



Identifying Protective Drugs for Parkinson's Disease in Health-Care Databases Using Machine Learning

Émeline Courtois, Thi Thu Ha Nguyen, Agnès Fournier, Laure Carcaillon-Bentata, Élodie Moutengou, Sylvie Escolano, Pascale Tubert-Bitter, Alexis Elbaz, Anne C.M. Thiébaut, Ismaïl Ahmed

► To cite this version:

Émeline Courtois, Thi Thu Ha Nguyen, Agnès Fournier, Laure Carcaillon-Bentata, Élodie Moutengou, et al.. Identifying Protective Drugs for Parkinson's Disease in Health-Care Databases Using Machine Learning. *Movement Disorders*, 2023, 37 (12), pp.2376-2385. 10.1002/mds.29205 . hal-03950138

HAL Id: hal-03950138

<https://hal.science/hal-03950138>

Submitted on 8 Mar 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.

RESEARCH ARTICLE

Identifying Protective Drugs for Parkinson's Disease in Health-Care Databases Using Machine Learning

CME

Émeline Courtois, PhD,^{1*} Thi Thu Ha Nguyen, PhD,² Agnès Fournier, PhD,² Laure Carcaillon-Bentata, PhD,³ Élodie Moutengou, MSc,³ Sylvie Escolano, PhD,¹ Pascale Tubert-Bitter, PhD,¹ Alexis Elbaz, PhD,² Anne C.M. Thiébaud, PhD,^{1†} and Ismaïl Ahmed, PhD¹

¹High-Dimensional Biostatistics for Drug Safety and Genomics, Université Paris-Saclay, UVSQ, Inserm, CESP, Villejuif, France

²Exposome, Heredity, Cancer and Health, Université Paris-Saclay, UVSQ, Inserm, CESP, Villejuif, France

³Santé publique France, Saint Maurice, France

ABSTRACT: Background: Available treatments for Parkinson's disease (PD) are only partially or transiently effective. Identifying existing molecules that may present a therapeutic or preventive benefit for PD (drug repositioning) is thus of utmost interest.

Objective: We aimed at detecting potentially protective associations between marketed drugs and PD through a large-scale automated screening strategy.

Methods: We implemented a machine learning (ML) algorithm combining subsampling and lasso logistic regression in a case-control study nested in the French national health data system. Our study population comprised 40,760 incident PD patients identified by a validated algorithm during 2016 to 2018 and 176,395 controls of similar age, sex, and region of residence, all followed since 2006. Drug exposure was defined at the chemical subgroup level, then at the substance level of the Anatomical Therapeutic Chemical (ATC) classification considering the frequency of prescriptions over a 2-year period starting 10 years before the index date to limit

reverse causation bias. Sensitivity analyses were conducted using a more specific definition of PD status.

Results: Six drug subgroups were detected by our algorithm among the 374 screened. Sulfonamide diuretics (ATC-C03CA), in particular furosemide (C03CA01), showed the most robust signal. Other signals included adrenergics in combination with anticholinergics (R03AL) and insulins and analogues (A10AD).

Conclusions: We identified several signals that deserve to be confirmed in large studies with appropriate consideration of the potential for reverse causation. Our results illustrate the value of ML-based signal detection algorithms for identifying drugs inversely associated with PD risk in health-care databases. © 2022 The Authors. *Movement Disorders* published by Wiley Periodicals LLC on behalf of International Parkinson and Movement Disorder Society

Key Words: Parkinson's disease; drug repositioning; machine learning; French national health data system; reverse causation bias

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](#) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

*Correspondence to: Dr. Émeline Courtois, High-Dimensional Biostatistics for Drug Safety and Genomics, Université Paris-Saclay, UVSQ, Inserm, CESP, 16, avenue Paul Vaillant Couturier, 94807 Villejuif Cedex, France; E-mail: emeline.courtois@inserm.fr

†Anne C.M. Thiébaud and Ismaïl Ahmed contributed equally to this work.

Received: 23 May 2022; Revised: 29 July 2022; Accepted: 12 August 2022

Published online 29 August 2022 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/mds.29205

Among neurological disorders, Parkinson's disease (PD) is the fastest growing in terms of prevalence, disability, and deaths.¹ Over the past generation, the global burden of PD has more than doubled as a result of population aging.^{2,3} PD is associated with increased risk of institutionalization and death⁴ and is an important source of health expenses.⁵ Currently available PD treatments are only partially or transiently effective and fail to restore lost dopaminergic neurons and delay disease progression.

In this context, there has been an increasing interest in how existing molecules could be repurposed as an accelerated route for drug discovery.⁶ Drug repurposing

or repositioning is the application of a known drug to new indications and can lead to shorter and inexpensive drug development cycles with increased probability of success. Expensive drug development, coupled with high clinical attrition rates, has fueled the interest of the pharmaceutical industry and academic teams in drug repurposing strategies.⁷ For example, recent studies suggested that salbutamol, a brain-penetrant β_2 -adrenergic agonist, could be associated with reduced PD risk.^{8,9} Other examples of drugs associated with lower PD risk include immunosuppressants¹⁰ or drugs used to treat symptoms of benign prostatic hyperplasia.¹¹

These works targeted a molecule or family of molecules based on prior knowledge or hypotheses to study their potential protective role. Our approach differs in that we propose to identify potential candidate drugs for repurposing in a fully agnostic manner. To our knowledge, only one study sought to do so by emulating randomized control trials, with PD progression as the outcome of interest.¹² Here, we focused on drugs that exhibit beneficial effects on PD incidence with a methodology inspired by automated signal detection in post-marketing pharmacovigilance.

The aim of post-marketing pharmacovigilance is to detect the adverse effects of marketed drugs as early as possible. Pharmacovigilance systems rely on large databases of individual case safety reports of adverse events suspected to be drug induced. Several methods, the most recent of which rely on multiple logistic regression machine learning (ML) algorithms,^{13,14} have been developed to mine this large amount of data and highlight suspicious drug-adverse events. These methods act as hypothesis generators, and signals must be further investigated.

We used an extensive large-scale automated drug screening ML-based strategy in a case-control study nested within the French national health data system (*Système National des Données de Santé* [SNDS]). Our aim was to highlight drugs inversely associated with PD incidence.

Patients and Methods

Data Source

We conducted a case-control study nested in the SNDS, which includes exhaustive individual information on demographic characteristics (age, sex, and place of residence), health-care consumption (drug claims, consultations with general practitioners or specialists, nursing care, and biological procedures), benefits for long-term diseases, and detailed information on hospital stays, for more than 97% of the French population since 2006.¹⁵ Our analyses are restricted to the general scheme that includes persons employed in the private sector and spouses if unemployed (76% of the French

population). As the SNDS was initially developed for the general scheme, data for its affiliates are more exhaustive than for other schemes, especially for historical data.

We used the Anatomical Therapeutic Chemical (ATC) classification for drugs. Diagnoses for hospitalizations were coded according to the *International Classification of Diseases and Related Health Problems*, 10th revision (ICD-10).

PD Patients

We identified incident PD patients in 2016, 2017, and 2018 as follows. We first identified individuals with at least one antiparkinsonian drug claim (ATC N04) between 2013 and 2018, after excluding persons aged below 20 years, those aged below 50 years reimbursed for bromocriptine alone (lactation suppression), and those only on anticholinergics and neuroleptics (drug-induced parkinsonism).

Using an algorithm that was previously validated against a clinical diagnosis of PD by a neurologist,¹⁶ we estimated the probability that a person who is reimbursed antiparkinsonian drugs is treated for PD, based on dose, regularity of use, and demographic variables. The probability cutoff (0.255) of this algorithm with the best combination of sensitivity (92.5%) and specificity (86.4%) according to the Youden index was used to identify PD patients as it is highly predictive of PD status (area under the curve, 0.95). Incident PD patients a given year were persons identified by the algorithm for the first time that year.

We then used several exclusion criteria to refine and increase the specificity of our incident case ascertainment. First, we increased the specificity of the algorithm by excluding PD patients with ICD-10 hospitalization codes for causes of parkinsonism other than PD (eg, supranuclear palsy and multisystem atrophy) after the incidence date. Second, we excluded PD patients with reimbursements of antiparkinsonian drugs, hospitalizations for PD, or benefits for long-term diseases related to PD (ICD-10 G20-G26 and F023) before the incidence date (prevalent PD patients). Third, to exclude PD patients with drug-induced parkinsonism, we excluded individuals with at least one prescription of neuroleptics (ATC N05A, except N05AN [lithium]) before the incidence date. Fourth, we excluded PD patients with a history of dementia before the incidence date, who are unlikely to have idiopathic PD as dementia preceded parkinsonism. These PD patients used antidementia drugs (ATC N06DA02/N06DA03/N06DA04/N06DX01), were hospitalized with dementia (ICD-10 F00/F01/F02/F03/F051/G30/G311), or had benefits for long-term diseases related to dementia before the incidence

date. We did not exclude PD patients who developed dementia after the incidence date.

In a sensitivity analysis, we increased the specificity of PD definition by excluding PD patients with a probability of the algorithm used for PD ascertainment of less than 0.383 (89.1% sensitivity and 89.2% specificity).

Controls

Three controls were randomly matched to each PD patient by incidence density sampling on age at their incidence year (index year = 2016, 2017, and 2018), sex, and region of residence (French *département*). Controls were selected among all subjects present in the SNDS without any antiparkinsonian drug reimbursement, hospitalization for PD, or benefits for long-term diseases related to PD between 2013 and 2018. As we excluded PD patients who used neuroleptics before the incidence date and those with dementia, we also excluded controls who used neuroleptics or controls with dementia before the index date. We also excluded a small number of controls who used antiparkinsonian drugs before 2013.

Drug Exposure Assessment

SNDS data were available from January 1, 2006, to December 31, 2018. Thus, at least a follow-up of 10 years was available before the index date for all PD patients and controls. PD has a long prodromal phase that could induce reverse causation bias.¹⁷ To minimize the risk of reverse causation, we considered a lag of 8 years before the index date, and we assessed drug exposure and covariates over the 2 years before this lag. In the following text, we refer to this 2-year period as the exposure period.

We first considered the penultimate level of the ATC classification, corresponding to chemical, pharmacological, or therapeutic subgroups of drugs. We then refined our analyses by considering drugs according to their active chemical substance (finest level of the ATC classification).

We assessed the number of drug deliveries per individual during the exposure period and, for each drug, generated three binary, embedded variables, following the approach used in the high-dimensional propensity score algorithm¹⁸: the drug is delivered (1) $\geq$ once (“ever” exposure in the remainder of the paper), (2) $\geq$ the median number of deliveries of that drug among exposed controls (“sporadic” exposure), and (3) $\geq$ the 75th percentile of the number of deliveries among exposed controls (“frequent” exposure). A “frequent” user of a given drug is assigned a value of 1 for all three variables. Each binary variable is thereby associated with a specific exposure cutoff; if any two of the cutoff values are equal, only the binary variable associated with the lowest frequency is created. See Section S1 for more details.

Covariates

Sociodemographic variables considered in our analyses were age (5-year groups), sex, and a surrogate for socioeconomic status defined at the *commune* (smallest administrative unit in France) of residence level, the French Deprivation Index (FDep, in quintiles).¹⁹ The higher the score, the greater the social disadvantage. We considered the proportion of land devoted to agriculture (*Surface Agricole Utile* [SAU], in quintiles) in each *commune* as a surrogate for pesticide exposure.²⁰

Additional covariates included benefits for long-term diseases; main/related diagnoses of hospital stays; consultations with general practitioners, psychiatrists and neurologists, dentists, any other specialists, or specialty not specified; use of nursing care; and biological procedures. See Section S2 for an extensive description of the data. Except for long-term diseases with benefits (one binary variable for each), other variables were all coded into three binary variables, as for drugs. All these covariates, including sociodemographic variables, were assessed during the exposure period.

Statistical Methods

We implemented a signal detection strategy combining multiple sample splitting and lasso logistic regression, an ML method used, notably, to perform variable selection in high-dimensional data (see Section S3 for more details and Section S6 for an R script).²¹ As the exclusion of some PD patients and controls led to break the matching, all analyses were adjusted for age, sex, FDep, and SAU. To avoid numerical instability, we removed binary variables, for both drugs and covariates (other than sociodemographic variables), with <100 occurrences.

Our approach involved (1) randomly splitting the data into two samples of equal size D_1 and D_2 ; (2) implementing a lasso logistic regression with fivefold cross-validation in D_1 to identify a subset of potentially relevant variables S associated with PD status among drug exposures and all covariates, while forcing age, sex, FDep, and SAU; (3) fitting an unconditional logistic regression model in D_2 including the variables in S as independent variables. For each binary drug variable present in S , we estimated its regression coefficient β in D_2 . For binary drug variables not in S , their regression coefficients were set to zero. We repeated this procedure from (1) to (3) 500 times and obtained 500 estimated regression coefficients for each binary drug variable. We considered as signals binary variables related to drugs selected more than half (selection percentage, SP = 50%) of the 500 repetitions of steps (1) and (2), and with a negative average regression coefficient, noted $\bar{\beta}$ hereafter, which indicates a protective effect toward the outcome. We also present as Supplementary

material the full list of signals generated with a less-stringent SP threshold of 10%.

Data Sharing

The use of the SNDS data in this research project was approved by *Commission Nationale de l'Informatique et des Libertés*. We are not allowed to share these data due to legal restrictions, but SNDS data are accessible to researchers who meet the criteria for access (request for access is evaluated by *Commission Nationale de l'Informatique et des Libertés*, <https://www.health-data-hub.fr/page/faq-english>).

Results

Figure 1 presents a flowchart for the selection of participants into the study. Our main analysis was based on 40,760 PD patients and 176,395 controls whose

main characteristics at the index date are shown in Supplementary Table C. PD patients and controls were equally distributed within age groups; about 50% were aged 70 to 85 years. The distributions of sex, FDep, and SAU were similar in PD patients and controls, who predominantly lived in urban areas.

Our set of potential covariates contained 1302 binary variables (Supplementary Table D): 15 pertained to long-term diseases, 392 to hospitalization diagnoses, 15 to consultations, 67 to biological procedures, and 20 to sociodemographic characteristics. We screened 374 chemical subgroups coded with 793 binary variables.

Eight signals were generated, involving six drug subgroups (Table 1). The most suggestive signal corresponded to plain sulfonamide diuretics (ATC C03CA), with frequent exposure (≥ 22 deliveries) showing the highest SP among our signals (78.8%) and a stronger association ($\bar{\beta} = -0.14$) than for ever exposure

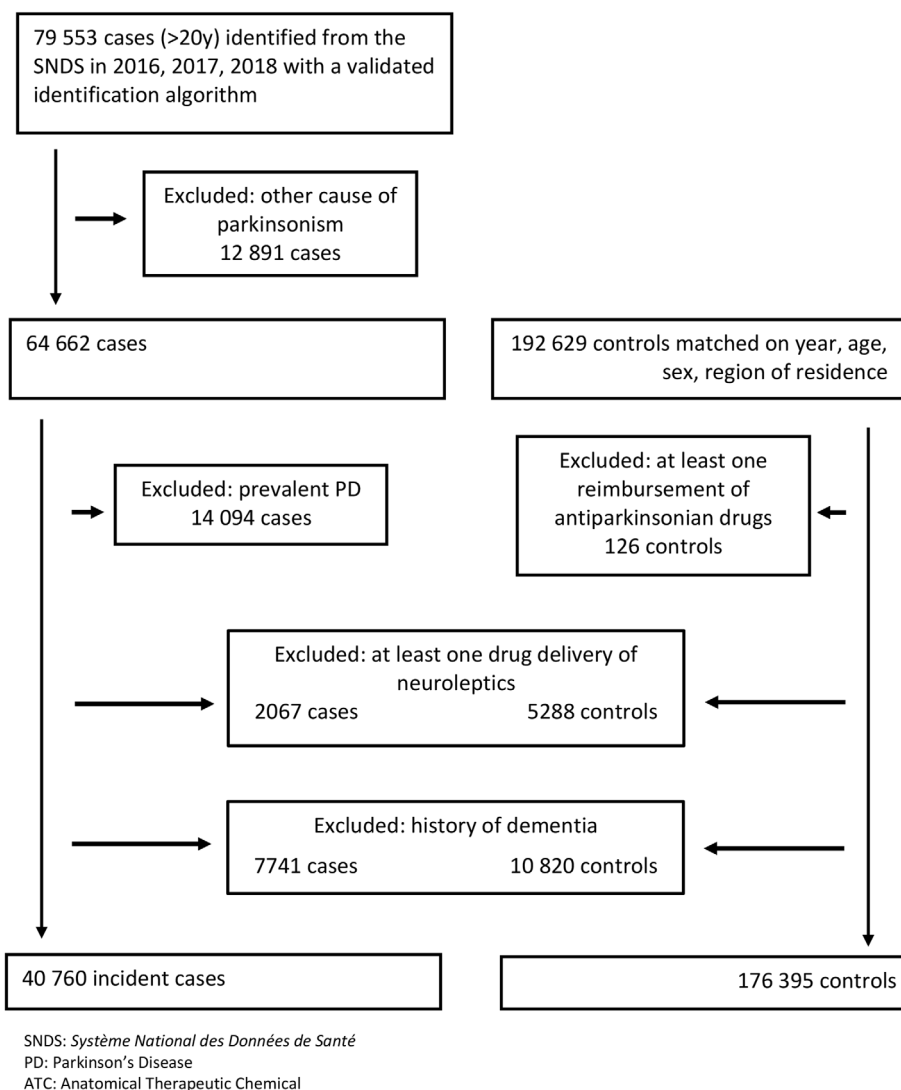


FIG. 1. Flow diagram of the study population.

TABLE 1 Characteristics of the generated signals (boldface) and corresponding binary variables: penultimate level of the Anatomical Therapeutic Chemical classification

Drugs	Drug chemical subgroup label	Exposed PD patients (%)	Exposed controls (%)	Exposure cutoff	Selection percentage	$\bar{\beta}$ ($\overline{\text{OR}}$)
A10AD-ever	Insulins and analogues for injection, intermediate- or long acting combined with fast acting	0.55	0.59	1	38.0	−0.039 (0.962)
A10AD-sporadic		0.24	0.30	15	58.6	−0.175 (0.840)
A10AD-frequent		0.14	0.16	21	7.4	0.020 (1.020)
C03CA-ever	Sulfonamides, plain	4.48	4.51	1	65.8	−0.053 (0.948)
C03CA-sporadic		2.15	2.27	10	22.2	0.004 (1.004)
C03CA-frequent		1.05	1.21	22	78.8	−0.144 (0.866)
D02AC-ever	Soft paraffin and fat products	9.68	9.33	1	65.2	−0.043 (0.958)
D02AC-frequent		2.86	2.70	3	5.4	0.003 (1.003)
M03C-ever	Muscle relaxants, directly acting agents	0.18	0.22	1	51.6	−0.147 (0.864)
M03C-frequent		0.06	0.08	2	9.4	0.006 (1.006)
N07BA-ever	Drugs used in nicotine dependence	0.75	0.84	1	67.0	−0.117 (0.889)
N07BA-frequent		0.20	0.26	2	54.0	−0.122 (0.886)
R03AL-ever	Adrenergics in combination with anticholinergics including triple combinations with corticosteroids	1.19	1.13	1	1.0	0.001 (1.001)
R03AL-sporadic		0.77	0.76	2	9.8	0.006 (1.006)
R03AL-frequent		0.26	0.31	7	60.6	−0.161 (0.851)

Signals are ordered by the alphabetical order of the Anatomical Therapeutic Chemical classification.

Abbreviations: $\bar{\beta}$, average regression coefficient; $\overline{\text{OR}}$, average odds ratio (exponential of the average regression coefficient).

($\bar{\beta} = -0.05$). Another suggestive signal corresponded to drugs used in nicotine dependence (N07BA), with a higher SP for ever (67.0%) than frequent exposure (≥ 2 deliveries, 54.0%) and a similar strength of association ($\bar{\beta} = -0.12$). Additional signals were detected for frequent exposure to adrenergics in combination with anticholinergics (R03AL), sporadic exposure to insulins and analogues for injection (A10AD), ever exposure to paraffin and fat products (D02AC), and direct-acting muscle relaxants (M03C).

Figure 2 shows the distribution of the estimated regression coefficients over 500 repetitions for 8 signals. Overall, the dispersion of these distributions was correlated with the proportion of exposed individuals: the smaller the proportion, the flatter the distribution. For sulfonamide diuretics (C03CA), the distribution of frequent exposure was shifted toward smaller negative values compared to that of ever exposure, suggesting a dose–effect relationship.

In Supplementary Section S4, we provide the list of 112 signals generated by relaxing the SP threshold to 10% instead of 50% (Supplementary Table E).

At the finest level of the ATC classification, we screened 831 substances coded with 1680 binary variables; this analysis included 2189 variables (Supplementary Table D). Six signals were generated

(Table 2), four of which involved substances belonging to the drug subgroups previously highlighted. Among plain sulfonamide diuretics, furosemide (C03CA01) was the most represented substance in our study population; both ever (SP = 58.6%, $\bar{\beta} = -0.05$) and frequent (SP = 56.2%, $\bar{\beta} = -0.08$) exposures to furosemide stood out. Insulin aspart (A10AD05) was the most represented substance among insulin and analogues; only sporadic use emerged as a signal (SP = 60.2%, $\bar{\beta} = -0.23$). Drugs used in nicotine dependence were approximately equally distributed between nicotine and varenicline users; of these two substances, only frequent exposure to varenicline (N07BA03) emerged (SP = 66.4%, $\bar{\beta} = -0.39$). Among adrenergics, a substance belonging to a slightly different drug subgroup from the one previously identified (R03AL) was highlighted with frequent exposure to formoterol and budesonide, a β_2 -agonist in combination with a corticosteroid (R03AK07, SP = 52.2%, $\bar{\beta} = -0.08$). The last signal corresponded to ever exposure to mycophenolic acid (L04AA06, SP = 55.6%, $\bar{\beta} = -0.33$). There were no signals for soft paraffin and fat products (D02AC) or direct-acting muscle relaxants (M03C).

Results of sensitivity analyses based on a more specific definition of PD (29,873 PD patients and 176,395

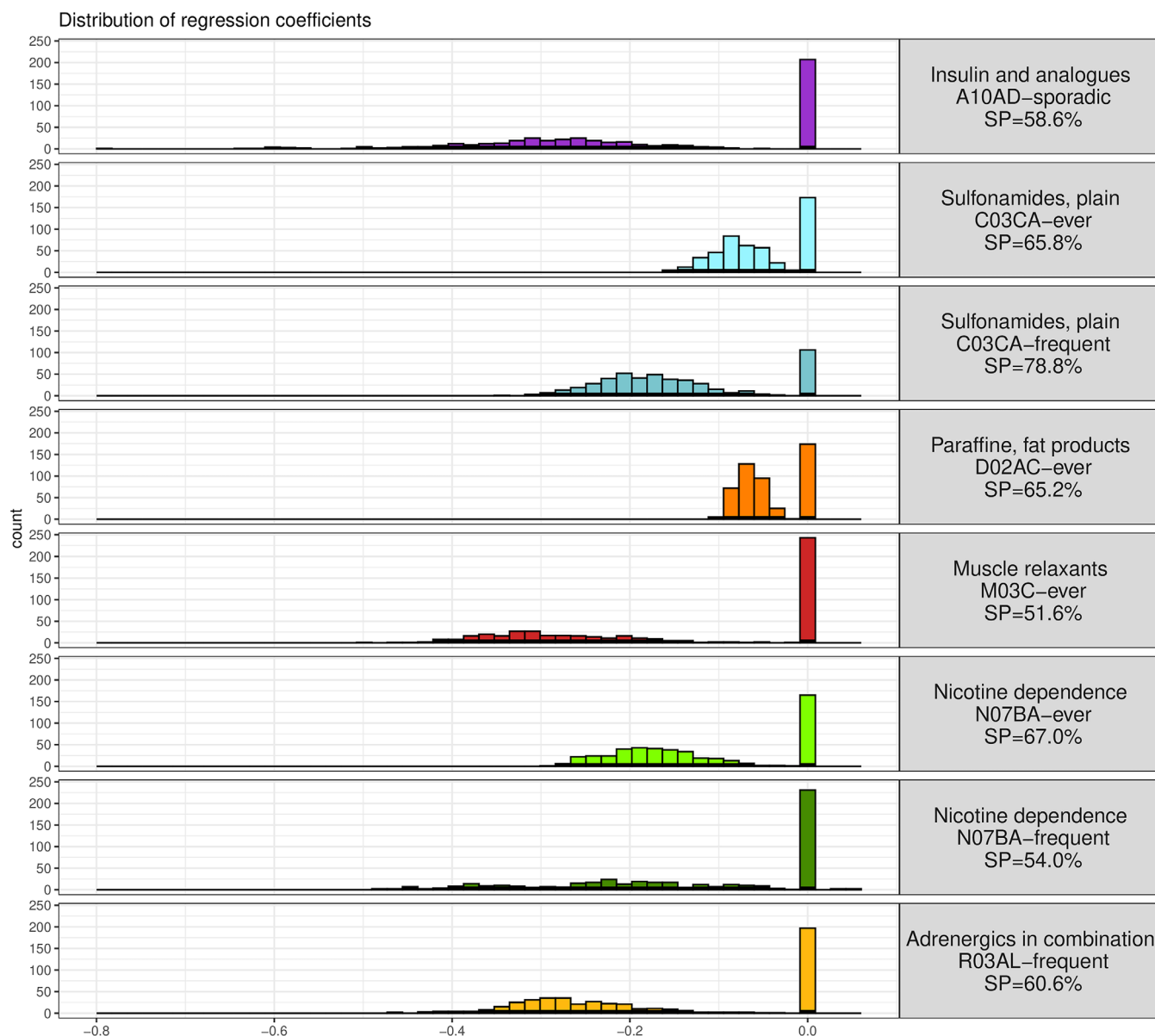


FIG. 2. Distribution of regression coefficients obtained over the 500 repetitions of our signal detection approach for eight generated signals: penultimate level of the Anatomical Therapeutic Chemical classification. The peaks of the distributions correspond to the number of repetitions where binary variables were not selected in the first step of our algorithm. SP: selection percentage. [Color figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1002/mds.25205)] [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1002/mds.25205)]

controls) are shown in Supplementary Section S5 (Tables F–I). Overall, these analyses yielded results generally consistent with those from the main analysis. At the penultimate level of the ATC classification, all signals identified in the main analysis were also generated except for insulins and analogues (Supplementary Table H). The signal for plain sulfonamide diuretics was stronger than in the main analysis (C03CA-ever, SP = 80.4%, $\bar{\beta} = -0.078$; C03CA-frequent, SP = 90.6%, $\bar{\beta} = -0.229$). This analysis also generated signals not identified in the main analysis, namely other emollients and protectives (D02AX-frequent), anti-

inflammatory preparations, nonsteroids for topical use (M02AA-frequent), anticholinergics (R03BB-ever), and mucolytics (R05CB-ever). Analysis conducted at the finest level of the ATC classification (Supplementary Table I) identified furosemide with a high SP (C03CA01-ever, SP = 77.0%, $\bar{\beta} = -0.086$; C03CA01-frequent, SP = 82.6%, $\bar{\beta} = -0.191$). Two signals were generated for the two main drugs used in nicotine dependence, that is, nicotine and varenicline. Signals were also generated for ever exposure to tiotropium bromide (R03BB04-ever) and for ever exposure to direct-acting muscle relaxants (M03C).

TABLE 2 Characteristics of the generated signals (boldface) and corresponding binary variables: finest level of the Anatomical Therapeutic Chemical classification

Drugs	Drug chemical subgroup label	Exposed PD patients (%)	Exposed controls (%)	Exposure cutoff	Selection percentage	$\bar{\beta}$ ($\overline{\text{OR}}$)
A10AD05-ever	Insulin aspart	0.36	0.38	1	14.0	0.002 (1.002)
A10AD05-sporadic		0.15	0.19	12	60.2	−0.230 (0.795)
A10AD05-frequent		0.09	0.10	19	0.6	0.005 (1.005)
C03CA01-ever	Furosemide	4.23	4.24	1	58.6	−0.050 (0.951)
C03CA01-sporadic		2.12	2.18	9	6.6	0.004 (1.004)
C03CA01-frequent		0.97	1.09	22	56.2	−0.077 (0.926)
L04AA06-ever	Mycophenolic acid	0.04	0.08	1	55.6	−0.333 (0.717)
N07BA03-ever	Varenicline	0.38	0.41	1	7.6	0.006 (1.006)
N07BA03-frequent		0.06	0.11	2	66.4	−0.391 (0.676)
R03AK07-ever	Formoterol and budesonide	3.46	3.33	1	6.2	0.001 (1.001)
R03AK07-sporadic		1.85	1.82	2	3.2	0.002 (1.002)
R03AK07-frequent		0.77	0.85	8	52.2	−0.083 (0.920)

Signals are ordered by the alphabetical order of the Anatomical Therapeutic Chemical classification.

Abbreviations: $\bar{\beta}$, average regression coefficient; $\overline{\text{OR}}$, average odds ratio (exponential of the average regression coefficient).

Discussion

This study is part of a research effort aimed at identifying already-developed compounds associated with reduced PD risk. Using an ML-based signal detection approach, we identified six subgroups of drugs in our main analysis: plain sulfonamide diuretics, drugs used in nicotine dependence, adrenergics in combination with anticholinergics including triple combinations with corticosteroids, insulin and analogues for injection, soft paraffin and fat products, and muscle relaxants. Our criteria for characterizing a signal were the proportion of selection over the repetitions of our algorithm (SP) and the average strength of the inverse association with PD. Analyses conducted at the active substance level of drugs confirmed our results by highlighting specific substances among subgroups previously identified: furosemide among plain sulfonamide diuretics, insulin aspart among insulin and analogues, and varenicline among drugs used in nicotine dependence.

According to our criteria, the signal pertaining to plain sulfonamide diuretics was the strongest. Ever and, even more, frequent exposures were associated with decreased PD risk in both main and sensitivity analyses. This pattern suggests a dose–effect relationship that strengthens the plausibility of a protective effect. Furosemide is a sulfonamide-derivative loop diuretic primarily used for the treatment of edema. Furosemide has been recently proposed as a probe molecule for Alzheimer's disease based on anti-inflammatory properties due to the inhibition of the secretion of

proinflammatory TNF- α , IL-6, and nitric oxide; it has also been shown to inhibit A β oligomer formation.^{22–24} Furosemide binds to mitoNEET, a mitochondrial outer-membrane protein that plays an important role in mitochondrial function and metabolism.²⁵ MitoNEET knockout mice show signs of striatal mitochondrial dysfunction and parkinsonian symptoms,²⁶ and mitoNEET inhibition attenuates lipopolysaccharide-induced inflammation and oxidative stress.²⁷ Zonisamide is another sulfonamide and antiepileptic that has also been proposed as a PD treatment based on several antiparkinsonian mechanisms, including blocking of calcium channels, modulation of dopamine metabolism, induction of neurotrophic factors, inhibition of monoamine oxidase-B, oxidative stress, apoptosis, and neuroinflammation.²⁸ In France, zonisamide is very rarely prescribed for epilepsy and has no marketing authorization for PD; the number of participants exposed was too small to allow analyses for this drug. One major difference between zonisamide and furosemide, however, is that whereas zonisamide readily crosses the blood–brain barrier (BBB), furosemide has poor BBB penetration. It is possible that age-related BBB alterations or high doses may facilitate its penetration into the brain.²⁹ Alternatively, whether peripheral changes in body fluid homeostasis due to furosemide have central effects remains to be determined.^{30,31}

Frequent exposure to adrenergics in combination with anticholinergics including triple combinations with corticosteroids was highlighted in both main and sensitivity analyses at the penultimate level of the ATC

classification (R03AL subgroup). It was also highlighted at the substance level (long-acting formoterol) in the main analysis. There has been recent interest in the association between β 2-adrenergic agonists and PD but with inconsistent findings. Discrepancies across studies may stem from differences in duration of follow-up and how important confounders (eg, smoking) or the prodromal phase of PD was considered; in addition all studies did not distinguish long- and short-acting adrenergic drugs.^{8,9,32-36}

In our main analysis, we also identified a signal for insulin use, although only sporadic exposure to insulin and analogues was selected. This signal was not highlighted in sensitivity analyses. Emerging evidence from epidemiological and laboratory studies is in favor of increased PD risk among diabetes mellitus patients.³⁷ However, these studies focused on type 2 diabetes, which is characterized by insulin resistance and is more frequent and occurs at a later age than insulin-dependent diabetes. Nevertheless, biological plausibility for an inverse association between insulin and PD may stem from intrinsic properties of insulin that crosses the BBB and influences a multitude of brain pathways, including neuronal survival and dopaminergic transmission.³⁸ The insulin/IGF-1 signaling pathway contributes to control neuronal excitability, and its dysfunction induces progressive neuronal loss in PD.³⁹

The signal corresponding to ever exposure to soft paraffin and fat products (used as emollients and protectives) is difficult to interpret because there is no straightforward mechanism involving these dermatological products that could influence PD risk. The signal corresponding to muscle relaxants is also unlikely to be plausible. The only substance in this group is dantrolene, which is used in specific situations (eg, after a stroke, paraplegia, cerebral palsy, or multiple sclerosis), therefore suggesting that this signal may be a proxy for a particular medical condition. These two signals were the only ones generated solely for ever exposure, which makes them less credible compared to signals for frequent exposures. Moreover, none of them were highlighted in analyses at substance level.

The main strengths of our study are its large size and long follow-up before the index date, which allowed us to address reverse causation bias by excluding exposures over several years preceding PD diagnosis.¹⁷ The prodromal phase of PD is characterized by the progressive emergence of nonmotor and motor symptoms in the years preceding PD diagnosis that are likely to lead to increased medical contacts and changes in prescriptions. Thus, the inclusion of a lag between drug exposure and disease incidence decreases the risk of biased associations reflecting changes in prescriptions in PD patients that are caused by prodromal symptoms and are unlikely to play a causal role in disease incidence. Moreover, our population-based case-control study

was nested within a nationwide database representative of the French population in which reimbursements of drugs prescribed by physicians are exhaustively recorded (except those used during hospital stays or delivered over the counter). Our approach, inspired from signal detection in pharmacovigilance, acts as a hypothesis generator. Unlike most hypothesis-driven studies, the purely agnostic nature of our approach allows highlighting unexpected associations and thus promotes new discoveries, which are immediately needed to develop disease-modifying treatments. Furthermore, the ML methodology is appropriate considering the large dimension of the data and allows handling multiple drug exposures and confounders. Alternative studies relied on artificial intelligence to identify a set of candidate drugs potentially associated with a reduced risk of Parkinson's disease.^{40,41}

Limitations include the lack of direct adjustment for potential confounders (eg, smoking and physical activity). However, we adjusted our analyses for a wide range of covariates available in the SNDS using an ML algorithm capable of handling numerous covariates that could serve as proxies for potential confounders. Furthermore, although we used a previously validated algorithm to identify PD patients, misclassification of some PD patients cannot be ruled out, but our main signal for plain sulfonamide diuretics was reinforced in sensitivity analyses with a more specific PD definition. The fact that drugs used in nicotine dependence were inversely associated with PD in our study, even though these drugs are not used frequently, is a strong argument in favor of the validity of our PD case ascertainment method and ML-based signal detection approach. The inverse association between smoking and PD is indeed one of the most consistent observations in PD epidemiology,⁴² and recent Mendelian randomization studies support a causal association.⁴³ Because drugs for nicotine dependence were retained in the models, our analyses are indirectly and partially adjusted for smoking.

The search for new PD therapies through drug repositioning has gained attention given the current lack of fully satisfactory therapeutic options. By mining a large-scale case-control study nested within the French SNDS using an ML algorithm developed to account for unmeasured confounding, we screened agnostically a large number of drugs and identified plain sulfonamide diuretics as a drug chemical subgroup potentially inversely associated with PD risk; weaker signals included insulin and β 2-adrenergic agonists. Our findings result in new hypotheses that deserve replication and could lead to developing new therapeutic or preventive strategies in PD. ■

Acknowledgment: This project was funded by The Michael J. Fox Foundation for Parkinson's Research.

Data Availability Statement

The use of the SNDS data in this research project was approved by Commission Nationale de l'Informatique et des Libertés. We are not allowed to share these data due to legal restrictions but SNDS data are accessible to researchers who meet the criteria for access (request for access are evaluated by Commission Nationale de l'Informatique et des Libertés; <https://www.health-data-hub.fr/page/faq-english>). "cd_value_code="text"

References

- GBD 2015 Neurological Disorders Collaborator Group. Global, regional, and national burden of neurological disorders during 1990–2015: a systematic analysis for the global burden of disease study 2015. *Lancet Neurol* 2017;16(11):877–897. [https://doi.org/10.1016/S1474-4422\(17\)30299-5](https://doi.org/10.1016/S1474-4422(17)30299-5)
- Ray Dorsey E, 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. *Lancet Neurol* 2018;17(11):939–953. [https://doi.org/10.1016/S1474-4422\(18\)30295-3](https://doi.org/10.1016/S1474-4422(18)30295-3)
- Wanneveich M, Moisan F, Jacqmin-Gadda H, Elbaz A, Joly P. Projections of prevalence, lifetime risk, and life expectancy of Parkinson's disease (2010–2030) in France. *Mov Disord* 2018;33(9):1449–1455. <https://doi.org/10.1002/mds.27447>
- Macleod AD, Taylor KSM, Counsell CE. Mortality in Parkinson's disease: a systematic review and meta-analysis. *Mov Disord* 2014;29(13):1615–1622. <https://doi.org/10.1002/mds.25898>
- Kowal SL, Dall TM, Chakrabarti R, Storm MV, Jain A. The current and projected economic burden of Parkinson's disease in the United States. *Mov Disord* 2013;28(3):311–318. <https://doi.org/10.1002/mds.25292>
- Kakkar AK, Singh H, Medhi B. Old wines in new bottles: repurposing opportunities for Parkinson's disease. *Eur J Pharmacol* 2018;830(April):115–127. <https://doi.org/10.1016/j.ejphar.2018.04.023>
- Ashburn TT, Thor KB. Drug repositioning: identifying and developing new uses for existing drugs. *Nat Rev Drug Discovery* 2004;3(8):673–683. <https://doi.org/10.1038/nrd1468>
- Mittal S, Bjørnevik K, Im DS, et al. β 2-Adrenoreceptor is a regulator of the α -synuclein gene driving risk of Parkinson's disease. *Science* 2017;357(6354):891–898. <https://doi.org/10.1126/science.aaf3934>
- Gronich N, Abernethy DR, Auriel E, Lavi I, Rennett G, Saliba W. β 2-adrenoreceptor agonists and antagonists and risk of Parkinson's disease. *Mov Disord* 2018;33(9):1465–1471. <https://doi.org/10.1002/mds.108>
- Racette BA, Gross A, Vouri SM, Camacho-Soto A, Willis AW, Searles NS. Immunosuppressants and risk of Parkinson disease. *Ann Clin Transl Neurol* 2018;5(7):870–875. <https://doi.org/10.1002/acn3.580>
- Gros P, Wang X, Guan J, et al. Exposure to phosphoglycerate kinase 1 activators and incidence of Parkinson's disease. *Mov Disord* 2021;36(10):2419–2425. <https://doi.org/10.1002/mds.28712>
- Laifenfeld D, Yanover C, Ozery-Flato M, et al. Emulated clinical trials from longitudinal real-world data efficiently identify candidates for neurological disease modification: examples from Parkinson's disease. *Front Pharmacol* 2021;12:1–11. <https://doi.org/10.3389/fphar.2021.631584>
- Ahmed I, Pariente A, Tubert-Bitter P. Class-imbalanced subsampling lasso algorithm for discovering adverse drug reactions. *Stat Methods Med Res* 2018;27(3):785–797. <https://doi.org/10.1177/0962280216643116>
- Caster O, Juhlin K, Watson S, Norén GN. Improved statistical signal detection in pharmacovigilance by combining multiple strength-of-evidence aspects in vigiRank: retrospective evaluation against emerging safety signals. *Drug Saf* 2014;37(8):617–628. <https://doi.org/10.1007/s40264-014-0204-5>
- Tuppin P, Rudant J, Constantinou P, et al. Value of a national administrative database to guide public decisions: From the système national d'information interrégimes de l'Assurance Maladie (SNIIRAM) to the système national des données de santé (SNDS) in France. *Rev Epidemiol Sante Publique* 2017;65:S149–S167. <https://doi.org/10.1016/j.respe.2017.05.004>
- Moisan F, Gourlet V, Mazurie JL, et al. Prediction model of parkinson's disease based on antiparkinsonian drug claims. *Am J Epidemiol* 2011;174(3):354–363. <https://doi.org/10.1093/aje/kwr081>
- Elbaz A. Prodromal symptoms of Parkinson's disease: implications for epidemiological studies of disease etiology. *Rev Neurol* 2016;172(8–9):503–511. <https://doi.org/10.1016/j.neurol.2016.07.001>
- Schneeweiss S, Rassen JA, Glynn RJ, Avorn J, Mogun H, Brookhart MA. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology* 2009;20(4):512–522. <https://doi.org/10.1097/EDE.0b013e3181a663cc>
- Rey G, Jouglé E, Fouillet A, Hémon D. Ecological association between a deprivation index and mortality in France over the period 1997–2001: variations with spatial scale, degree of urbanicity, age, gender and cause of death. *BMC Public Health*

34. Cheng CM, Wu YH, Tsai SJ, et al. Risk of developing Parkinson's disease among patients with asthma: a nationwide longitudinal study. *Allergy* 2015;70(12):1605–1612. <https://doi.org/10.1111/all.12758>
35. de Gernay S, Conte C, Rascol O, Montastruc JL, Lapeyre-Mestre M. β -adrenoceptor drugs and Parkinson's disease: a nationwide nested case-control study. *CNS Drugs* 2020;34(7):763–772. <https://doi.org/10.1007/s40263-020-00736-2>
36. Marras C, Pequeno P, Austin PC, et al. Beta agonists and progression of Parkinson's disease in older adults: a retrospective cohort study. *Mov Disord* 2020;35(7):1275–1277. <https://doi.org/10.1002/mds.28085>
37. Cheong JLY, de Pablo-Fernandez E, Foltynie T, Noyce AJ. The association between type 2 diabetes mellitus and Parkinson's disease. *J Parkinsons Dis* 2020;10(3):775–789. <https://doi.org/10.3233/jpd-191900>
38. Athauda D, Foltynie T. Insulin resistance and Parkinson's disease: a new target for disease modification? *Prog Neurobiol* 2016;145:98–120.
39. Bassil F, Fernagut PO, Bezard E, Meissner WG. Insulin, IGF-1 and GLP-1 signaling in neurodegenerative disorders: targets for disease modification? *Prog Neurobiol* 2014;118:1–18. <https://doi.org/10.1016/j.pneurobio.2014.02.005>
40. Maclagan LC, Visanji NP, Cheng Y, et al. Identifying drugs with disease-modifying potential in Parkinson's disease using artificial intelligence and pharmacoepidemiology. *Pharmacoepidemiol Drug Saf* 2020;29(8):864–872. <https://doi.org/10.1002/pds.5015>
41. Visanji NP, Madan P, Lacoste AMB, et al. Using artificial intelligence to identify anti-hypertensives as possible disease modifying agents in Parkinson's disease. *Pharmacoepidemiol Drug Saf* 2021;30(2):201–209. <https://doi.org/10.1002/pds.5176>
42. Noyce AJ, Bestwick JP, Silveira-Moriyama L, et al. Meta-analysis of early nonmotor features and risk factors for Parkinson disease. *Ann Neurol* 2012;72(6):893–901. <https://doi.org/10.1002/ana.23687>
43. Domenighetti C, Sugier P-E, Sreelatha AAK, 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. <https://doi.org/10.3233/jpd-212851>

Supporting Data

Additional Supporting Information may be found in the online version of this article at the publisher's web-site.