or in combination, in such trials, while adequately addressing the risk of reverse causation.



# REFERENCES



1. Parkinson J. *An essay on the shaking palsy. 1817*. The Journal of neuropsychiatry and clinical neurosciences, 2002. 14(2): p. 223-36.
2. Obeso JA, Stamelou M, Goetz CG, et al. *Past, present, and future of Parkinson's disease: A special essay on the 200th Anniversary of the Shaking Palsy*. Mov Disord, 2017. 32(9): p. 1264-1310.
3. Kalia LV, Lang AE. *Parkinson's disease*. The Lancet, 2015. 386(9996): p. 896-912.
4. HAS. *Guide du parcours de soins: la maladie de Parkinson*, Haute Autorité de Santé. 2016.
5. Savica R, Boeve BF, Mielke MM. *When Do alpha-Synucleinopathies Start? An Epidemiological Timeline: A Review*. JAMA Neurol, 2018. 75(4): p. 503-509.
6. Schrag A, Horsfall L, Walters K, et al. *Prediagnostic presentations of Parkinson's disease in primary care: a case-control study*. The Lancet Neurology, 2015. 14(1): p. 57-64.
7. Postuma RB, Berg D. *Advances in markers of prodromal Parkinson disease*. Nat Rev Neurol, 2016. 12(11): p. 622-634.
8. Elbaz A. *Prodromal symptoms of Parkinson's disease: Implications for epidemiological studies of disease etiology*. Rev Neurol, 2016. 172(8-9): p. 503-511.
9. Tolosa E, Garrido A, Scholz SW, et al. *Challenges in the diagnosis of Parkinson's disease*. The Lancet Neurology, 2021. 20(5): p. 385-397.
10. Berg D, Postuma RB, Adler CH, et al. *MDS research criteria for prodromal Parkinson's disease*. Mov Disord, 2015. 30(12): p. 1600-11.
11. Poewe W, Seppi K, Tanner CM, et al. *Parkinson disease*. Nat Rev Dis Primers, 2017. 3: p. 17013.
12. Stoker TB, Greenland JC. *Parkinson's Disease: Pathogenesis and Clinical Aspects*. Brisbane (AU): Codon Publications. 2018.
13. Borghammer P. *The alpha-Synuclein Origin and Connectome Model (SOC Model) of Parkinson's Disease: Explaining Motor Asymmetry, Non-Motor Phenotypes, and Cognitive Decline*. J Parkinsons Dis, 2021. 11(2): p. 455-474.
14. Borghammer P, Van Den Berge N. *Brain-First versus Gut-First Parkinson's*

*Disease: A Hypothesis*. J Parkinsons Dis, 2019. 9(s2): p. S281-S295.

15. Berg D, Borghammer P, Fereshtehnejad SM, et al. *Prodromal Parkinson disease subtypes - key to understanding heterogeneity*. Nat Rev Neurol, 2021. 17(6): p. 349-361.
16. Carvey PM, Punati A, Newman MB. *Progressive dopamine neuron loss in Parkinson's disease: the multiple hit hypothesis*. Cell Transplant, 2006. 15(3): p. 239-50.
17. Gaig C, Tolosa E. *When does Parkinson's disease begin?* Mov Disord, 2009. 24 Suppl 2: p. S656-64.
18. Fereshtehnejad SM, Yao C, Pelletier A, et al. *Evolution of prodromal Parkinson's disease and dementia with Lewy bodies: a prospective study*. Brain, 2019. 142(7): p. 2051-2067.
19. Camacho-Soto A, Gross A, Searles Nielsen S, et al. *Fractures in the prodromal period of Parkinson disease*. Neurology, 2020. 94(23): p. e2448-e2456.
20. Hustad E, Aasly JO. *Clinical and Imaging Markers of Prodromal Parkinson's Disease*. Front Neurol, 2020. 11: p. 395.
21. Armstrong MJ, Okun MS. *Diagnosis and Treatment of Parkinson Disease: A Review*. JAMA, 2020. 323(6): p. 548-560.
22. Munhoz RP, Picillo M, Fox SH, et al. *Eligibility Criteria for Deep Brain Stimulation in Parkinson's Disease, Tremor, and Dystonia*. Can J Neurol Sci, 2016. 43(4): p. 462-71.
23. Bloem BR, de Vries NM, Ebersbach G. *Nonpharmacological treatments for patients with Parkinson's disease*. Mov Disord, 2015. 30(11): p. 1504-20.
24. Pupikova M, Rektorova I. *Non-pharmacological management of cognitive impairment in Parkinson's disease*. J Neural Transm (Vienna), 2019.
25. Mak MK, Wong-Yu IS, Shen X, et al. *Long-term effects of exercise and physical therapy in people with Parkinson disease*. Nat Rev Neurol, 2017. 13(11): p. 689-703.
26. Janssen Daalen JM, Schootemeijer S, Richard E, et al. *Lifestyle Interventions for the Prevention of Parkinson Disease: A Recipe for Action*. Neurology, 2022. 99(7 Supplement 1): p. 42-51.
27. Zhen K, Zhang S, Tao X, et al. *A systematic review and meta-analysis on effects of aerobic exercise in people with Parkinson's disease*. NPJ Parkinsons Dis, 2022. 8(1): p.

146.

28. Domingos J, Keus SHJ, Dean J, et al. *The European Physiotherapy Guideline for Parkinson's Disease: Implications for Neurologists*. J Parkinsons Dis, 2018. 8(4): p. 499-502.

29. Radder DLM, Lígia Silva de Lima A, Domingos J, et al. *Physiotherapy in Parkinson's Disease: A Meta-Analysis of Present Treatment Modalities*. Neurorehabilitation and Neural Repair, 2020. 34(10): p. 871-880.

30. Schootemeijer S, van der Kolk NM, Bloem BR, et al. *Current Perspectives on Aerobic Exercise in People with Parkinson's Disease*. Neurotherapeutics, 2020. 17(4): p. 1418-1433.

31. Cristini J, Weiss M, De Las Heras B, et al. *The effects of exercise on sleep quality in persons with Parkinson's disease: A systematic review with meta-analysis*. Sleep Med Rev, 2020. 55: p. 101384.

32. Johansson ME, Cameron IGM, Van der Kolk NM, et al. *Aerobic Exercise Alters Brain Function and Structure in Parkinson's Disease: A Randomized Controlled Trial*. Ann Neurol, 2022. 91(2): p. 203-216.

33. Kim R, Lee TL, Lee H, et al. *Effects of Exercise on Depressive Symptoms in Patients With Parkinson Disease: A Meta-analysis*. Neurology, 2022: p. 10.1212/WNL.0000000000201453.

34. Tsukita K, Sakamaki-Tsukita H, Takahashi R. *Long-term Effect of Regular Physical Activity and Exercise Habits in Patients With Early Parkinson Disease*. Neurology, 2022. 98(8): p. e859-e871.

35. Paul KC, Chuang YH, Shih IF, et al. *The association between lifestyle factors and Parkinson's disease progression and mortality*. Mov Disord, 2019. 34(1): p. 58-66.

36. Sunwoo MK, Lee JE, Hong JY, et al. *Premorbid exercise engagement and motor reserve in Parkinson's disease*. Parkinsonism Relat Disord, 2017. 34: p. 49-53.

37. Zhang X, Molsberry SA, Schwarzschild MA, et al. *Association of Diet and Physical Activity With All-Cause Mortality Among Adults With Parkinson Disease*. JAMA Netw Open, 2022. 5(8): p. e2227738.



38. Camicioli R, Majumdar SR. *Relationship between mild cognitive impairment and falls in older people with and without Parkinson's disease: 1-Year Prospective Cohort Study*. *Gait Posture*, 2010. 32(1): p. 87-91.
39. Macleod AD, Taylor KS, Counsell CE. *Mortality in Parkinson's disease: a systematic review and meta-analysis*. *Mov Disord*, 2014. 29(13): p. 1615-22.
40. Zhao N, Yang Y, Zhang L, et al. *Quality of life in Parkinson's disease: A systematic review and meta-analysis of comparative studies*. *CNS Neurosci Ther*, 2021. 27(3): p. 270-279.
41. Crotty GF, Keavney JL, Alcalay RN, et al. *Planning for Prevention of Parkinson Disease*. *Neurology*, 2022. 99(7 Supplement 1): p. 1-9.
42. Berg D, Crotty GF, Keavney JL, et al. *Path to Parkinson Disease Prevention: Conclusion and Outlook*. *Neurology*, 2022. 99(7 Suppl 1): p. 76-83.
43. Elbaz A, Artaud F, Canonico M, et al. *Épidémiologie descriptive et étiologique de la maladie de Parkinson*, in *Pathologies du mouvement*, C Tranchant, J P Azulay, Editors: Doin. 2020: 37-48.
44. Dorsey ER, Elbaz A, Nichols E, et al. *Global, regional, and national burden of Parkinson's disease, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016*. *The Lancet Neurology*, 2018. 17(11): p. 939-953.
45. Collaborators GD. *Global, regional, and national burden of Alzheimer's disease and other dementias, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016*. *The Lancet Neurology*, 2019. 18(1): p. 88-106.
46. Ou Z, Pan J, Tang S, et al. *Global Trends in the Incidence, Prevalence, and Years Lived With Disability of Parkinson's Disease in 204 Countries/Territories From 1990 to 2019*. *Front Public Health*, 2021. 9: p. 776847.
47. Wanneveich M, Moisan F, Jacqmin-Gadda H, et al. *Projections of prevalence, lifetime risk, and life expectancy of Parkinson's disease (2010-2030) in France*. *Mov Disord*, 2018.
48. Santé publique France. *Communication personnelle*. 2022.
49. Wooten GF. *Are men at greater risk for Parkinson's disease than women?* *Journal*

of Neurology, Neurosurgery & Psychiatry, 2004. 75(4): p. 637-639.

50. Moisan F, Kab S, Mohamed F, et al. *Parkinson disease male-to-female ratios increase with age: French nationwide study and meta-analysis*. J Neurol Neurosurg Psychiatry, 2016. 87(9): p. 952-7.

51. Darweesh SKL, Ikram MK, Faber MJ, et al. *Professional occupation and the risk of Parkinson's disease*. Eur J Neurol, 2018. 25(12): p. 1470-1476.

52. Gunnarsson LG, Bodin L. *Occupational Exposures and Neurodegenerative Diseases-A Systematic Literature Review and Meta-Analyses*. Int J Environ Res Public Health, 2019. 16(3).

53. Vlaar T, Kab S, Schwaab Y, et al. *Association of Parkinson's disease with industry sectors: a French nationwide incidence study*. Eur J Epidemiol, 2018. 33(11): p. 1101-1111.

54. Elbaz A, Carcaillon L, Kab S, et al. *Epidemiology of Parkinson's disease*. Rev Neurol (Paris), 2016. 172(1): p. 14-26.

55. Cerri S, Mus L, Blandini F. *Parkinson's Disease in Women and Men: What's the Difference?* J Parkinsons Dis, 2019. 9(3): p. 501-515.

56. Bourque M, Di Paolo T. *Neuroactive steroids and Parkinson's disease*. Current Opinion in Endocrine and Metabolic Research, 2022. 22: p. 100312.

57. Del Rey NL, Quiroga-Varela A, Garbayo E, et al. *Advances in Parkinson's Disease: 200 Years Later*. Front Neuroanat, 2018. 12: p. 113.

58. Pang SY, Ho PW, Liu HF, et al. *The interplay of aging, genetics and environmental factors in the pathogenesis of Parkinson's disease*. Transl Neurodegener, 2019. 8: p. 23.

59. Chen Y, Sun X, Lin Y, et al. *Non-Genetic Risk Factors for Parkinson's Disease: An Overview of 46 Systematic Reviews*. J Parkinsons Dis, 2021. 11(3): p. 919-935.

60. Flather MD, Farkouh ME, Pogue JM, et al. *Strengths and limitations of meta-analysis: larger studies may be more reliable*. Controlled clinical trials, 1997. 18(6): p. 568-79; discussion 661-6.

61. Deeks JJ, Higgins JPT, Altman DG. *Analysing data and undertaking meta-analyses.*, in *Cochrane Handbook for Systematic Reviews of Interventions version 6.3*, J P

T Higgins, J Thomas, J Chandler, et al, Editors: Cochrane. 2022.

62. Westreich D. *Epidemiology by design: a causal approach to the health sciences*: Oxford University Press. 2020.
63. Liestøl K, Andersen PK. *Updating of covariates and choice of time origin in survival analysis: problems with vaguely defined disease states*. Stat Med, 2002. 21(23): p. 3701-14.
64. Lee DH, Rezende LFM, Ferrari G, et al. *Physical activity and all-cause and cause-specific mortality: assessing the impact of reverse causation and measurement error in two large prospective cohorts*. Eur J Epidemiol, 2021. 36(3): p. 275-285.
65. Tamim H, Monfared AA, LeLorier J. *Application of lag-time into exposure definitions to control for protopathic bias*. Pharmacoepidemiol Drug Saf, 2007. 16(3): p. 250-8.
66. Schulz KF, Grimes DA. *Case-control studies: research in reverse*. The Lancet, 2002. 359(9304): p. 431-434.
67. Rothman K, Greenland S, Lash T. *Modern Epidemiology*. 3rd ed, ed. L W Wilkins. 2008.
68. Chen H, Zhang SM, Hernán MA, et al. *Weight loss in Parkinson's disease*. Ann Neurol, 2003. 53(5): p. 676-9.
69. Chen H, Zhang SM, Schwarzschild MA, et al. *Physical activity and the risk of Parkinson disease*. Neurology, 2005. 64(4): p. 664-669.
70. Driver JA, Smith A, Buring JE, et al. *Prospective cohort study of type 2 diabetes and the risk of Parkinson's disease*. Diabetes Care, 2008. 31(10): p. 2003-5.
71. van der Kolk NM, de Vries NM, Kessels RPC, et al. *Effectiveness of home-based and remotely supervised aerobic exercise in Parkinson's disease: a double-blind, randomised controlled trial*. The Lancet Neurology, 2019. 18(11): p. 998-1008.
72. Schenkman M, Moore CG, Kohrt WM, et al. *Effect of High-Intensity Treadmill Exercise on Motor Symptoms in Patients With De Novo Parkinson Disease: A Phase 2 Randomized Clinical Trial*. JAMA Neurol, 2018. 75(2): p. 219-226.
73. Reynolds GO, Otto MW, Ellis TD, et al. *The Therapeutic Potential of Exercise to*

*Improve Mood, Cognition, and Sleep in Parkinson's Disease.* *Mov Disord*, 2016. 31(1): p. 23-38.

74. Galetta S, Ganesh A, Lewis A, et al. *Editors' Note: Long-term Effect of Regular Physical Activity and Exercise Habits in Patients With Early Parkinson Disease.* *Neurology (Ovid)*, 2022. 99(9): p. 401-401.

75. Fang X, Han D, Cheng Q, et al. *Association of Levels of Physical Activity With Risk of Parkinson Disease: A Systematic Review and Meta-analysis.* *JAMA Netw Open*, 2018. 1(5): p. e182421.

76. Henderson GC. *Sexual dimorphism in the effects of exercise on metabolism of lipids to support resting metabolism.* *Front Endocrinol (Lausanne)*, 2014. 5: p. 162.

77. Palasz E, Niewiadomski W, Gasiorowska A, et al. *Exercise-Induced Neuroprotection and Recovery of Motor Function in Animal Models of Parkinson's Disease.* *Front Neurol*, 2019. 10: p. 1143.

78. Xu X, Fu Z, Le W. *Exercise and Parkinson's disease.* *Int Rev Neurobiol*, 2019. 147: p. 45-74.

79. Gleeson M, Bishop NC, Stensel DJ, et al. *The anti-inflammatory effects of exercise: mechanisms and implications for the prevention and treatment of disease.* *Nat Rev Immunol*, 2011. 11(9): p. 607-15.

80. Hou L, Chen W, Liu X, et al. *Exercise-Induced Neuroprotection of the Nigrostriatal Dopamine System in Parkinson's Disease.* *Front Aging Neurosci*, 2017. 9: p. 358.

81. Erickson HI, Gildengers AG, Butters MA. *Physical activity and brain plasticity in late adulthood.* *Dialogues in clinical neuroscience*, 2013. 15(1): p. 99-108.

82. Pialoux V, Brown AD, Leigh R, et al. *Effect of cardiorespiratory fitness on vascular regulation and oxidative stress in postmenopausal women.* *Hypertension*, 2009. 54(5): p. 1014-20.

83. Thacker EL, Chen H, Patel AV, et al. *Recreational physical activity and risk of Parkinson's disease.* *Mov. Disord.*, 2008. 23(1): p. 69-74.

84. Yang F, Trolle Lagerros Y, Belloc R, et al. *Physical activity and risk of Parkinson's disease in the Swedish National March Cohort.* *Brain*, 2015. 138(2): p. 269-75.

85. Saaksjarvi K, Knekt P, Mannisto S, et al. *Reduced risk of Parkinson's disease associated with lower body mass index and heavy leisure-time physical activity*. Eur J Epidemiol, 2014. 29(4): p. 285-92.
86. Xu Q, Park Y, Huang X, et al. *Physical activities and future risk of Parkinson disease*. Neurology, 2010. 75(4): p. 341–348.
87. Logroscino G, Sesso HD, Paffenbarger RS, Jr., et al. *Physical activity and risk of Parkinson's disease: a prospective cohort study*. J Neurol Neurosurg Psychiatry, 2006. 77(12): p. 1318-22.
88. Llamas-Velasco S, Contador I, Mendez-Guerrero A, et al. *Physical activity and risk of Parkinson's disease and parkinsonism in a prospective population-based study (NEDICES)*. Prev Med Rep, 2021. 23: p. 101485.
89. van der Marck MA, Dicke HC, Uc EY, et al. *Body mass index in Parkinson's disease: a meta-analysis*. Parkinsonism Relat Disord, 2012. 18(3): p. 263-7.
90. Fasano A, Visanji NP, Liu LWC, et al. *Gastrointestinal dysfunction in Parkinson's disease*. The Lancet Neurology, 2015. 14(6): p. 625-639.
91. Wang YL, Wang YT, Li JF, et al. *Body Mass Index and Risk of Parkinson's Disease: A Dose-Response Meta-Analysis of Prospective Studies*. PLoS One, 2015. 10(6): p. e0131778.
92. Rahmani J, Roudsari AH, Bawadi H, et al. *Body mass index and risk of Parkinson, Alzheimer, Dementia, and Dementia mortality: a systematic review and dose-response meta-analysis of cohort studies among 5 million participants*. Nutr Neurosci, 2020. 25(3): p. 423-431.
93. Noyce AJ, Kia DA, Hemani G, et al. *Estimating the causal influence of body mass index on risk of Parkinson disease: A Mendelian randomisation study*. PLoS Med, 2017. 14(6): p. e1002314.
94. Heilbron K, Jensen MP, Bandres-Ciga S, et al. *Unhealthy Behaviours and Risk of Parkinson's Disease: A Mendelian Randomisation Study*. J Parkinsons Dis, 2021. 11(4): p. 1981-1993.
95. Riso L, Kaaks R, Kuhn T, et al. *General and abdominal adiposity and the risk of*

- Parkinson's disease: A prospective cohort study.* Parkinsonism Relat Disord, 2019. 62: p. 98-104.
96. Chen H, Zhang SM, Schwarzschild MA, et al. *Obesity and the risk of Parkinson's disease.* Am J Epidemiol, 2004. 159(6): p. 547-55.
97. Palacios N, Gao X, McCullough ML, et al. *Obesity, diabetes, and risk of Parkinson's disease.* Mov Disord, 2011. 26(12): p. 2253-9.
98. Park JH, Choi Y, Kim H, et al. *Association between body weight variability and incidence of Parkinson disease: A nationwide, population-based cohort study.* Eur J Neurol, 2021. 28(11): p. 3626-3633.
99. Jeong SM, Han K, Kim D, et al. *Body mass index, diabetes, and the risk of Parkinson's disease.* Mov Disord, 2019. 35(2): p. 236-244.
100. Nam GE, Kim SM, Han K, et al. *Metabolic syndrome and risk of Parkinson disease: A nationwide cohort study.* PLoS Med, 2018. 15(8): p. e1002640.
101. Noyce AJ, Bandres-Ciga S, Kim J, et al. *The Parkinson's Disease Mendelian Randomization Research Portal.* Mov Disord, 2019. 34(12): p. 1864-1872.
102. Foltynie T, Athauda D. *Diabetes, BMI, and Parkinson's.* Mov Disord, 2020. 35(2): p. 201-203.
103. Hu G, Jousilahti P, Nissinen A, et al. *Body mass index and the risk of Parkinson disease.* Neurology, 2006. 67(11): p. 1955-9.
104. Logroscino G, Sesso HD, Paffenbarger RS, Jr., et al. *Body mass index and risk of Parkinson's disease: a prospective cohort study.* Am J Epidemiol, 2007. 166(10): p. 1186-90.
105. Roos E, Grotta A, Yang F, et al. *Body mass index, sitting time, and risk of Parkinson disease.* Neurology, 2018. 90(16): p. e1413-e1417.
106. Saaksjarvi K, Knekt P, Mannisto S, et al. *Prospective study on the components of metabolic syndrome and the incidence of Parkinson's disease.* Parkinsonism Relat Disord, 2015. 21(10): p. 1148-55.
107. Kahn SE, Hull RL, Utzschneider KM. *Mechanisms linking obesity to insulin resistance and type 2 diabetes.* Nature, 2006. 444(7121): p. 840-6.

108. Athauda D, Foltynie T. *Insulin resistance and Parkinson's disease: A new target for disease modification?* Prog Neurobiol, 2016. 145-146: p. 98-120.
109. Craft S, Stennis Watson G. *Insulin and neurodegenerative disease: shared and specific mechanisms.* The Lancet Neurology, 2004. 3(3): p. 169-178.
110. Yang D, Zhao D, Ali Shah SZ, et al. *The Role of the Gut Microbiota in the Pathogenesis of Parkinson's Disease.* Front Neurol, 2019. 10: p. 1155.
111. Peters BA, Shapiro JA, Church TR, et al. *A taxonomic signature of obesity in a large study of American adults.* Sci Rep, 2018. 8(1): p. 9749.
112. Kyrozis A, Ghika A, Stathopoulos P, et al. *Dietary and lifestyle variables in relation to incidence of Parkinson's disease in Greece.* Eur J Epidemiol, 2013. 28(1): p. 67-77.
113. De Pablo-Fernandez E, Breen DP, Bouloux PM, et al. *Neuroendocrine abnormalities in Parkinson's disease.* J Neurol Neurosurg Psychiatry, 2017. 88(2): p. 176-185.
114. Chohan H, Senkevich K, Patel RK, et al. *Type 2 Diabetes as a Determinant of Parkinson's Disease Risk and Progression.* Mov Disord, 2021. 36(6): p. 1420-1429.
115. Liu W, Tang J. *Association between diabetes mellitus and risk of Parkinson's disease: A prisma-compliant meta-analysis.* Brain Behav, 2021. 11(8): p. e02082.
116. Lu L, Fu DL, Li HQ, et al. *Diabetes and risk of Parkinson's disease: an updated meta-analysis of case-control studies.* PLoS One, 2014. 9(1): p. e85781.
117. Cheong JLY, de Pablo-Fernandez E, Foltynie T, et al. *The Association Between Type 2 Diabetes Mellitus and Parkinson's Disease.* J Parkinsons Dis, 2020.
118. Rhee SY, Han KD, Kwon H, et al. *Association Between Glycemic Status and the Risk of Parkinson Disease: A Nationwide Population-Based Study.* Diabetes Care, 2020.
119. Yang YW, Hsieh TF, Li CI, et al. *Increased risk of Parkinson disease with diabetes mellitus in a population-based study.* Medicine (Baltimore), 2017. 96(3): p. e5921.
120. Xu Q, Park Y, Huang X, et al. *Diabetes and risk of Parkinson's disease.* Diabetes Care, 2011. 34(4): p. 910-5.
121. De Pablo-Fernandez E, Goldacre R, Pakpoor J, et al. *Association between diabetes and subsequent Parkinson disease: A record-linkage cohort study.* Neurology, 2018.

91(2): p. e139-e142.

122. Lee SE, Han K, Baek JY, et al. *Association Between Diabetic Retinopathy and Parkinson Disease: The Korean National Health Insurance Service Database*. *The Journal of clinical endocrinology and metabolism*, 2018. 103(9): p. 3231-3238.

123. Gross A, Racette BA, Camacho-Soto A, et al. *Use of medical care biases associations between Parkinson disease and other medical conditions*. *Neurology*, 2018. 90(24): p. e2155-e2165.

124. Camargo Maluf F, Feder D, Alves de Siqueira Carvalho A. *Analysis of the Relationship between Type II Diabetes Mellitus and Parkinson's Disease: A Systematic Review*. *Parkinsons Dis*, 2019. 2019: p. 4951379.

125. Hu G, Jousilahti P, Bidel S, et al. *Type 2 diabetes and the risk of Parkinson's disease*. *Diabetes Care*, 2007. 30(4): p. 842-7.

126. Simon KC, Chen H, Schwarzschild M, et al. *Hypertension, hypercholesterolemia, diabetes, and risk of Parkinson disease*. *Neurology*, 2007. 69(17): p. 1688-95.

127. Sun Y, Chang YH, Chen HF, et al. *Risk of Parkinson disease onset in patients with diabetes: a 9-year population-based cohort study with age and sex stratifications*. *Diabetes Care*, 2012. 35(5): p. 1047-9.

128. Pupillo E, Cricelli C, Mazzoleni F, et al. *Epidemiology of Parkinson's Disease: A Population-Based Study in Primary Care in Italy*. *Neuroepidemiology*, 2016. 47(1): p. 38-45.

129. Kizza J, Lewington S, Mappin-Kasirer B, et al. *Cardiovascular risk factors and Parkinson's disease in 500,000 Chinese adults*. *Ann Clin Transl Neurol*, 2019. 6(4): p. 624-632.

130. Kummer BR, Diaz I, Wu X, et al. *Associations between cerebrovascular risk factors and parkinson disease*. *Ann Neurol*, 2019. 86(4): p. 572-581.

131. Sanchez-Gomez A, Diaz Y, Duarte-Salles T, et al. *Prediabetes, type 2 diabetes mellitus and risk of Parkinson's disease: A population-based cohort study*. *Parkinsonism Relat Disord*, 2021. 89: p. 22-27.

132. Subramanian I, Mathur S, Oosterbaan A, et al. *Unmet Needs of Women Living*



- with Parkinson's Disease: Gaps and Controversies*. *Mov Disord*, 2022. 37(3): p. 444-455.
133. Clavel-Chapelon F, Group ENS. *Cohort Profile: The French E3N Cohort Study*. *Int J Epidemiol*, 2015. 44(3): p. 801-9.
134. Canonico M, Artaud F, Degaey I, et al. *Incidence of Parkinson's disease in French women from the E3N cohort study over 27 years of follow-up*. *Eur. J. Epidemiol.*, 2022. 37(5): p. 513-523.
135. de Jonge SW, Boldingh QJJ, Solomkin JS, et al. *Effect of postoperative continuation of antibiotic prophylaxis on the incidence of surgical site infection: a systematic review and meta-analysis*. *The Lancet Infectious Diseases*, 2020.
136. Bower JH, Maraganore DM, McDonnell SK, et al. *Incidence and distribution of parkinsonism in Olmsted County, Minnesota, 1976–1990*. *Neurology*, 1999. 52(6): p. 1214.
137. Moisan F, Gourlet V, Mazurie J-L, et al. *Prediction Model of Parkinson's Disease Based on Antiparkinsonian Drug Claims*. *American Journal of Epidemiology*, 2011. 174(3): p. 354-363.
138. de Rijk MC, Tzourio C, Breteler MM, et al. *Prevalence of parkinsonism and Parkinson's disease in Europe: the EUROPARKINSON Collaborative Study*. *European Community Concerted Action on the Epidemiology of Parkinson's disease*. *J Neurol Neurosurg Psychiatry*, 1997. 62(1): p. 10-5.
139. Frigerio R, Elbaz A, Sanft KR, et al. *Education and occupations preceding Parkinson disease: a population-based case-control study*. *Neurology*, 2005. 65(10): p. 1575-83.
140. Insee. *Base des unités urbaines*. 2019 11/05/2019]; Available from: <https://www.insee.fr/fr/information/2115018>.
141. Bellou V, Belbasis L, Tzoulaki I, et al. *Environmental risk factors and Parkinson's disease: An umbrella review of meta-analyses*. *Parkinsonism Relat Disord*, 2016. 23: p. 1-9.
142. Pesce G, Artaud F, Roze E, et al. *Reproductive characteristics, use of exogenous hormones, and Parkinson's disease in women from the E3N cohort study*. *Brain*. in press.
143. Domenighetti C, Sugier PE, Sreelatha AAK, et al. *Mendelian Randomisation Study*

*of Smoking, Alcohol, and Coffee Drinking in Relation to Parkinson's Disease.* J Parkinsons Dis, 2022. 12(1): p. 267-282.

144. Fowler JS, Volkow ND, Wang GJ, et al. *Inhibition of monoamine oxidase B in the brains of smokers.* Nature, 1996. 379(6567): p. 733-6.

145. Hong DP, Fink AL, Uversky VN. *Smoking and Parkinson's disease: does nicotine affect alpha-synuclein fibrillation?* Biochimica et biophysica acta, 2009. 1794(2): p. 282-90.

146. Domenighetti C, Douillard V, Sugier PE, et al. *The Interaction between HLA-DRB1 and Smoking in Parkinson's Disease Revisited.* Mov Disord, 2022. 37(9): p. 1929-1937.

147. Chuang YH, Lee PC, Vlaar T, et al. *Pooled analysis of the HLA-DRB1 by smoking interaction in Parkinson disease.* Ann Neurol, 2017. 82(5): p. 655-664.

148. Van Liere MJ, Lucas F, Clavel F, et al. *Relative validity and reproducibility of a French dietary history questionnaire.* International Journal of Epidemiology, 1997. 26(suppl\_1): p. S128-S128.

149. Favier J-C, Ireland-Ripert J, Toque C, et al. *Répertoire général des aliments : table de composition (General repertory of foods. Composition tables).* Paris: INRA. 1995: 927.

150. Hu FB. *Dietary pattern analysis: a new direction in nutritional epidemiology.* Current opinion in lipidology, 2002. 13 1: p. 3-9.

151. Fung TT, Willett WC, Stampfer MJ, et al. *Dietary Patterns and the Risk of Coronary Heart Disease in Women.* Archives of internal medicine, 2001. 161(15): p. 1857-1862.

152. Yin W, Lof M, Pedersen NL, et al. *Mediterranean Dietary Pattern at Middle Age and Risk of Parkinson's Disease: A Swedish Cohort Study.* Mov Disord, 2020.

153. Sanchez-Sanchez ML, Garcia-Vigara A, Hidalgo-Mora JJ, et al. *Mediterranean diet and health: A systematic review of epidemiological studies and intervention trials.* Maturitas, 2020. 136: p. 25-37.

154. Domenighetti C, Sugier PE, Ashok Kumar Sreelatha A, et al. *Dairy Intake and Parkinson's Disease: A Mendelian Randomization Study.* Mov Disord, 2022. 37(4): p. 857-864.

155. LaHue SC, Comella CL, Tanner CM. *The best medicine? The influence of physical*

- activity and inactivity on Parkinson's disease. Mov Disord, 2016. 31(10): p. 1444-1454.*
156. Ainsworth BE, Haskell WL, Whitt MC, et al. *Compendium of Physical Activities: an update of activity codes and MET intensities. Med. Sci. Sports Exerc., 2000. 32(9): p. S498-S516.*
157. Tehard B, Friedenreich CM, Oppert JM, et al. *Effect of physical activity on women at increased risk of breast cancer: results from the E3N cohort study. Cancer Epidemiol Biomarkers Prev, 2006. 15(1): p. 57-64.*
158. Pols MA, Peeters PHM, Ocke MC, et al. *Estimation of reproducibility and relative validity of the questions included in the EPIC physical activity questionnaire. International Journal of Epidemiology, 1997. 26: p. S181-S189.*
159. Consortium TI. *Validity of a short questionnaire to assess physical activity in 10 European countries. Eur. J. Epidemiol., 2012. 27(1): p. 15-25.*
160. Cust AE, Smith BJ, Chau J, et al. *Validity and repeatability of the EPIC physical activity questionnaire: a validation study using accelerometers as an objective measure. Int J Behav Nutr Phys Act, 2008. 5: p. 33.*
161. Desai RJ, Glynn RJ, Wang S, et al. *Performance of Disease Risk Score Matching in Nested Case-Control Studies: A Simulation Study. Am J Epidemiol, 2016. 183(10): p. 949-57.*
162. Richardson DB. *An incidence density sampling program for nested case-control analyses. Occup Environ Med, 2004. 61(12): p. e59.*
163. Wagner M, Dartigues JF, Samieri C, et al. *Modeling Risk-Factor Trajectories When Measurement Tools Change Sequentially During Follow-up in Cohort Studies: Application to Dietary Habits in Prodromal Dementia. Am J Epidemiol, 2018. 187(4): p. 845-854.*
164. Proust-Lima C, Amieva H, Jacqmin-Gadda H. *Analysis of multivariate mixed longitudinal data: a flexible latent process approach. Br J Math Stat Psychol, 2013. 66(3): p. 470-87.*
165. Thiebaut AC, Benichou J. *Choice of time-scale in Cox's model analysis of epidemiologic cohort data: a simulation study. Stat Med, 2004. 23(24): p. 3803-20.*
166. Lv M, Zhang Y, Chen GC, et al. *Reproductive factors and risk of Parkinson's disease*

- in women: A meta-analysis of observational studies.* Behav Brain Res, 2017. 335: p. 103-110.
167. Sun F, Norman IJ, While AE. *Physical activity in older people: a systematic review.* BMC Public Health, 2013. 13: p. 449.
168. McPhee JS, French DP, Jackson D, et al. *Physical activity in older age: perspectives for healthy ageing and frailty.* Biogerontology, 2016. 17(3): p. 567-80.
169. Yoon SY, Suh JH, Yang SN, et al. *Association of Physical Activity, Including Amount and Maintenance, With All-Cause Mortality in Parkinson Disease.* JAMA Neurol, 2021.
170. Hughes KC, Gao X, Molsberry S, et al. *Physical activity and prodromal features of Parkinson disease.* Neurology, 2019. 93(23): p. e2157-e2169.
171. Evenson KR, Wen F, Herring AH. *Associations of Accelerometry-Assessed and Self-Reported Physical Activity and Sedentary Behavior With All-Cause and Cardiovascular Mortality Among US Adults.* Am J Epidemiol, 2016. 184(9): p. 621-632.
172. LaMonte MJ, Lee IM, Rillamas-Sun E, et al. *Comparison of Questionnaire and Device Measures of Physical Activity and Sedentary Behavior in a Multi-Ethnic Cohort of Older Women.* Journal for the Measurement of Physical Behaviour, 2019. 2(2): p. 82-93.
173. Rothman KJ, Gallacher JE, Hatch EE. *Why representativeness should be avoided.* Int J Epidemiol, 2013. 42(4): p. 1012-4.
174. Batty GD, Shipley M, Tabak A, et al. *Generalizability of occupational cohort study findings.* Epidemiology, 2014. 25(6): p. 932-933.
175. Song S, Luo Z, Li C, et al. *Changes in Body Composition Before and After Parkinson's Disease Diagnosis.* Mov Disord, 2021. 36(7): p. 1617-1623.
176. WHO. *Obesity : preventing and managing the global epidemic : report of a WHO consultation,* World Health Organization: Geneva. 2000.
177. Visscher TL, Seidell JC, Molarius A, et al. *A comparison of body mass index, waist-hip ratio and waist circumference as predictors of all-cause mortality among the elderly: the Rotterdam study.* International journal of obesity and related metabolic disorders : journal of the International Association for the Study of Obesity, 2001. 25(11): p. 1730-5.

178. Tehard B, Liere MJV, Noug  CC, et al. *Anthropometric Measurements and Body Silhouette of Women*. Journal of the American Dietetic Association, 2002. 102(12): p. 1779-1784.
179. Yan J, Qiao L, Tian J, et al. *Effect of statins on Parkinson's disease: A systematic review and meta-analysis*. Medicine (Baltimore), 2019. 98(12): p. e14852.
180. Costello H, Berry AJ, Reeves S, et al. *Disrupted reward processing in Parkinson's disease and its relationship with dopamine state and neuropsychiatric syndromes: a systematic review and meta-analysis*. J Neurol Neurosurg Psychiatry, 2021. 93(5): p. 555-62.
181. Vazquez G, Duval S, Jacobs DR, Jr., et al. *Comparison of body mass index, waist circumference, and waist/hip ratio in predicting incident diabetes: a meta-analysis*. Epidemiol Rev, 2007. 29: p. 115-28.
182. Labrecque JA, Swanson SA. *Interpretation and Potential Biases of Mendelian Randomization Estimates With Time-Varying Exposures*. Am J Epidemiol, 2019. 188(1): p. 231-238.
183. Winkler TW, Justice AE, Graff M, et al. *The Influence of Age and Sex on Genetic Associations with Adult Body Size and Shape: A Large-Scale Genome-Wide Interaction Study*. PLoS Genet, 2015. 11(10): p. e1005378.
184. Lawlor DA, Tilling K, Davey Smith G. *Triangulation in aetiological epidemiology*. Int J Epidemiol, 2016. 45(6): p. 1866-1886.
185. Bassil F, Fernagut PO, Bezdard E, et al. *Insulin, IGF-1 and GLP-1 signaling in neurodegenerative disorders: targets for disease modification?* Prog Neurobiol, 2014. 118: p. 1-18.
186. O'Brien PD, Hinder LM, Callaghan BC, et al. *Neurological consequences of obesity*. The Lancet Neurology, 2017. 16(6): p. 465-477.
187. Ross R, Neeland IJ, Yamashita S, et al. *Waist circumference as a vital sign in clinical practice: a Consensus Statement from the IAS and ICCR Working Group on Visceral Obesity*. Nat Rev Endocrinol, 2020. 16(3): p. 177-189.
188. Yue X, Li H, Yan H, et al. *Risk of Parkinson Disease in Diabetes Mellitus: An Updated*

*Meta-Analysis of Population-Based Cohort Studies*. Medicine (Baltimore), 2016. 95(18): p. e3549.

189. de Lauzon-Guillain B, Balkau B, Charles MA, et al. *Birth weight, body silhouette over the life course, and incident diabetes in 91,453 middle-aged women from the French Etude Epidemiologique de Femmes de la Mutuelle Generale de l'Education Nationale (E3N) Cohort*. Diabetes Care, 2010. 33(2): p. 298-303.

190. Fagherazzi G, Vilier A, Affret A, et al. *The association of body shape trajectories over the life course with type 2 diabetes risk in adulthood: a group-based modeling approach*. Ann Epidemiol, 2015. 25(10): p. 785-7.

191. Fagherazzi G, Vilier A, Lajous M, et al. *Wine consumption throughout life is inversely associated with type 2 diabetes risk, but only in overweight individuals: results from a large female French cohort study*. Eur J Epidemiol, 2014. 29(11): p. 831-9.

192. Mancini FR, Affret A, Dow C, et al. *Dietary antioxidant capacity and risk of type 2 diabetes in the large prospective E3N-EPIC cohort*. Diabetologia, 2017. 61(2): p. 308-316.

193. Mancini FR, Affret A, Dow C, et al. *High dietary phosphorus intake is associated with an increased risk of type 2 diabetes in the large prospective E3N cohort study*. Clin Nutr, 2018. 37(5): p. 1625-1630.

194. Mancini FR, Dow C, Affret A, et al. *Micronutrient dietary patterns associated with type 2 diabetes mellitus among women of the E3N-EPIC (Etude Epidemiologique aupres de femmes de l'Education Nationale) cohort study*. J Diabetes, 2018. 10(8): p. 665-674.

195. Laouali N, Mancini FR, Hajji-Louati M, et al. *Dietary inflammatory index and type 2 diabetes risk in a prospective cohort of 70,991 women followed for 20 years: the mediating role of BMI*. Diabetologia, 2019.

196. Mancini FR, Rajaobelina K, Dow C, et al. *High iodine dietary intake is associated with type 2 diabetes among women of the E3N-EPIC cohort study*. Clin Nutr, 2019. 38(4): p. 1651-1656.

197. Laouali N, Berrandou T, J AR, et al. *Profiles of Polyphenol Intake and Type 2 Diabetes Risk in 60,586 Women Followed for 20 Years: Results from the E3N Cohort Study*. Nutrients, 2020. 12(7).

198. Federation WHOID. *Definition and diagnosis of diabetes mellitus and intermediate hyperglycaemia : report of a WHO/IDF consultation*, World Health Organization: Geneva. 2006.
199. Kilpatrick ES. *Haemoglobin A1c in the diagnosis and monitoring of diabetes mellitus*. J Clin Pathol, 2008. 61(9): p. 977-82.
200. Saudek CD, Herman WH, Sacks DB, et al. *A new look at screening and diagnosing diabetes mellitus*. The Journal of clinical endocrinology and metabolism, 2008. 93(7): p. 2447-53.
201. US Preventive Services Task Force, Mangione CM, Barry MJ, et al. *Statin Use for the Primary Prevention of Cardiovascular Disease in Adults: US Preventive Services Task Force Recommendation Statement*. JAMA, 2022. 328(8): p. 746-753.
202. Komici K, Femminella GD, Bencivenga L, et al. *Diabetes Mellitus and Parkinson's Disease: A Systematic Review and Meta-Analyses*. J Parkinsons Dis, 2021. 11(4): p. 1585-1596.
203. Cereda E, Barichella M, Pedrolli C, et al. *Diabetes and risk of Parkinson's disease*. Mov Disord, 2013. 28(2): p. 257.
204. Noyce AJ, Bestwick JP, Silveira-Moriyama L, et al. *Meta-analysis of early nonmotor features and risk factors for Parkinson disease*. Annals of Neurology, 2012. 72(6): p. 893-901.
205. Cereda E, Barichella M, Pedrolli C, et al. *Diabetes and risk of Parkinson's disease: a systematic review and meta-analysis*. Diabetes care, 2011. 34(12): p. 2614-2623.
206. Foltynie T, Athauda D. *Repurposing anti-diabetic drugs for the treatment of Parkinson's disease: Rationale and clinical experience*. Prog Brain Res, 2020. 252: p. 493-523.
207. Kent DM, Leung LY, Puttock EJ, et al. *Development of Parkinson Disease and Its Relationship with Incidentally Discovered White Matter Disease and Covert Brain Infarction in a Real-World Cohort*. Ann Neurol, 2022. 92(4): p. 620-630.
208. Lawlor DA, Harbord RM, Sterne JA, et al. *Mendelian randomization: using genes as instruments for making causal inferences in epidemiology*. Stat. Med, 2008. 27(8): p.

1133-1163.

209. Smit RAJ, Trompet S, Dekkers OM, et al. *Survival Bias in Mendelian Randomization Studies: A Threat to Causal Inference*. *Epidemiology*, 2019. 30(6): p. 813-816.

210. Athauda D, Maclagan K, Skene SS, et al. *Exenatide once weekly versus placebo in Parkinson's disease: a randomised, double-blind, placebo-controlled trial*. *The Lancet*, 2017. 390(10103): p. 1664-1675.

211. Brauer R, Wei L, Ma T, et al. *Diabetes medications and risk of Parkinson's disease: a cohort study of patients with diabetes*. *Brain*, 2020. 143(10): p. 3067-3076.

212. Sunnarborg K, Tiihonen M, Huovinen M, et al. *Association between different diabetes medication classes and risk of Parkinson's disease in people with diabetes*. *Pharmacoepidemiol Drug Saf*, 2022. 31(8): p. 875-882.





# **ANNEXES**



## **Annex 1: Résumé substantiel.**

### **Contexte et objectifs**

La maladie de Parkinson (MP) se caractérise par des symptômes moteurs, notamment le tremblement de repos, la rigidité, la bradykinésie et l'instabilité posturale, qui représentent les principales caractéristiques de la MP.[1] De plus, des symptômes non-moteurs, tels que des troubles de l'odorat, une constipation, des troubles du sommeil paradoxal (RBD), des troubles urinaires ou une dépression peuvent également être présents, voire précéder chez certains patients l'apparition des symptômes moteurs de plusieurs années ou décennies.[2-6] La présence de symptômes non-moteurs avant le diagnostic de la MP caractérise la phase prodromale de la maladie.[7]

La MP est une maladie liée à l'âge, rare avant 50 ans, dont l'incidence augmente nettement après cet âge.[8] La MP est la deuxième maladie neurologique la plus fréquente après la maladie d'Alzheimer et celle caractérisée par la plus forte croissance au niveau mondial en termes de prévalence, d'invalidité et de décès.[9] Les taux d'incidence de la MP sont environ 1,5 plus élevés chez les hommes que chez les femmes.[10] Cette différence pourrait s'expliquer par des expositions professionnelles différentes chez les hommes et les femmes[11-13] et/ou par un effet neuro-protecteur des œstrogènes.[14-16] En raison de cette différence en fonction du sexe, les femmes constituent un groupe qui a été moins étudié dans la recherche sur la MP.

La MP est une maladie multifactorielle causée dans la plupart des cas par une combinaison de facteurs génétiques et environnementaux. La majorité des cas de MP sont considérés comme des cas idiopathiques. Plusieurs expositions ont été associées au risque de MP mais, en raison de la longue phase prodromale de la MP, un biais de causalité inverse pourrait contribuer à expliquer certaines des associations précédemment rapportées dans les études épidémiologiques.[17] Par exemple, des symptômes non moteurs de la MP, tels que la constipation, peuvent entraîner des changements dans le régime alimentaire et le poids ; de même, des symptômes

moteurs discrets pourraient amener certaines personnes à réduire leur niveau d'activité physique dans les années précédant le diagnostic de la MP. Cette question représente un risque majeur pour les études épidémiologiques, en particulier pour celles avec un suivi court et qui ont évalué les expositions à proximité du diagnostic de la MP.[5]

Des méthodes statistiques appropriées sont donc nécessaires pour prendre en compte le risque de causalité inverse. En particulier, l'inclusion dans les études de cohorte d'un *lag*, c'est-à-dire un délai entre l'exposition et l'incidence de MP, permet de réduire le risque de biais dû à la causalité inverse,[18, 19] mais peu d'études de cohorte ont un suivi suffisamment long pour pouvoir le faire tout en maintenant un nombre suffisant de cas. Les analyses de trajectoires permettent également d'examiner si une causalité inverse pourrait modifier les trajectoires des expositions chez les cas de MP avant l'apparition de la maladie, mais peu d'études ont utilisé cette approche.

Dans cette thèse, notre objectif était d'examiner la relation entre deux comportements de santé potentiellement modifiables, l'activité physique (premier objectif) et l'adiposité (deuxième objectif), et le diabète avec l'incidence de la MP chez environ 100 000 femmes de l'étude de cohorte française E3N suivies pendant 29 ans, tout en tenant compte du risque de causalité inverse. Nous avons également utilisé une étude cas-témoins nichée dans la cohorte pour comparer les trajectoires des différentes expositions avant le diagnostic de MP.

## **Population d'étude**

L'étude de cohorte E3N est une étude de cohorte française prospective, toujours en cours, portant sur 98 995 femmes, nées entre 1925 et 1950 et recrutées en 1990, qui étaient affiliées à un régime d'assurance maladie national français couvrant principalement les enseignants (Mutuelle Générale de l'Education Nationale, MGEN). Des questionnaires de suivi ont été envoyés tous les 2 ou 3 ans. A ce jour, 11 vagues de collecte de données sont disponibles (la dernière en 2014, questionnaire 11). De

plus, depuis 2004, les femmes sont suivies passivement grâce aux bases de données de la MGEN qui incluent les remboursements de médicaments et les consultations.

Les cas de MP potentiels ont été identifiés grâce aux diagnostics de MP auto-déclarés par les femmes dans les questionnaires de suivi, aux remboursements de médicaments antiparkinsoniens et aux certificats de décès. Lorsque cela était possible, les cas de MP potentiels ont été contactés par courrier pour confirmer le diagnostic. Pour les femmes qui ont confirmé un diagnostic de MP ou de syndrome parkinsonien et pour celles qui n'ont pas pu être contactées (en cas de décès ou de refus de contact), nous avons sollicité leurs neurologues traitants, ou à défaut leurs médecins traitants, pour obtenir des renseignements cliniques détaillés (année d'apparition des symptômes et de diagnostic, signes moteurs et autres symptômes neurologiques, utilisation de neuroleptiques, type de traitement, réactivité au traitement, etc). Sur la base de la documentation médicale disponible, le statut vis-à-vis de la MP (certain, probable, possible, pas de MP) a été déterminé par un groupe d'experts. Seules les patientes atteintes de MP certaine ou probable ont été retenues dans les analyses. Lorsque la documentation médicale n'était pas disponible, nous avons utilisé un algorithme validé qui repose sur les demandes de remboursement de médicaments antiparkinsoniens et les consultations neurologiques.[20, 21] Parmi les femmes dont la MP a été confirmée, le diagnostic était basé sur les dossiers médicaux pour 62% et sur l'algorithme pour 38%.[21] L'année du diagnostic de la MP a été définie comme l'année du diagnostic disponible dans les dossiers médicaux ou, par ordre décroissant de priorité, l'année du diagnostic auto-déclaré, l'année de la première utilisation de médicaments antiparkinsoniens et l'année du premier questionnaire où la MP a été auto-déclarée. Les taux d'incidence de la MP dans l'E3N sont en accord avec ceux observés chez les femmes d'Europe occidentale entre 1992-2018 selon le *Global Burden of Disease*, une observation qui est en faveur de la validité de notre stratégie d'identification et de validation des cas.[21]

# Objectif I - Association entre l'activité physique et la maladie de Parkinson

## Introduction

Des essais cliniques et des études observationnelles récents sont en faveur de l'effet bénéfique de l'exercice sur les symptômes moteurs et non-moteurs chez les patients atteints de MP,[22, 23] et l'exercice est considéré comme l'intervention non pharmacologique la plus prometteuse dans la MP.[24, 25] Des études antérieures ont également montré une association inverse entre l'activité physique et le risque ultérieur de MP, mais cette association était uniquement présente chez les hommes et pas chez les femmes dans les analyses stratifiées par sexe.[26] En outre, la plupart des études ont utilisé une seule mesure d'activité physique à l'inclusion et des analyses avec des *lag* courts ( $\leq 10$  ans) ont été réalisées pour prendre en compte la causalité inverse dans quelques-unes de ces études. Par conséquent, des études prospectives avec un suivi plus long sont nécessaires chez les femmes pour étudier cette question.

## Méthodes

Le niveau d'activité physique a été évalué dans six questionnaires (Q1, Q3, Q5, Q7, Q8, Q11). Les questions étant différentes d'un questionnaire à l'autre, nous avons utilisé un modèle mixte à processus latent avec intercept et pente aléatoires pour créer une variable latente (catégorisée en quartiles) qui capture la trajectoire longitudinale non gaussienne de l'activité physique.[27]

Des modèles de Cox pour variables dépendantes du temps avec l'âge comme échelle de temps ont été utilisés et ajustés sur les facteurs de confusion (lieu de résidence, âge aux premières règles, parité, tabagisme, statut ménopausique). Notre analyse principale repose sur l'activité physique (en quartiles) avec un *lag* de 10 ans pour prendre en compte la causalité inverse ; dans des analyses de sensibilité, nous avons utilisé des *lags* de 5, 15 et 20 ans. Des modèles linéaires mixtes ajustés sur les

facteurs de confusion ont été mis en œuvre afin d'examiner les trajectoires d'activité physique (variable continue) au cours de 29 ans de suivi avant le diagnostic de MP dans le cadre d'une étude cas-témoins nichée dans E3N.

## Résultats

Dans notre analyse principale (avec un *lag* de 10 ans), nous avons suivi 95 354 femmes pendant 19 ans (moyenne = 17,2, écart-type = 3,3), dont 1 074 ont développé une MP. Par rapport aux études précédentes, la nôtre a inclus le plus grand nombre de patientes atteintes de MP et a le plus long suivi.

Les analyses des trajectoires d'activité physique latente ont montré qu'après une augmentation initiale de l'activité physique chez les cas et les témoins, l'activité physique diminuait, avec un déclin plus marqué chez les cas que chez les témoins. L'activité physique était significativement plus faible chez les cas que chez les témoins tout au long du suivi. La différence entre les cas et les témoins commençait à augmenter environ 10 ans avant le diagnostic de MP ; par conséquent, la différence était plus importante à la date du diagnostic de la MP qu'au début de l'étude.

Dans notre analyse principale (*lag* de 10 ans), le risque de MP diminuait avec l'augmentation de l'activité physique (*P*-tendance = 0,001), avec une incidence inférieure de 25 % chez les participantes du quartile le plus élevé par rapport à celles du quartile le plus bas (rapport de risque [HR]=0,75, intervalle de confiance [IC] 95% = 0,63-0,89). L'utilisation de *lags* plus longs a conduit à des associations similaires; l'association inverse était proche de la signification pour le *lag* de 20 ans (*P*-tendance=0,06), probablement en raison d'un nombre plus faible de cas de MP.

## Discussion

Dans cette étude de cohorte de ~100 000 femmes avec six mesures d'activité physique au cours de 29 ans de suivi, l'augmentation de l'activité physique était associée à une réduction de l'incidence de la MP, tout en tenant compte du risque de causalité inverse.



Les analyses des trajectoires d'activité physique ont montré que les cas de MP avaient des niveaux d'activité physique inférieurs à ceux des témoins 29 ans avant le diagnostic, et que les différences entre cas et témoins augmentaient environ 10 ans avant le diagnostic de la MP, en accord avec l'hypothèse selon laquelle la MP prodromale entraînerait une réduction de l'activité physique.

Ces résultats renforcent l'évidence en faveur des bénéfices de l'activité physique pour la santé et justifient des études supplémentaires pour comprendre quel type et quel niveau d'activité physique seraient les plus bénéfiques pour la MP. Ces résultats sont importants pour la planification des études d'intervention dont l'objectif serait de prévenir ou de retarder la survenue de la MP.

## **Objectif II - Association entre l'adiposité et la maladie de Parkinson**

### **Introduction**

La relation entre l'indice de masse corporelle (IMC) et le risque de MP est complexe et les études précédentes ont conduit à des résultats contradictoires. Une méta-analyse de 10 études de cohorte n'a montré aucune association entre l'IMC continu et la MP, mais avec toutefois une grande hétérogénéité.[28] De plus, des études de randomisation mendélienne (RM) ont montré qu'un IMC plus élevé serait associé à un risque de MP plus faible, en faveur d'une association inverse causale.[29, 30] Des études moins nombreuses portant sur d'autres mesures d'adiposité (tour de taille, WC ; rapport taille-hanche, WHR ; rapport taille-hauteur, WHtR) ont également donné des résultats peu cohérents.[31-35]

La perte de poids est fréquente chez les patients atteints de MP, qui ont un poids moyen inférieur à celui des témoins, avec une différence qui augmente avec la durée de la maladie.[36] En outre, comme discuté précédemment, les symptômes

prodromaux de la MP (par exemple, la constipation ou la fatigue) peuvent entraîner des changements dans le régime alimentaire, l'activité physique ou d'autres comportements conduisant à des changements de poids avant le diagnostic. Par conséquent, les résultats des études antérieures pourraient avoir été influencés par la causalité inverse. En outre, toutes les études de cohorte précédentes étaient basées sur une seule mesure de l'IMC à l'inclusion.

Des études de cohorte avec un long suivi sont donc nécessaires pour examiner la relation entre l'IMC et la MP, tout en essayant de minimiser l'influence de la causalité inverse. De plus, des études avec des mesures répétées seraient utiles pour étudier les trajectoires avant la survenue de la MP.

## Méthodes

Le poids a été auto-déclaré à tous les questionnaires (11 mesures). La taille a été auto-déclarée dans huit questionnaires (Q1, Q4, Q6-Q11) et standardisée en utilisant la valeur la plus fréquente. L'IMC a été calculé comme le poids divisé par la taille au carré ( $\text{kg/m}^2$ ) et catégorisé selon quatre niveaux (OMS) : maigre,  $<18,5$  ; normal (catégorie de référence),  $[18,5-25,0[$  ; surpoids,  $[25,0-30,0[$  ; obésité,  $\geq 30,0 \text{ kg/m}^2$ ).[37] Le WC et la circonférence de la hanche (HC) ont été évalués dans six questionnaires (Q4, Q7-Q11). Nous avons calculé le WHR comme le WC divisé par l'HC et le WHtR comme le WC divisé par la taille (tous en cm). Le WC a été catégorisé en trois catégories :  $\leq 80$ ,  $]80-88]$ ,  $>88 \text{ cm}$  ; le WHR en deux catégories :  $<0,85$  (référence),  $\geq 0,85$  ; le WHtR en trois catégories :  $<0,5$ ,  $[0,5-0,6[$ ,  $\geq 0,6$ . Pour le WC et le WHtR, nous avons utilisé la catégorie intermédiaire comme référence pour examiner les relations quadratiques.[38]

Nous avons utilisé des modèles de Cox pour variables dépendantes du temps, avec l'âge comme échelle de temps et ajustés sur les facteurs de confusion (parité, âge aux premières règles, lieu de résidence, tabagisme, statut ménopausique, activité physique) ; les analyses pour le WC ont également été ajustées sur la taille. Pour tenir compte du risque de causalité inverse, nos analyses principales ont utilisé un *lag* de 5

ans ; dans des analyses de sensibilité, nous avons utilisé des *lags* de 10 et 20 ans. De plus, les trajectoires des catégories d'IMC sur 29 ans de suivi (échelle de temps rétrospective) ont été examinées dans une étude cas-témoins nichée dans E3N à l'aide de modèles logistiques multinomiaux à équations d'estimation généralisées et ajustés sur les facteurs de confusion.

## Résultats

Dans notre analyse principale (*lag* de 5 ans), nous avons suivi 96 702 femmes pendant 24 ans (moyenne = 22,0, écart-type = 4,0), dont 1 164 ont développé une MP. L'incidence de la MP était inférieure de 24% dans le groupe obèse par rapport au groupe avec un IMC normal (HR = 0,76, IC 95% = 0,59-0,98). Les femmes en surpoids présentaient également une diminution du risque de MP qui n'était pas statistiquement significative (HR = 0,89, IC 95% = 0,78-1,03), et il n'y avait pas d'association pour le groupe en maigreur (HR = 0,99, IC 95% = 0,71-1,37). L'incidence de la MP diminuait avec l'augmentation de l'IMC (*P*-linéaire = 0,016). Des *lags* plus longs ont donné des résultats similaires.

Les femmes avec un WC plus élevé ( $\geq 88$  cm) avaient une diminution non statistiquement significative du risque de MP (HR = 0,84, IC 95% = 0,70-1,03) par rapport à celles du groupe de référence ([80-88] cm). Les femmes avec un WC faible ( $\leq 80$  cm) présentaient également un risque diminué, mais non statistiquement significatif, de MP (HR = 0,88, IC 95% = 0,76-1,04). Dans l'analyse avec un *lag* de 10 ans, l'association inverse est devenue significative pour les participantes avec le WC le plus élevé (HR = 0,74, IC 95% = 0,59-0,94). Des résultats similaires ont été observés pour le WHtR tandis qu'il n'y avait pas d'association avec le WHR.

Les analyses des trajectoires d'IMC ont montré que la fréquence de l'obésité augmentait chez les témoins au cours du suivi, alors qu'elle commençait à diminuer cinq à dix ans avant le diagnostic chez les patientes parkinsoniennes (*P*-interaction=0,019). L'obésité était significativement moins fréquente chez les patientes

que chez les témoins chaque année, y compris 29 ans avant la date de diagnostic de la MP. Pour les femmes en maigreur et en surpoids, il n'y avait pas de différence dans les trajectoires des cas par rapport à celles des témoins.

## **Discussion**

L'obésité était associée à une plus faible incidence de la MP chez les femmes de E3N, même lorsqu'elle a été évaluée 20 ans avant le diagnostic de MP. Des résultats similaires étaient observés pour un WC et un WHtR plus élevés, avec des associations significatives dans les analyses avec un *lag* de 10 ans. Les analyses des trajectoires d'IMC ont montré que l'obésité était moins fréquente chez les cas que chez les témoins chaque année, et la différence était statistiquement significative 29 ans avant le diagnostic. De plus, l'IMC commençait à diminuer 5 à 10 ans avant le diagnostic chez les cas.

Ces résultats sont cohérents avec les études de RM récentes et justifient des investigations supplémentaires pour comprendre les mécanismes sous-jacents. Nos analyses soulignent l'importance des analyses avec *lags* en raison des changements de l'IMC au cours de la période prodromale de la MP.

## **Objectif III - Association entre le diabète et la maladie de Parkinson**

### **Introduction**

Plusieurs études de cohorte ont examiné la relation entre le diabète et le risque de MP, avec des résultats contradictoires. Une méta-analyse de neuf études prospectives a montré que le diabète était associé à une augmentation de 29% de l'incidence de la MP (RR = 1,29, IC 95% = 1,15-1,45), mais avec une hétérogénéité marquée.[39] Parmi ces neuf études, cinq étaient des études de cohorte prospectives, tandis que quatre

s'appuyaient sur des bases de données administratives. Toutes les études de cohorte prospectives ont validé les diagnostics de MP et de diabète, sauf une dans laquelle le diagnostic de MP reposait uniquement sur des remboursements de médicaments.[40] Dans les études basées sur des bases de données administratives, les diagnostics de MP et de diabète reposaient sur différents algorithmes mais n'étaient pas validés. A l'inverse, d'après une méta-analyse de 14 études cas-témoins, le diabète était associé à un risque de MP diminué. Ces résultats doivent être interprétés avec prudence en raison d'un certain nombre de biais potentiels, notamment un biais de surveillance, la causalité inverse, la confusion résiduelle, des erreurs de classement sur le diagnostic, un biais de survie ou une définition inadéquate de la date de diagnostic de MP dans les bases de données administratives. Des études avec un long suivi qui examinent le rôle du diabète incident, la durée du diabète et du recours aux soins médicaux sont nécessaires.

## Méthodes

Les participantes diabétiques ont été identifiées à l'aide de deux sources de données : les déclarations des participantes dans les questionnaires E3N jusqu'en 2003 et les bases de remboursements de médicaments entre 2004 et 2016. Au total, 6 658 femmes diabétiques ont été identifiées. Bien que le processus de validation n'ait pas permis de distinguer le diabète de type 1 et type 2, la tranche d'âge de notre population implique que les cas incidents de diabète sont principalement des cas de diabète de type 2.

Nous avons utilisé des modèles de Cox pour variables dépendantes du temps avec l'âge comme échelle de temps et ajustés sur les facteurs de confusion (parité, âge aux premières règles, lieu de résidence, tabagisme, statut ménopausique, activité physique, IMC, hypertension, hypercholestérolémie). Pour nos analyses principales, nous n'avons pas inclus de *lag* entre l'exposition et l'incidence de la MP ; nous avons examiné l'association entre le statut diabétique, la durée du diabète et l'âge au début du diabète avec l'incidence de la MP. Nous avons ensuite examiné l'association entre le statut diabétique et l'incidence de la MP en incluant des *lags* de 5, 10 et 15 ans. Afin

d'examiner le rôle du diabète incident, nous avons répété ces analyses après avoir exclu les femmes présentant un diabète prévalent (diagnostiqué avant le début du suivi en 1990). Dans les analyses débutant en 2004, il a été possible de procéder à des ajustements supplémentaires sur le nombre de consultations médicales, les statines lipophiles et les maladies cardiovasculaires.

Les trajectoires du diabète avant le diagnostic de la MP sur 29 ans de suivi (échelle de temps rétrospective) ont été examinées dans une étude cas-témoins nichée dans E3N à l'aide de modèles logistiques à équations d'estimation généralisées et ajustés sur les variables de confusion.

Une méta-analyse a également été réalisée à l'aide de modèles à effets aléatoires afin de mettre à jour les estimations des méta-analyses antérieures en incluant nos propres résultats. Nous avons effectué des analyses distinctes pour les études de cohorte prospectives et celles basées sur des bases de données administratives. Nous avons également examiné l'influence des études qui incluaient un *lag* entre le diabète et l'incidence de la MP ou qui examinaient le rôle de la durée du diabète.

## Résultats

Dans notre analyse principale, nous avons suivi 98 069 femmes pendant 29 ans (moyenne = 26,7, écart-type = 4,8), dont 1 200 ont développé une MP. Soixante-trois femmes diabétiques ont développé une MP. Le diabète (prévalent et incident) n'était pas associé à l'incidence de la MP (HR = 1,04, IC 95% = 0,80-1,35) dans les analyses sans *lag* ; l'inclusion de *lags* croissants n'a pas modifié les résultats. La durée du diabète et l'âge au diagnostic du diabète ne semblaient pas jouer de rôle. Les analyses limitées aux cas incidents ou à la période 2004-2018 (lorsque les bases de remboursements de médicaments sont devenues disponibles) et l'ajustement sur le nombre de consultations médicales, l'utilisation de statines ou les antécédents de maladies cardiovasculaires ont abouti à des résultats similaires.

Dans l'analyse de trajectoires de la prévalence du diabète, aucune différence n'a été retrouvée entre cas et témoins.

Une méta-analyse de 14 études, dont huit études basées sur des bases de données administratives et six études de cohorte prospectives, a montré que le diabète était associé à un risque plus élevé de MP (RR = 1,35, IC 95% = 1,33-1,37,  $I^2 = 93,6\%$ ,  $P$ -hétérogénéité =  $<0,001$ ). L'association était légèrement plus faible dans les études prospectives (RR = 1,26, IC 95% = 1,13-1,40) que dans celles basées sur des bases de données administratives (RR = 1,35, IC 95% = 1,34-1,37), mais la différence n'était pas statistiquement significative ( $P = 0,182$ ). Après avoir exclu une étude de cohorte prospective dans laquelle les cas de MP étaient identifiés à partir des remboursements de médicaments,[40] l'association diminuait dans les études prospectives (RR = 1,22, IC 95% = 1,10-1,37) et la différence entre les deux types d'études était proche de la signification ( $P = 0,078$ ). En prenant en compte les études qui incluaient un *lag* entre le diabète et l'incidence de la MP ou qui examinaient le rôle de la durée du diabète, l'association a diminué parmi les études prospectives de 7 % (HR = 1,19, IC 95% = 1,04-1,36) et de 12% (RR = 1,14, IC 95% = 1,00-1,31) après exclusion de l'étude de cohorte prospective dans laquelle les cas de MP ont été identifiés à partir des remboursements de médicaments.[40] La différence entre les études prospectives et celles basées sur des bases de données administratives devenait alors significative ( $P = 0,029$ ).

## Discussion

Le diabète n'était pas associé à l'incidence de la MP parmi les participantes d'E3N, même lorsqu'il était évalué 5, 10 ou 15 ans avant le diagnostic de MP. Les analyses des trajectoires du diabète ont montré que la fréquence du diabète était similaire chez les patientes atteintes de MP et les témoins au cours des 29 années précédant le diagnostic de MP. Notre méta-analyse suggère que les études de cohorte prospectives et celles basées sur des bases de données administratives aboutissent à des résultats différents, et différents phénomènes pourraient expliquer les différences entre les deux types d'études, notamment un biais de surveillance, la causalité inverse, la confusion

résiduelle, des erreurs de diagnostic, un biais de survie ou la définition inadéquate de la date de diagnostic de MP dans les études à partir de bases de données administratives. En conclusion, des études supplémentaires, de grande envergure et bien conçues, sont nécessaires pour examiner si le diabète est associé à la MP.

## **Perspectives et conclusions**

En raison de l'absence de traitement curatif de la MP et de l'augmentation du nombre de patients atteints de cette maladie dans le monde, il est urgent de trouver des caractéristiques modifiables du mode de vie qui pourraient prévenir ou retarder l'apparition de la maladie. Toutefois, la causalité inverse représente un écueil majeur pour les études épidémiologiques visant à identifier des mesures préventives en raison de la longue phase prodromale de la MP, et doit être prise en compte de manière adéquate.

À l'heure actuelle, l'activité physique semble être l'intervention présentant la meilleure combinaison de faisabilité et de sécurité qui pourrait être testée chez des patients présentant des symptômes prodromaux ou à risque génétique de MP. Des études complémentaires sont nécessaires pour mieux comprendre quels types d'activités pourraient être plus bénéfiques et les mécanismes biologiques impliqués dans cette association.

Nos résultats sur l'association entre l'adiposité et la MP n'ont pas de traduction immédiate en termes de prévention, étant donné les nombreux autres effets néfastes de l'obésité, mais ils peuvent fournir des pistes quant aux mécanismes impliqués dans la MP, et des études supplémentaires sont nécessaires pour examiner cette question. Enfin, d'autres études sont nécessaires pour examiner la relation entre le diabète et la MP.

Au cours des dernières années, notre connaissance de la MP s'est considérablement améliorée, en particulier avec la reconnaissance d'une longue phase



prodromale et l'identification de personnes à risque élevé de MP (symptômes prodromaux, porteurs de mutations génétiques). Bien que l'on commence à envisager des essais de prévention de la MP, de nombreuses questions demeurent, et des études épidémiologiques supplémentaires sont nécessaires pour aider à identifier les interventions (p. ex. comportements liés au mode de vie, médicaments) qui pourraient être testées, seules ou en combinaison, dans de tels essais, tout en tenant compte de manière adéquate du risque de causalité inverse.

## Références

1. Kalia, L.V., Lang, A.E. *Parkinson's disease*. The Lancet, 2015. 386(9996): p. 896-912.
2. Savica, R., Boeve, B.F., Mielke, M.M. *When Do alpha-Synucleinopathies Start? An Epidemiological Timeline: A Review*. JAMA Neurol, 2018. 75(4): p. 503-509.
3. Schrag, A., Horsfall, L., Walters, K., et al. *Prediagnostic presentations of Parkinson's disease in primary care: a case-control study*. The Lancet Neurology, 2015. 14(1): p. 57-64.
4. Postuma, R.B., Berg, D. *Advances in markers of prodromal Parkinson disease*. Nat Rev Neurol, 2016. 12(11): p. 622-634.
5. Elbaz, A. *Prodromal symptoms of Parkinson's disease: Implications for epidemiological studies of disease etiology*. Rev Neurol, 2016. 172(8-9): p. 503-511.
6. Tolosa, E., Garrido, A., Scholz, S.W., et al. *Challenges in the diagnosis of Parkinson's disease*. The Lancet Neurology, 2021. 20(5): p. 385-397.
7. Berg, D., Postuma, R.B., Adler, C.H., et al. *MDS research criteria for prodromal Parkinson's disease*. Mov Disord, 2015. 30(12): p. 1600-11.
8. Elbaz, A., Artaud, F., Canonico, M., et al. *Épidémiologie descriptive et étiologique de la maladie de Parkinson*, in *Pathologies du mouvement*, C. Tranchant, J.P. Azulay, Editors: Doin. 2020: 37-48.
9. Dorsey, E.R., Elbaz, A., Nichols, E., et al. *Global, regional, and national burden of Parkinson's disease, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016*. The Lancet Neurology, 2018. 17(11): p. 939-953.
10. Moisan, F., Kab, S., Mohamed, F., et al. *Parkinson disease male-to-female ratios increase with age: French nationwide study and meta-analysis*. J Neurol Neurosurg Psychiatry, 2016. 87(9): p. 952-7.
11. Darweesh, S.K.L., Ikram, M.K., Faber, M.J., et al. *Professional occupation and the risk of Parkinson's disease*. Eur J Neurol, 2018. 25(12): p. 1470-1476.
12. Gunnarsson, L.G., Bodin, L. *Occupational Exposures and Neurodegenerative Diseases-A Systematic Literature Review and Meta-Analyses*. Int J Environ Res Public Health, 2019. 16(3).
13. Vlaar, T., Kab, S., Schwaab, Y., et al. *Association of Parkinson's disease with industry sectors: a French nationwide incidence study*. Eur J Epidemiol, 2018. 33(11): p. 1101-1111.
14. Elbaz, A., Carcaillon, L., Kab, S., et al. *Epidemiology of Parkinson's disease*. Rev Neurol (Paris), 2016. 172(1): p. 14-26.
15. Cerri, S., Mus, L., Blandini, F. *Parkinson's Disease in Women and Men: What's the Difference?* J Parkinsons Dis, 2019. 9(3): p. 501-515.
16. Bourque, M., Di Paolo, T. *Neuroactive steroids and Parkinson's disease*. Current Opinion in Endocrine and Metabolic Research, 2022. 22: p. 100312.
17. Westreich, D. *Epidemiology by design: a causal approach to the health sciences*: Oxford University Press. 2020.
18. Liestøl, K., Andersen, P.K. *Updating of covariates and choice of time origin in survival analysis: problems with vaguely defined disease states*. Stat Med, 2002. 21(23): p. 3701-14.

19. Lee, D.H., Rezende, L.F.M., Ferrari, G., et al. *Physical activity and all-cause and cause-specific mortality: assessing the impact of reverse causation and measurement error in two large prospective cohorts.* Eur J Epidemiol, 2021. 36(3): p. 275-285.
20. Moisan, F., Gourlet, V., Mazurie, J.-L., et al. *Prediction Model of Parkinson's Disease Based on Antiparkinsonian Drug Claims.* American Journal of Epidemiology, 2011. 174(3): p. 354-363.
21. Canonico, M., Artaud, F., Degaey, I., et al. *Incidence of Parkinson's disease in French women from the E3N cohort study over 27 years of follow-up.* Eur. J. Epidemiol., 2022. 37(5): p. 513-523.
22. Mak, M.K., Wong-Yu, I.S., Shen, X., et al. *Long-term effects of exercise and physical therapy in people with Parkinson disease.* Nat Rev Neurol, 2017. 13(11): p. 689-703.
23. Zhen, K., Zhang, S., Tao, X., et al. *A systematic review and meta-analysis on effects of aerobic exercise in people with Parkinson's disease.* NPJ Parkinsons Dis, 2022. 8(1): p. 146.
24. Janssen Daalen, J.M., Schootemeijer, S., Richard, E., et al. *Lifestyle Interventions for the Prevention of Parkinson Disease: A Recipe for Action.* Neurology, 2022. 99(7 Supplement 1): p. 42-51.
25. Berg, D., Crotty, G.F., Keavney, J.L., et al. *Path to Parkinson Disease Prevention: Conclusion and Outlook.* Neurology, 2022. 99(7 Suppl 1): p. 76-83.
26. Fang, X., Han, D., Cheng, Q., et al. *Association of Levels of Physical Activity With Risk of Parkinson Disease: A Systematic Review and Meta-analysis.* JAMA Netw Open, 2018. 1(5): p. e182421.
27. Wagner, M., Dartigues, J.F., Samieri, C., et al. *Modeling Risk-Factor Trajectories When Measurement Tools Change Sequentially During Follow-up in Cohort Studies: Application to Dietary Habits in Prodromal Dementia.* Am J Epidemiol, 2018. 187(4): p. 845-854.
28. Wang, Y.L., Wang, Y.T., Li, J.F., et al. *Body Mass Index and Risk of Parkinson's Disease: A Dose-Response Meta-Analysis of Prospective Studies.* PLoS One, 2015. 10(6): p. e0131778.
29. Noyce, A.J., Kia, D.A., Hemani, G., et al. *Estimating the causal influence of body mass index on risk of Parkinson disease: A Mendelian randomisation study.* PLoS Med, 2017. 14(6): p. e1002314.
30. Heilbron, K., Jensen, M.P., Bandres-Ciga, S., et al. *Unhealthy Behaviours and Risk of Parkinson's Disease: A Mendelian Randomisation Study.* J Parkinsons Dis, 2021. 11(4): p. 1981-1993.
31. Riso, L., Kaaks, R., Kuhn, T., et al. *General and abdominal adiposity and the risk of Parkinson's disease: A prospective cohort study.* Parkinsonism Relat Disord, 2019. 62: p. 98-104.
32. Chen, H., Zhang, S.M., Schwarzschild, M.A., et al. *Obesity and the risk of Parkinson's disease.* Am J Epidemiol, 2004. 159(6): p. 547-55.
33. Palacios, N., Gao, X., McCullough, M.L., et al. *Obesity, diabetes, and risk of Parkinson's disease.* Mov Disord, 2011. 26(12): p. 2253-9.
34. Park, J.H., Choi, Y., Kim, H., et al. *Association between body weight variability and incidence of Parkinson disease: A nationwide, population-based cohort study.* Eur J Neurol, 2021. 28(11): p. 3626-3633.
35. Jeong, S.M., Han, K., Kim, D., et al. *Body mass index, diabetes, and the risk of Parkinson's disease.* Mov Disord, 2019. 35(2): p. 236-244.
36. van der Marck, M.A., Dicke, H.C., Uc, E.Y., et al. *Body mass index in Parkinson's disease: a meta-analysis.* Parkinsonism Relat Disord, 2012. 18(3): p. 263-7.
37. Bigaard, J., Tjønneland, A., Thomsen, B.L., et al. *Waist circumference, BMI, smoking, and mortality in middle-aged men and women.* Obesity research, 2003. 11(7): p. 895-903.
38. Visscher, T.L., Seidell, J.C., Molarius, A., et al. *A comparison of body mass index, waist-hip ratio and waist circumference as predictors of all-cause mortality among the elderly: the Rotterdam study.* International journal of obesity and related metabolic disorders : journal of the International Association for the Study of Obesity, 2001. 25(11): p. 1730-5.
39. Liu, W., Tang, J. *Association between diabetes mellitus and risk of Parkinson's disease: A prisma-compliant meta-analysis.* Brain Behav, 2021. 11(8): p. e02082.
40. Hu, G., Jousilahti, P., Bidel, S., et al. *Type 2 diabetes and the risk of Parkinson's disease.* Diabetes Care, 2007. 30(4): p. 842-7.



**Annex 2: Results from the linear mixed model for the trajectories of physical activity with a retrospective time scale in the nested case-control study (1,196 PD cases, 23,879 controls).**

<b>Characteristics</b>	<b>Beta (95% CI)<sup>a</sup></b>	<b>P-value</b>	<b>P-value<sup>b</sup></b>
Intercept	-0.048 (-0.098, 0.001)	0.06	
Parkinson's disease status	-0.164 (-0.230, -0.097)	<0.001	
Time	-0.372 (-0.383, -0.361)	<0.001	
Time <sup>2</sup>	-0.119 (-0.122, -0.115)	<0.001	<0.001
Parkinson's disease status × Time	-0.074 (-0.122, -0.025)	0.003	
Parkinson's disease status × Time <sup>2</sup>	-0.017 (-0.033, -0.000)	0.05	0.003
Age	-0.046 (-0.048, -0.044)	<0.001	
Age × Time	-0.061 (-0.063, -0.060)	<0.001	
Age × Time <sup>2</sup>	-0.016 (-0.017, -0.016)	<0.001	<0.001

Age was centered at 73 years; time was in decades.

Please see Figure 6 for a graphical representation.

<sup>a</sup> Models were adjusted for baseline place of residence (rural/urban), age at menarche ( $\leq 11/12-13/\geq 14$  years), parity (nulliparous/one child/two children/ $\geq$ three children), and for time-varying smoking (never/ex/current), and menopausal status (no/natural/artificial/unknown type).

<sup>b</sup> P-value for the combined effect of the terms including Time and Time<sup>2</sup>.

**Annex 3: Results from the multinomial GEE logistic model for trajectories of the categories of body mass index with a backward time-scale (1,196 patients, 23,876 controls).**

Characteristics	BMI (kg/m <sup>2</sup> )					
	<18.5 vs [18.5-25[		[25-30[ vs [18.5-25[		≥30 vs [18.5-25[	
	OR (95% CI) <sup>a</sup>	P-value	OR (95% CI) <sup>a</sup>	P-value	OR (95% CI) <sup>a</sup>	P-value
Intercept	0.12 (0.09-0.16)	<0.001	0.43 (0.37-0.49)	<0.001	0.16 (0.12-0.21)	<0.001
PD	1.17 (0.79-1.71)	0.433	1.05 (0.86-1.27)	0.628	0.64 (0.43-0.95)	0.025
Time	1.50 (1.31-1.72)	<0.001	1.13 (1.06-1.20)	<0.001	1.14 (1.02-1.27)	0.017
Time <sup>2</sup>	1.09 (1.04-1.15)	0.001	0.93 (0.91-0.95)	<0.001	0.89 (0.85-0.94)	<0.001
		<0.001 <sup>b</sup>		<0.001 <sup>b</sup>		<0.001 <sup>b</sup>
PD×Time	1.34 (0.66-2.74)	0.415	1.25 (0.93-1.70)	0.145	0.57 (0.31-1.07)	0.079
PD×Time <sup>2</sup>	1.06 (0.79-1.42)	0.694	1.10 (0.97-1.24)	0.141	0.73 (0.57-0.94)	0.014
		0.295 <sup>b</sup>		0.329 <sup>b</sup>		0.019 <sup>b</sup>
Age	1.02 (1.01-1.03)	0.001	1.03 (1.02-1.03)	<0.001	1.02 (1.01-1.03)	<0.001
Age <sup>2</sup>	1.00 (1.00-1.00)	0.228	1.00 (1.00-1.00)	0.022	1.00 (1.00-1.00)	0.011
		0.006 <sup>b</sup>		<0.001 <sup>b</sup>		<0.001 <sup>b</sup>

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

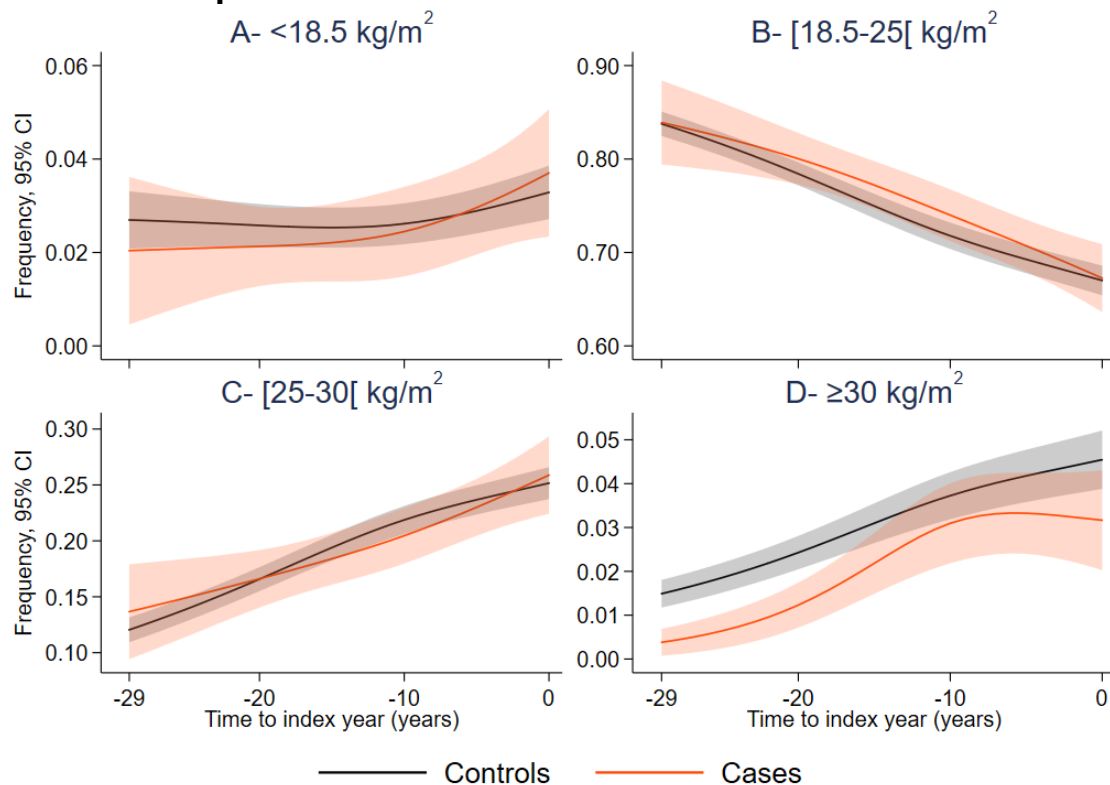
Time is in decades.

The value of the Akaike information criterion (AIC) was of 292,527.9.

<sup>a</sup> Models were adjusted for baseline place of residence (rural/urban), age at menarche ( $\leq 11/12-13/\geq 14$  years), parity (nulliparous/one child/two children/ $\geq$ three children), and for time-varying smoking (never/ex/current), menopausal status (no/natural/artificial/unknown type), and physical activity (quartiles).

<sup>b</sup> P-values for the combined effect of linear and quadratic terms.

**Annex 4: Trajectories of the frequency of the four categories of body mass index in cases and controls up to 29 years before the index date: time modelled using restricted cubic splines.**



The figures show the frequencies (solid lines) and 95% confidence intervals (CI, shaded areas) of the four BMI categories in cases (orange) and controls (blue) according to retrospective time before the index date ( $T_0$ , time=0).

Frequencies were predicted by a multinomial GEE logistic model, using a retrospective time scale, with  $T_0$  representing the age at PD diagnosis in cases and age at the index date in matched controls. The model includes retrospective time modelled with restricted cubic splines (3 knots), case-control status, and the interactions of case-control status with time. The model was adjusted for baseline age at menarche, parity, and place of residence, and time-varying smoking, menopausal status, and physical activity. The model's coefficients are shown in Annex 3.

Trajectories were plotted for the most common profile of E3N participants: 73 years of age at the index date ( $T_0$ ), never smokers, age at menarche at 12-13 years, natural menopause, 2 children, living in urban areas, third quartile (Q3) of physical activity.

The value of the Akaike information criterion (AIC) was of 292,532.7 and slightly higher than the AIC of the model based on a quadratic function time (Figure 11, AIC=292,527.9, Annex 3). Because lower values correspond to a better fit, we retained the model based on quadratic time for our main analysis. However the difference in AIC is not large and both approaches yield very similar trajectories.

**Annex 5: Results from the GEE logistic model for trajectories of diabetes with a backward time-scale (1,200 patients, 25,200 controls).**

Characteristics	Diabetes	
	OR (95% CI) <sup>a</sup>	P-value
Intercept	0.03 (0.02, 0.04)	<0.001
Parkinson' s disease	1.18 (0.79, 1.77)	0.418
Time	1.57 (1.31, 1.89)	<0.001
Time <sup>2</sup>	0.89 (0.81, 0.97)	0.007
		<0.001 <sup>b</sup>
Parkinson' s disease × Time	1.74 (0.83, 3.64)	0.139
Parkinson' s disease × Time <sup>2</sup>	1.28 (0.91, 1.81)	0.152
		0.322 <sup>b</sup>
Age	1.06 (1.05, 1.07)	0.001
Age <sup>2</sup>	1.00 (1.00, 1.00)	0.008
		<0.001 <sup>b</sup>

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

Time is in decades.

The value of the Akaike information criterion (AIC) was of 43,994.42.

<sup>a</sup> Model was adjusted for baseline place of residence (rural/urban), age at menarche ( $\leq 11/12-13/\geq 14$  years), parity (nulliparous/one child/two children/ $\geq$ three children), and for time-varying smoking (never/ex/current), menopausal status (no/natural/artificial/unknown type), body mass index ( $<18.5/[18.5-25/[25-30]/\geq 30$  kg/m<sup>2</sup>), physical activity (quartiles), hypercholesterolemia (Yes/No) and hypertension (Yes/No).

<sup>b</sup> P-values for the combined effect of linear and quadratic terms.





**Titre :** Facteurs de risque de la maladie de Parkinson chez les femmes de la cohorte E3N : rôle de l'activité physique, de l'adiposité et du diabète

**Mots clés :** Epidémiologie, Maladie de Parkinson, activité physique, adiposité, diabète, études de cohorte

**Résumé :** Plusieurs études épidémiologiques ont identifié des associations entre des expositions liées au mode de vie et la maladie de Parkinson (MP). Cependant, compte tenu de la longue phase prodromale de la MP, un biais de causalité inverse pourrait expliquer certaines des associations précédemment rapportées. De plus, les femmes constituent un groupe sous-étudié dans la recherche sur la MP, car la MP est plus fréquente chez les hommes que chez les femmes.

L'objectif de cette thèse était d'examiner la relation entre des comportements de santé potentiellement modifiables (activité physique, adiposité) et le diabète avec l'incidence de la MP chez environ 100 000 femmes participant à l'étude de cohorte française E3N, au cours de 29 ans de suivi, tout en tenant compte du risque de causalité inverse grâce à des méthodes statistiques adaptées.

Nos résultats ont montré que des niveaux plus élevés d'activité physique et d'adiposité étaient associés à une incidence plus faible de la MP chez les femmes, alors qu'aucune association n'a été retrouvée entre le diabète et l'incidence de la MP. Les analyses de trajectoires suggèrent que la causalité inverse n'explique vraisemblablement pas nos résultats.

Ces travaux soulignent l'importance des études de cohortes de grande taille avec un long suivi pour prendre en compte le risque de causalité inverse dans l'étude des maladies neurodégénératives. L'activité physique nous semble être l'intervention non pharmacologique la plus prometteuse en termes de potentiel de prévention de la MP. D'autres études sont nécessaires pour comprendre les mécanismes sous-jacents à l'association inverse entre obésité et MP et pour confirmer l'absence d'association entre le diabète et la MP.

**Title :** Risk factors of Parkinson's disease in women from the E3N cohort: role of physical activity, adiposity, and diabetes

**Keywords:** Epidemiology, Parkinson's disease, physical activity, adiposity, diabetes, cohort studies

**Abstract:** Several epidemiological studies have identified associations between lifestyle exposures and Parkinson's disease (PD). However, given the long prodromal phase of PD, reverse causation bias may explain some of the associations previously reported. In addition, women are an understudied group in PD research since PD is more frequent in men than women.

The aim of this thesis was to examine the relation of potentially modifiable health behaviours (physical activity, adiposity) and diabetes, with PD incidence in about 100,000 women from the French E3N cohort study over 29 years of follow-up, while accounting for the risk of reverse causation using appropriate statistical methods.

We showed that higher levels of physical activity and adiposity were associated with a lower incidence of PD in women, while no association was found between diabetes and PD incidence. Analyses of trajectories suggest that reverse causation is unlikely to explain our results.

This work underscores the importance of large cohort studies with long follow-ups to address the potential for reverse causation. Physical activity seems to be the most promising lifestyle intervention in terms of its potential to prevent PD. Further studies are needed to understand the mechanisms underlying the inverse association between obesity and PD and to confirm the lack of association between diabetes and PD.