Parkinson's disease case ascertainment in the EPIC cohort: the NeuroEPIC4PD study

Valentina Gallo, MD, LSHTM-MSc, PhD[1,2], Carol Brayne, MD, FRCP, FFPH[3], Lars Forsgren, MD, PhD[4], Roger A Barker, MD, PhD[5], Jesper Petersson, MD[6], Oskar Hansson, MD, PhD[6], Daniel Lindqvist, MD[6], Claudio Ruffmann, MD[7,8], Lianna Ishihara, PhD[9], Robert Luben, PhD[10], Larraitz Arriola, MD[11,12,15], Alberto Bergareche, MD[12,13], Diana Gavrila, MD[14,15], Maria Elena Erro, MD[16], Nicola Vanacore, MD, PhD[17], Carlotta Sacerdote, MD, PhD[18,19], Bas Bueno-de-Mesquita, MD, MPH, PhD[1,20,21,22], Roel Vermeulen, PhD[23,24], Meinie Seelen, MD[25], Sabina Sieri, PhD[26], Giovanna Masala, MD[27], Silvia Ramat, MD, PhD[28], Andreas Kyrozis, MD[29,30], Antonia Trichopoulou, PhD[29], Salvatore Panico, MD[31], Amalia Mattiello, MD[31], Rudolf Kaaks, PhD[32], Birgit Teucher, PhD[32], Verena Katzke, PhD[32], Manja Kloss, MD[32], Lisa Curry[8], Federico Calboli, PhD[8], Elio Riboli, MD, PhD[1], Paolo Vineis, MD, MPH, FFPH[1], Lefkos Middleton, MD, FRCP[8]

1. School of Public Health, Imperial College London, London, UK
2. Centre of Primary Care and Public Health, Queen Mary University of London, London, UK
3. University of Cambridge
4. Department of Pharmacology and Clinical Neuroscience, Umeå University, Umeå, Sweden
5. Department of Neurology, University of Cambridge, Cambridge, UK
6. Department of Clinical Sciences, Lund University, Sweden
7. Parkinson Institute, Istituti Clinici di Perfezionamento, Milan, Italy
8. Neuroepidemiology and Ageing Unit, School of Public Health, Imperial College London, London, UK
9. Department of World Wide Epidemiology, GlaxoSmithKline, Stockley Park, Uxbridge, UK
10. Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
11. Public Health Department of Gipuzkoa. Basque Government. Spain
12. Biodonostia Research Institute, Neurosciences area, Hospital Universitario Donostia, Donostia, Spain
13. Centro de investigación biomédica en Red para enfermedades Neurodegenerativas (CIBERNED), Carlos III Health Institute, Madrid, Spain
14. Department of Epidemiology, Murcia Regional Health Council, Murcia, Spain
15. CIBER Epidemiología y Salud Pública (CIBERESP), Spain
16. Neurology Department, Complejo Hospitalario de Navarra, Pamplona, Spain.
17. Italian National Institute of Health, Rome, Italy
18. Centre for Cancer Prevention (CPO-Piemonte), Turin, Italy
19. Human Genetic Foundation (HuGeF), Turin, Italy
20. National Institute for Public health and the Environment, Bilthoven, The Netherlands
21. Department of Gastroenterology and Hepatology, University Medical Centre, Utrecht, The Netherlands
22. Dt. of Social & Preventive Medicine, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia.
23. Julius Centre for Public Health Sciences and Primary Care, University Medical Centre, Utrecht, the Netherlands
24. Institute for Risk Assessment Science, Division of Epidemiology, Utrecht University, Utrecht, The Netherlands
25. Department of Neurology, Rudolf Magnus Institute of Neuroscience, University Medical Centre Utrecht, Utrecht, The Netherlands
26. Epidemiology and Prevention Unit, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy
27. Molecular and Nutritional Epidemiology Unit, Cancer Research and Prevention Institute ISPO, Florence, Italy
28. Neuroscience Department, Careggi Hospital, Florence, Italy
29. Hellenic Health Foundation, Athens, Greece
30. First department of Neurology, University of Athens, Athens, Greece
31. Dipartimento di Medicina Clinica e Chirurgia, Federico II University, Naples, Italy
32. German Cancer Research Centre (DKFZ), Division of Cancer Epidemiology, Heidelberg, Germany
33. Department of Neurology, University Heidelberg, Heidelberg, Germany.

Please address all correspondence to:
Dr Valentina Gallo
Imperial College London
School of Public Health
St Mary's Campus
Norfolk place
NW1 2PG London (UK)
Tel +44/(0)20/7594 3463
Email: v.gallo@imperial.ac.uk or valentina.gallo@gmail.com

Word count: 2,962

Running title: NeuroEPIC4PD

Keywords: Parkinson's disease, cohort study, case ascertainment, record linkage, validation, EPIC, incidence.

Abstract

Background/Aims: Large epidemiological prospective studies represent an important opportunity for investigating risk factors for rare diseases such as Parkinson's disease (PD). Here we describe the procedures we used for ascertaining PD cases in the European Prospective Investigation into Cancer and Nutrition (EPIC).

Methods: A three-phase procedure was used: elaboration of a NeuroEPIC4PD template for clinical data collection, identification all potential PD cases via record linkage, and validation of the diagnosis through clinical record revision, in a population of 220,494 subjects recruited in 7 European countries. All cases were labelled with a NeuroEPIC4PD diagnoses of "definite", "very likely", "probable", or "possible" PD.

Results: A total of 881 PD cases were identified, over 2,741,780 person/years follow-up (199 definite, 275 very likely, 146 probable, and 261 possible). Of these, 734 were incident cases. Their mean age at diagnosis was 67.9 years (SD 9.2), and 458 (52.0%) were men. Bradykinesia was the most frequent presenting motor sign (76.5%). Tremor-dominant and akinetic rigid forms of PD were the most common types of PD. A total of 289 (32.8%) were dead at the time of last follow-up.

Conclusions: this exercise proved that it is feasible to ascertain PD in large population-based cohort studies, and offer a potential framework to be replicated in similar studies.

Keywords: Parkinson's disease, cohort study, case ascertainment, record linkage, validation, EPIC, incidence.

Introduction

Parkinson's disease (PD) is the second most common neurodegenerative disorder after Alzheimer's disease. PD is characterised by motor dysfunction, such as bradykinesia, resting tremor, rigidity and postural instability, but also affects the autonomic nervous system, cognition, and a range of other non-motor functions (1). The incidence of the disease rises with age, peaking at 119.5 per 100,000 person years in the seventieth decade, with a lifetime risk of developing the disease of 1.5-3% (2-5). The median age of onset is in the late 60s (4-6) and the mean duration of the disease from diagnosis to death is 15 years (3). Because of ageing of western populations, an increase in incidence can be anticipated.

Large established prospective population studies offer an important opportunity for investigating the role of risk factors for rarer diseases such as PD, with relatively small additional effort for ascertaining cases. In particular, data coming from such studies can shed light on potential mechanisms of action of factors known to be negatively associated with PD (i.e. cigarette smoking) using modern advanced epidemiological techniques (such as Mendelian randomisation). To date PD has been investigated in relatively small cohorts (7-17) (with the numbers of PD cases ranging from 41 (15) to 656 (16)) or using self-reported diagnosis only (17). Cohort studies with large sample sizes and well characterised and validated case definitions have a high degree of power for detecting weak associations with environmental factors.

The aim of this study was to describe the procedures used to ascertain PD cases in the EPIC cohort in order to enable investigators to prospectively assess the association between pre-diagnostic risk factors and the incidence of the disease: the NeuroEPIC4PD study. The methods used in this study are expected to be generalizable to other cohorts.

Materials and Methods

The EPIC study

The European Prospective Investigation into Cancer and Nutrition (EPIC) study is a large, well established, multicentre cohort study on-going in ten Western European countries(18). At baseline, information on lifestyle (education, occupation, lifetime tobacco and alcohol consumption, physical activity, oral contraceptive and hormone replacement therapy) menstrual and reproductive history, and past medical history was collected through a questionnaire. Anthropometric assessments and blood pressure were measured in a standardised way (18). Dietary intake was assessed through self-administered quantitative dietary questionnaires and a 24-hour dietary recall questionnaire (19). EPIC was originally designed to investigate the role of nutrition in cancer aetiology, but it has also demonstrated its value for investigating other medical conditions such as cardiovascular diseases (20), and diabetes (21).

The NeuroEPIC4PD population

The NeuroEPIC4PD study is based on an source population of 220,494 subjects recruited in Sweden (Umea and Malmo), UK (Cambridge), Netherlands (Utrecht), Germany (Heidelberg), Spain (Navarra, San Sebastian, and Murcia), Italy (Turin, Varese, Florence and Naples), and Greece from the general population residing in defined geographical areas between 1992 and 2002 within the EPIC study (18). Exception was the Utrecht-Prospect cohort which was based on breast cancer screening participants (18). The Naples and Utrecht-Prospect cohorts were restricted to women, whereas all other cohorts included both sexes. In EPIC, follow-up for mortality and specific causes of death is carried out actively or through linkage with mortality registries at regional and national levels (18). To date, follow-up is 98.5% complete.

Expected PD cases

Numbers of expected PD cases were calculated in order to compare completeness of procedures used in different centres, and to estimate the workload of phase III in each centre. They were estimated applying sex- and age-specific incidence rates coming from weighted averages of population-based studies investigating PD incidence (4, 22-27). Sex- and age-specific cumulative incidence rates were calculated over the sex- and centre-specific follow-up periods in each EPIC sub-cohort resulting in estimates of expected cases during the current follow-up.

Case ascertainment methods

Case ascertainment was organised in three phases: in Phase I the NeuroEPIC4PD template for clinical data collection - including the NeuroEPIC4PD label to be attached to each diagnosis - was elaborated by a group of experienced neurologists and epidemiologists. During Phase II, each centre identified *potential cases* through record linkage with one or more local sources of information. Clinical records of *potential cases* were then reviewed by experts in movement disorders and given a final diagnosis with a NeuroEPIC4PD label, during Phase III (Online Figure 1). The three phases will be described more in detail below.

Phase I: NeuroEPIC4PD template for clinical data collection

The first NeuroEPIC meeting took place on November 29th 2010. During the meeting the strategy for PD case ascertainment was discussed in detail by a group of neurologists and epidemiologists. The final outcome was the elaboration of the template for clinical data collection (Appendix 1) and the definition of the NeuroEPIC4PD label to be given to every case ascertained in EPIC. The template included sections on general information, diagnostic criteria according to the UK Brain Bank(28), additional clinical data, medication, surgical treatment, autopsy, source of information and quality of data, and final diagnosis. Possible final diagnoses included: PD, Multiple System Atrophy (MSA), progressive supranuclear palsy (PSP), vascular parkinsonism (VD), dementia with Lewy bodies (LBD), essential tremor (ET), Parkinson's disease with essential tremor (PDET), unclassifiable parkinsonism (UP), and other diagnosis.

Each diagnosis was labelled with a NeuroEPIC4PD label, which was based on a matrix combining two variables: the amount and quality of data available, and the degree of confidence of the neurologist expert in movement disorders reviewing the evidence (Online Figure 2, Appendix 1). The amount and quality of data available could be rated as "poor", "good" or "excellent", where "excellent" data was defined as a complete set of clinical data able to give a clear picture of the case (including detailed neurological examinations), with scattered non-essential missing information; "good" data was defined as a set of data giving a fairly complete idea of the case with scattered essential information missing; "poor" data was defined as an incomplete set of data, with much essential information missing. The degree of confidence of the neurologist could be rated as "high", "medium", or "low" on the basis of his/her final judgment of the clinical history of the single case. It was stated very clearly that this judgement was independent of the amount of information available (a neurologist could have a high degree of confidence despite very poor information, or a low degree of confidence despite very detailed information). Diagnoses were defined as "definite" only when the degree of confidence of the neurologist was high and data quality excellent; diagnoses were defined as "very likely" when the degree of confidence of the neurologist was high, but data quality was either good or poor; and defined as "probable" when the degree of confidence of the neurologist was medium and data quality was either excellent or good; finally, diagnoses were defined as "possible" in all remaining cases.

Phase II: *Potential PD cases* identification

Potential PD cases were identified by centre-specific strategies, in order to optimise local sources of data. The general principle was to increase sensitivity as much as possible in order to minimise the number of false negatives given that the subsequent clinical record review would maximise specificity. In addition, patterns of possible referrals pathways for PD cases in each centre were collected in order to evaluate the adequacy of the sources of information used. Specific sources of information by centre used in Phase II are shown in Online Figure 1; details of the procedures can be found in Appendix 2.

Phase III – Case validation and collection of additional clinical data

For all *potential cases* identified in Phase II, clinical records were searched for and reviewed by a neurologist expert in movement disorders. For each subject for whom at least some information was available, a clinical data form was filled in (Appendix 1). Each *potential case* was then attached a final diagnosis (PD, MSA, PSP, VD, LBD, ET, PDET, UP, and other diagnosis) and labelled with a NeuroEPIC4PD labels (Online Figure 2).

Results

The source population used for the NeuroEPIC4PD study includes 220,494 subjects recruited in 13 EPIC centres across 7 European countries (Online Table 1). Mean age of the population at recruitment was 53.1 years (SD 10.0). The recruitment framework oversampled women by design, resulting in a total of 83,320 (37.8%) men recruited, exceptions being Utrecht and Naples which recruited women only. The entire population was followed up for a mean of 12.4 years (SD 2.8) generating a total 2,741,780 person years. Of the entire cohort, 19,473 subjects (8.8%) died according to the updated vital status by the last follow-up visit.

Applying the sex- and age-specific incidence rates of PD derived from population based studies, a total of 309 PD cases were expected over the current follow-up period. These were distributed according to figures shown in Table 1.

Phase II: Potential PD cases identification

A total of 1,723 *potential PD cases* were identified across all centres (Online Figure 3 and Table 1). Overall *potential cases* were about 5.6 times more than expected cases. In Utrecht and Varese *potential cases* were very close to expected. In Murcia and Turin *potential cases* were more than ten times higher than expected; while in the remaining centres they were between about 250 and 650% with the exception of Cambridge where they were 800%. Differences are likely due to differences in sources of information used by centres.

Phase III – Case validation and collection of additional clinical data

Additional clinical information was collected using the NeuroEPIC4PD template in 1,336 out of 1,723 potential cases (77.5%) ; few centres (Umea, Navarra, Murcia, Varese, Naples, and Greece) were able to verify all potential cases; others (Malmo, San Sebastian, and Florence) could verify the great majority of potential cases (87.7 to 98.0%); and other centres (Utrecht, Cambridge, Heidelberg, and Turin) only a smaller proportion of potential cases (47.5 to 66.9%) (Table 1, Figure 1). The variation in these proportions is mainly due to ability of retrieving clinical records of *potential cases*; and to the extent to which the criteria used in phase II had a high sensitivity (including a higher proportion of false positive who would not have a clinical record for PD artificially inflating the number of *potential cases*).

On the basis of the clinical information available, 881 PD cases (65.9%), 230 parkinsonian-related disorders (17.2%), and 225 unrelated conditions (16.8%) were identified (Table 1). Cases who received a diagnosis after the date of recruitment were defined as *incident cases* (N=734). Out of the 881 PD cases, 199 were labelled as definite PD cases (22.6%), 275 as very likely (31.2%), 146 as probable (29.6%), and 261 as possible (29.6%). Among the parkinsonian-related disorders, 26 MSA, 22 PSP, 34 VP, 34 LBD, 30 ET, and 9 PDET were identified. Additionally, 75 cases (5.6% of the total) were defined as having UP (Online Table 2).

PD cases are more than double of those originally expected. All centres have ascertained more than expected cases a part from Utrecht and Varese where about half of expected cases has been ascertained. For all other centres, proportion range from 200% in Florence to 757% in Murcia (Table 1, Figure 1). In each country, the shape of the age- and sex-specific cumulative incidence peaks at 70-79 years with exception of Spain and Germany where it peaks at 60-69 years in both men and women (Figure 2-3). A part from Netherlands, differences between observed and expected curves are mainly due to age-truncation of the cohorts, in fact in all countries, the incidence rates at older ages are much lower than expected, but this may simply reflect the age-composition of the EPIC cohort (Online Table 1).

General and clinical characteristics of PD patients are described in Table 2. Despite women outnumbering men in this study, there is slight predominance of men among PD patients, except for the definite PD category. Individuals with PD were recruited when they were on average 61.9 years of age (SD 8.2) they noticed their first motor symptoms at a mean age of 66.8 years (SD 8.3), and were diagnosed on average one year later, i.e. at 67.9 years (SD 9.2). Relaxing criteria of certainty of the diagnosis, the gap between first motor symptom and diagnosis shortens. Bradykinesia is the most frequent cardinal sign (present in 76.5% of all PD patients). The prevalence of bradykinesia, resting tremor, and rigidity increases the certainty of the diagnosis; this is not observable for postural instability which remains roughly constant across categories, and slightly more frequent among probable cases (Table 2). Tremor-dominant and akinetic rigid forms are the most common types of PD across all NeuroEPIC4PD labels and their relative difference decreases, decreasing the certainty of the diagnosis (Figure 2). Levodopa and dopamine agonists are the most common drugs taken by these PD patients, and all drugs are proportionally taken more frequently by patients with a more certain diagnosis. A total of 16 PD patients had a surgical treatment with deep brain stimulation. A total of 289 (32.8% of entire sample) were dead at the time of last follow-up within the NeuroEPIC4PD study (Table 2).

Discussion

This exercise proved that it is feasible to ascertain PD cases in large population-based multicentre studies. Maximising the local expertise and sources of information, and through involving epidemiologists and neurologists in the project, we were able to ascertain a considerable number of cases, labelled with a different degree of certainty, which will allow us to exploit the invaluable resource of a prospective study in investigating risk factors for this relatively rare neurodegenerative disorder.

Using a three-phase approach, we were able to optimise our search by maximising sensitivity in phase II (thus minimising the false negative rate), and specificity in phase III (therefore minimising the false positive rate). The only limit to this approach remains the proportion of potential cases which could not be ascertained, thus introducing potential false negatives in our sample. However, in the present case this proportion was low and only seen in few centres for practical reasons (Utrecht and Heidelberg); in Cambridge and Turin this is more likely due to having relaxed criteria for identifying potential cases in phase II; in fact in both centres, the number of ascertained cases is well above the number of expected cases (Figure 1). Also, it is important to note that in the context of studying a rare disease in a large cohort study it is of greater relevance to minimise false positive diagnoses which may bias results towards the null, than to minimise false negatives which will be diluted in a very large set of non-cases making their impact negligible.

The finding of a higher number of cases than expected may be due to two reasons: firstly, expected cases were calculated on the follow-up period for which data had been centralised in EPIC, while in each centre, case ascertainment was conducted on a longer period of follow-up, counting the follow-up time which was available locally. Secondly, we expect more cases according to the healthy cohort effect given that cigarette smoking is inversely associated with PD(29): if less smokers than in the general population enrol in epidemiological studies (as occurred in EPIC), this would lead to a less than expected number of smoking-related diseases, but to a higher number of PD cases.

Finally, the large differences between potential and ascertained cases reflect the different methods used in phase I. When more specific methods were used (i.e. Umea, Heidelberg, Greece), it was more likely that a potential case was a true case. Overall, the plots of age-specific incident rates by country confirm that the procedures used were appropriate for the population observed, and that the quality of our case ascertainment procedures was quite high. Moreover, the NeuroEPIC4PD label is able to effectively discriminate between a high and low certainty of diagnosis: the NeuroEPIC4PD labelling was studied in order to optimise the trade-off between power and specificity. By including possible cases, the power for detecting associations with potential risk/protective factors will be maximised; the presence of the label, however, will allow sensitivity analyses on those cases with a higher degree of certainty only.

Clinical characteristics of PD cases using the NeuroEPIC4PD label, reflect the fact that the label is also calculated on the basis of the amount and quality of data available. The crude mortality among PD cases is higher than in the rest of the population, a part from the definite PD. Age at diagnosis seems to be inversely correlated with degree of certainty of diagnosis. This may reflect reduced specificity of symptoms such as bradykinesia, rigidity and postural instability, which are more common (less sensitive) and multifactorial in origin (arthritis, cerebrovascular diseases, etc.), especially in the aged.

The results of this exercise are expected to inform case ascertainment of PD in other cohort studies. As can easily be deduced by the complexity and heterogeneity of the procedures used, maximising the local resources and sources of information is of paramount importance in this exercise. The epidemiological effort is to harmonise data coming from such different sources and build a method which allows for the trade-off between power and specificity that studies of this nature encounter.

This type of study should coordinate closely with other population-based PD cohort studies, such as the CamPaIGN study (30). These can potentially produce complementary information for investigating the complex interactions of risk factors underlying PD. While cohort studies can shed light on the role of potential risk factors (and molecular markers) for developing PD, the patient cohorts can investigate the role of the same risk factors (and the same biological markers) on PD survival, complementing the causal inference process.

Acknowledgments

Author Contributions: Dr Gallo had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: V Gallo, C Brayne, L Forsgren, RA Barker, L Middleton, E Riboli, P Vineis

Analysis and interpretation of data: V Gallo

Drafting of the manuscript: V Gallo, RA Barker, B Bueno-de-Mesquita, P Vineis

Data collection: L Forsgren, J Petersson, O Hansson, D Lindqvist, C Ruffmann, L Ishihara, R Luben, L Arriola, A Bergareche, D Gavrila, ME Erro, N Vanacore, C Sacerdote, R Vermeulen, M Seleen, S Sieri, G Masala, S Ramat, A Kyrozis, A Thricopolou, S Panico, A Mattiello, R Kaaks, B Teucher, V Grote, M Kloss

Critical revision of the manuscript for important intellectual content: C Brayne, L Forsgren, J Petersson, O Hansson, D Lindqvist, C Ruffmann, L Ishihara, R Luben, L Arriola, A Bergareche, D Gavrila, ME Erro, N Vanacore, C Sacerdote, R Vermeulen, M Seleen, S Sieri, G Masala, S Ramat, A Kyrozis, A Thricopolou, S Panico, A Mattiello, R Kaaks, B Teucher, V Grote, M Kloss, L Middleton, E Riboli

Conflict of interest statement: we declare that we have no conflict of interests.

Acknowledgments: This research has been made possible thanks to a grant of the European Community (5th Framework Programme) to Prof. Paolo Vineis (grant QLK4CT199900927); and a grant of the Compagnia di San Paolo to the ISI Foundation. All authors are independent form founders. Mortality data from the Netherlands are obtained from "Statistics Netherlands". In addition we would like to thank for their financial support: Europe Against cancer Program of the European Commission (SANCO); ISCIII, Red de Centros RCESP, C03/09; Spanish Ministry of Health (ISCIII RETICC RD06/0020); Deutsche Krebshilfe; Deutsches Krebsforschungszentrum; German Federal Ministry of Education and Research; Danish Cancer Society; Health Research Fund (FIS) of the Spanish Ministry of Health; Spanish Regional Governments of Andalucia, Asturias, Basque Country, Murcia and Navarra; Spanish Ministry of Health (ISCIII RETICC RD06/0020)Cancer Research U.K.; Medical Research Council, United Kingdom; Stroke Association, United Kingdom; British Heart Foundation; Department of Health, United Kingdom; Food Standards Agency, United Kingdom; Wellcome Trust, United Kingdom Greek Ministry of Health; Greek Ministry of Education; Italian Association for Research on Cancer (AIRC); Italian National Research Council; Dutch Ministry of Public Health, Welfare and Sports (VWS); Netherlands Cancer Registry (NKR); LK Research Funds; Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland); World Cancer Research Fund (WCRF); Statistics Netherlands (The Netherlands); Swedish Cancer; Swedish Scientific Council; Regional Government of Skåne and Västerbotten, Sweden; Norwegian Cancer Society; Research Council of Norway; French League against cancer, Inserm, Mutuelle Generale l'Education National and IGR. Claudio Ruffmann received funding from 'Fondazione Grigioni per la lotta al Morbo di Parkinson

References

1. Poewe W. Non-motor symptoms in Parkinson's disease. European journal of neurology : the official journal of the European Federation of Neurological Societies. 2008;15 Suppl 1:14-20. Epub 2008/04/11.
2. de Rijk MC, Tzourio C, Breteler MM, Dartigues JF, Amaducci L, Lopez-Pousa S, 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. Journal of neurology, neurosurgery, and psychiatry. 1997;62(1):10-5. Epub 1997/01/01.
3. Lees AJ, Hardy J, Revesz T. Parkinson's disease. Lancet. 2009;373(9680):2055-66. Epub 2009/06/16.
4. Bower JH, Maraganore DM, McDonnell SK, Rocca WA. Incidence and distribution of parkinsonism in Olmsted County, Minnesota, 1976-1990. Neurology. 1999;52(6):1214-20. Epub 1999/04/24.
5. Linder J, Stenlund H, Forsgren L. Incidence of Parkinson's disease and parkinsonism in northern Sweden: a population-based study. Movement disorders : official journal of the Movement Disorder Society. 2010;25(3):341-8. Epub 2010/01/29.
6. Alves G, Muller B, Herlofson K, HogenEsch I, Telstad W, Aarsland D, et al. Incidence of Parkinson's disease in Norway: the Norwegian ParkWest study. Journal of neurology, neurosurgery, and psychiatry. 2009;80(8):851-7. Epub 2009/02/28.
7. Chen H, Zhang SM, Schwarzschild MA, Hernan MA, Logroscino G, Willett WC, et al. Folate intake and risk of Parkinson's disease. American journal of epidemiology. 2004;160(4):368-75. Epub 2004/08/03.
8. Bornebroek M, de Lau LM, Haag MD, Koudstaal PJ, Hofman A, Stricker BH, et al. Nonsteroidal anti-inflammatory drugs and the risk of Parkinson disease. Neuroepidemiology. 2007;28(4):193-6. Epub 2007/09/14.
9. Logroscino G, Sesso HD, Paffenbarger RS, Jr., Lee IM. Body mass index and risk of Parkinson's disease: a prospective cohort study. American journal of epidemiology. 2007;166(10):1186-90. Epub 2007/08/22.
10. Driver JA, Smith A, Buring JE, Gaziano JM, Kurth T, Logroscino G. Prospective cohort study of type 2 diabetes and the risk of Parkinson's disease. Diabetes Care. 2008;31(10):2003-5. Epub 2008/07/05.
11. Tan LC, Koh WP, Yuan JM, Wang R, Au WL, Tan JH, et al. Differential effects of black versus green tea on risk of Parkinson's disease in the Singapore Chinese Health Study. American journal of epidemiology. 2008;167(5):553-60. Epub 2007/12/25.
12. Chen H, Mosley TH, Alonso A, Huang X. Plasma urate and Parkinson's disease in the Atherosclerosis Risk in Communities (ARIC) study. American journal of epidemiology. 2009;169(9):1064-9. Epub 2009/03/21.
13. Knekt P, Kilkkinen A, Rissanen H, Marniemi J, Saaksjarvi K, Heliovaara M. Serum vitamin D and the risk of Parkinson disease. Arch Neurol. 2010;67(7):808-11. Epub 2010/07/14.
14. Qiu C, Hu G, Kivipelto M, Laatikainen T, Antikainen R, Fratiglioni L, et al. Association of blood pressure and hypertension with the risk of Parkinson disease: the National FINRISK Study. Hypertension. 2011;57(6):1094-100. Epub 2011/05/04.
15. Huang X, Abbott RD, Petrovitch H, Mailman RB, Ross GW. Low LDL cholesterol and increased risk of Parkinson's disease: prospective results from Honolulu-Asia Aging Study. Mov Disord. 2008;23(7):1013-8. Epub 2008/04/03.
16. Palacios N, Gao X, McCullough ML, Jacobs EJ, Patel AV, Mayo T, et al. Obesity, diabetes, and risk of Parkinson's disease. Movement disorders : official journal of the Movement Disorder Society. 2011;26(12):2253-9. Epub 2011/07/09.
17. Xu Q, Park Y, Huang X, Hollenbeck A, Blair A, Schatzkin A, et al. Diabetes and risk of Parkinson's disease. Diabetes Care. 2011;34(4):910-5. Epub 2011/03/08.
18. Riboli E, Hunt KJ, Slimani N, Ferrari P, Norat T, Fahey M, et al. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. Public Health Nutr. 2002;5(6B):1113-24. Epub 2003/03/18.
19. Ferrari P, Day NE, Boshuizen HC, Roddam A, Hoffmann K, Thiebaut A, et al. The evaluation of the diet/disease relation in the EPIC study: considerations for the calibration and the disease models. International journal of epidemiology. 2008;37(2):368-78. Epub 2008/01/09.
20. Danesh J, Saracci R, Berglund G, Feskens E, Overvad K, Panico S, et al. EPIC-Heart: the cardiovascular component of a prospective study of nutritional, lifestyle and biological factors

in 520,000 middle-aged participants from 10 European countries. Eur J Epidemiol. 2007;22(2):129-41. Epub 2007/02/14.
21. Langenberg C, Sharp S, Forouhi NG, Franks PW, Schulze MB, Kerrison N, et al. Design and cohort description of the InterAct Project: an examination of the interaction of genetic and lifestyle factors on the incidence of type 2 diabetes in the EPIC Study. Diabetologia. 2011;54(9):2272-82. Epub 2011/07/01.
22. Rajput AH, Offord KP, Beard CM, Kurland LT. Epidemiology of parkinsonism: incidence, classification, and mortality. Ann Neurol. 1984;16(3):278-82. Epub 1984/09/01.
23. Granieri E, Carreras M, Casetta I, Govoni V, Tola MR, Paolino E, et al. Parkinson's disease in Ferrara, Italy, 1967 through 1987. Arch Neurol. 1991;48(8):854-7. Epub 1991/08/01.
24. Marttila RJ, Rinne UK. Epidemiology of Parkinson's disease in Finland. Acta Neurol Scand. 1976;53(2):81-102. Epub 1976/02/01.
25. Kuopio AM, Marttila RJ, Helenius H, Rinne UK. Changing epidemiology of Parkinson's disease in southwestern Finland. Neurology. 1999;52(2):302-8. Epub 1999/02/05.
26. Mayeux R, Marder K, Cote LJ, Denaro J, Hemenegildo N, Mejia H, et al. The frequency of idiopathic Parkinson's disease by age, ethnic group, and sex in northern Manhattan, 1988-1993. American journal of epidemiology. 1995;142(8):820-7. Epub 1995/10/15.
27. Vines JJ, Larumbe R, Gaminde I, Artazcoz MT. [Incidence of idiopathic and secondary Parkinson disease in Navarre. Population-based case registry]. Neurologia. 1999;14(1):16-22. Epub 1999/03/18. Incidencia de la enfermedad de Parkinson idiopatica y secundaria en Navarra. Registro poblacional de casos.
28. Gibb WR, Lees AJ. A comparison of clinical and pathological features of young- and old-onset Parkinson's disease. Neurology. 1988;38(9):1402-6. Epub 1988/09/01.
29. Hernan MA, Takkouche B, Caamano-Isorna F, Gestal-Otero JJ. A meta-analysis of coffee drinking, cigarette smoking, and the risk of Parkinson's disease. Ann Neurol. 2002;52(3):276-84. Epub 2002/09/03.
30. Foltynie T, Brayne CE, Robbins TW, Barker RA. The cognitive ability of an incident cohort of Parkinson's patients in the UK. The CamPaIGN study. Brain : a journal of neurology. 2004;127(Pt 3):550-60. Epub 2003/12/24.

Table 1: Total population, expected PD cases, potential PD cases, and final diagnoses of the subjects recruited in the centres participating in the NeuroEPIC4PD study

Centre	Population	Expected PD cases	Potential cases (phase II)	Cases reviewed in phase III (% of potential)	Incident PD	Prevalent PD	Total PD	Parkinsonian -related disorders	Other diseases
Umea (SW)	25,717	23-	62	62 (100.0)	55	1	56	6	0
Malmo (SW)	28,097	63	305	299 (98.0)	140	17	166*	79	54
Utrecht (NL)	17,031	27	31	17 (54.8)	13	1	14	3	0
Cambridge (UK)	30,440	75	622	416 (66.9)	213	70	307*	18	91
Heidelberg (DE)	25,538	25	162	77 (47.5)	50	12	62	12	3
San Sebastian (ES)	8,417	8	46	39 (84.8)	31	1	32	6	1
Navarra (ES)	8,084	8	41	41 (100.0)	18	3	21	13	7
Murcia (ES)	8,516	7	99	99 (100.0)	52	1	53	24	22
Turin (IT)	10,587	10	143	81 (56.6)	28	1	29	44	8
Varese (IT)	11,896	11	13	13 (100.0)	3	3	7*	6	0
Florence (IT)	13,596	13	57	50 (87.7)	23	3	26	3	21
Naples (IT)	5,061	4	22	22 (100.0)	13	0	13	0	9
Greece (GR)	27,514	35	120	120 (100.0)	95	0	95	16	9
Total	**220,494**	**309**	**1,723**	**1,336**	**734**	**113**	**881**	**230**	**225**

* incident and prevalent PD cases do not total all PD cases, as there are missing data on date of diagnosis

Table 2: General and clinical characteristics of the PD cases, using the NeuroEPIC4PD labels

	Definite PD	Very likely PD	Probable PD	Possible PD	All PD
N	199	275	146	261	881
Male (%)	77 (38.7)	164 (59.6)	73 (50.0)	144 (55.2)	458 (52.0)
Mean age at recruitment (SD)	59.1 (8.4)	61.5 (8.0)	61.6 (6.6)	64.8 (8.1)	61.9 (8.2)
Mean age at symptom onset (SD)	64.9 (8.1)	65.4 (9.4)	69.0 (7.1)	71.0 (8.1)	66.8 (8.3)
Mean age at diagnosis (SD)	65.7 (7.8)	66.2 (10.2)	69.3 (7.5)	71.0 (9.2)	67.9 (9.2)
Cardinal signs^					
Resting tremor (%)	151 (75.9)	121 (63.0)	91 (62.3)	29 (24.8)	392 (59.9)
Bradykinesia (%)	192 (96.5)	147 (76.6)	118 (80.8)	43 (36.8)	500 (76.5)
Rigidity (%)	191 (96.0)	132 (68.8)	114 (78.1)	42 (35.9)	479 (73.2)
Postural instability (%)	33 (16.6)	31 (16.2)	31 (21.2)	19 (16.2)	114 (17.4)
Treatment					
Dopamine receptor agonists (%)	133 (66.8)	128 (46.5)	67 (45.9)	42 (16.1)	370 (42.0)
Levodopa (%)	177 (88.9)	163 (59.3)	114 (78.1)	79 (30.3)	533 (60.5)
COMT inhibitors (%)	85 (42.7)	84 (30.5)	35 (24.0)	25 (9.6)	229 (26.0)
MAO-B inhibitors (%)	61 (30.7)	65 (23.6)	36 (24.7)	24 (9.2)	186 (21.1)
Amantadine (%)	16 (8.0)	16 (5.8)	8 (5.5)	5 (1.9)	45 (5.1)
Antimuscarinic (%)	2 (1.0)	5 (1.8)	0 (0.0)	5 (1.9)	12 (1.4)
DBS -STN (%)	9 (4.5)	3 (1.1)	1 (0.9)	0 (0.0)	13 (0.3)
DBS-VIM (%)	1 (0.5)	1 (0.4)	0 (0.0)	0 (0.0)	2 (0.2)
DBS-GP (%)	1 (0.5)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.1)
Death (%)	16 (8.0)	99 (36.0)	46 (31.5)	128 (49.0)	289 (32.8)

^ coded as cardinal sign, i.e. symptom recorded during the first 3 years of disease; * figures do not add up to 100% as the option Not Available was possible for each of the questions; PIGD: Postural instability/gait disturbance; DBS: deep brain stimulation; STN: substantia nigra; VIM: ventral intermediate nucleus of the thalamus; GP: Globus Pallidus.

Figure legends:

Figure 1: Number of expected PD cases, potential cases, cases reviewed in phase II, and ascertained PD cases by centre, in the NeuroEPIC4PD study

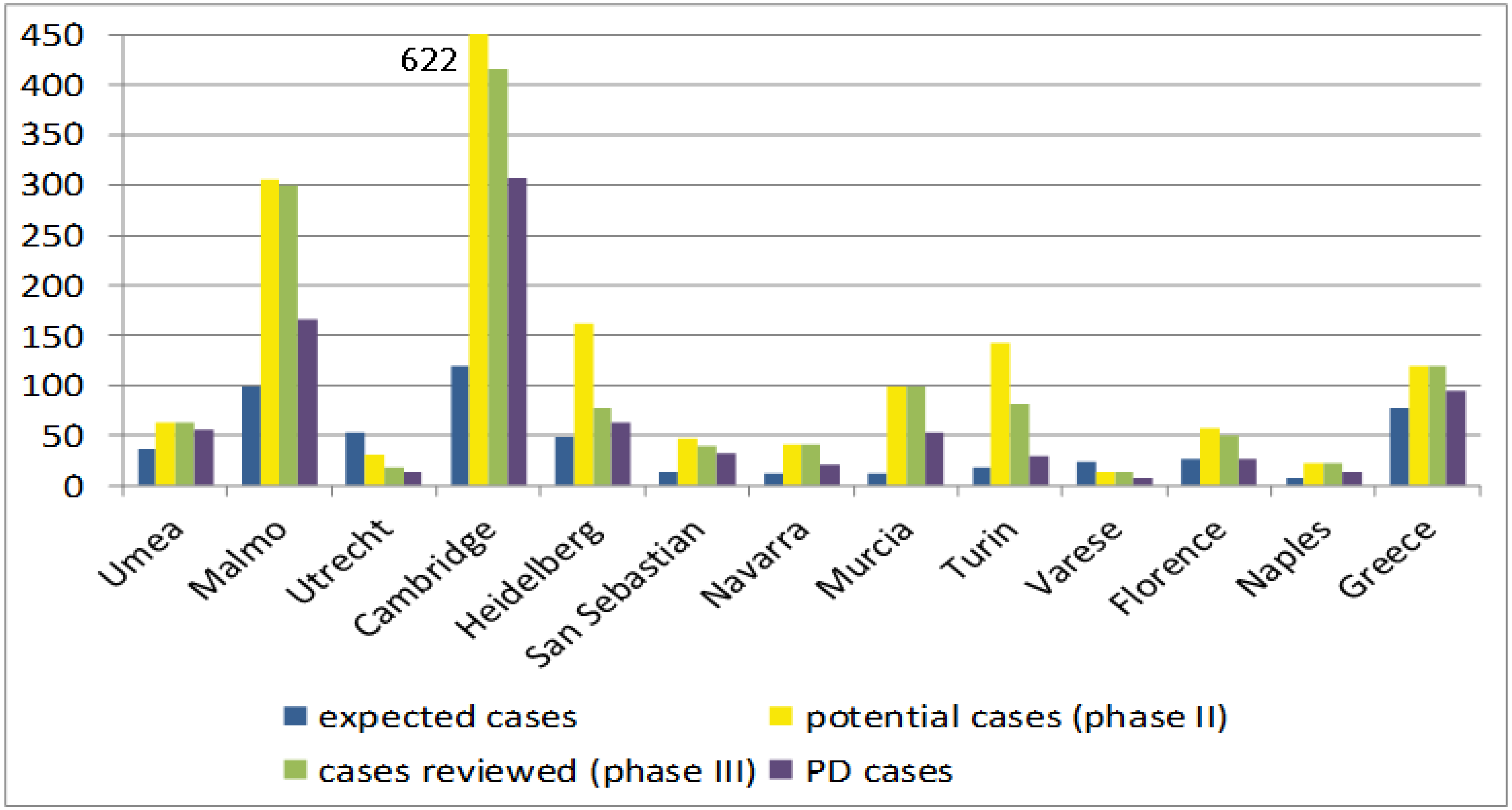


Figure 2: Age-specific cumulative incidence of PD by country, in men.

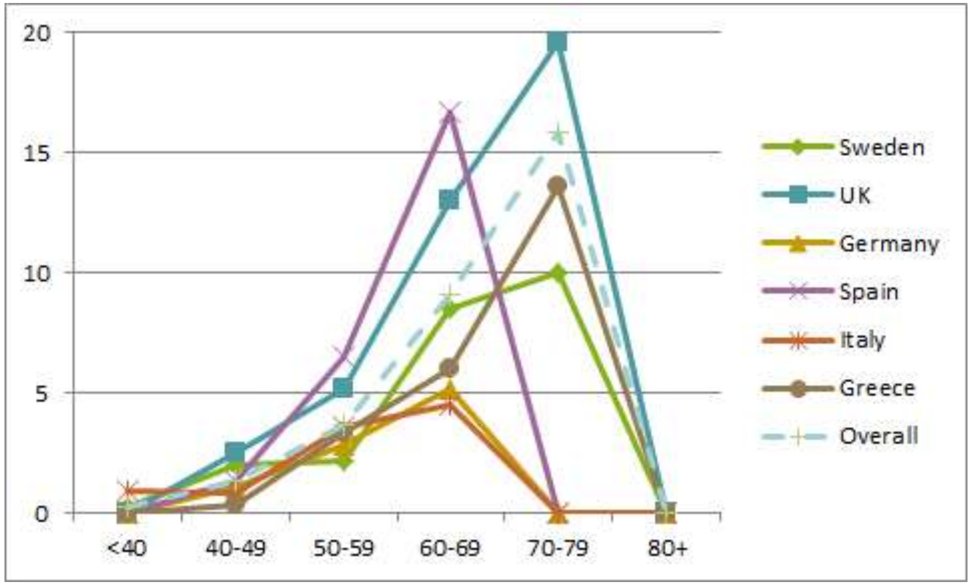
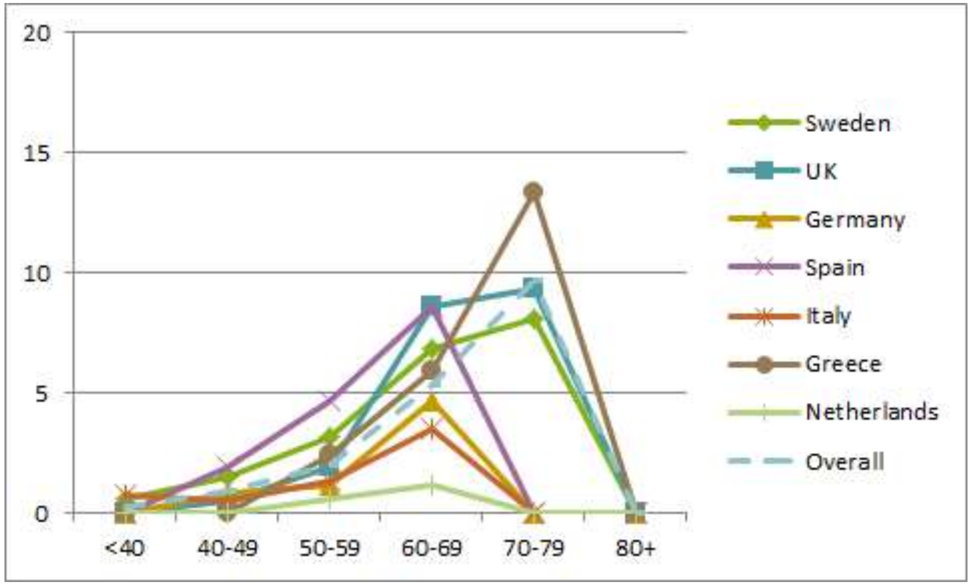


Figure 3: Age-specific cumulative incidence of PD by country, in women (y-axis is deliberately on the same scale as in Figure 2 for facilitating comparison).



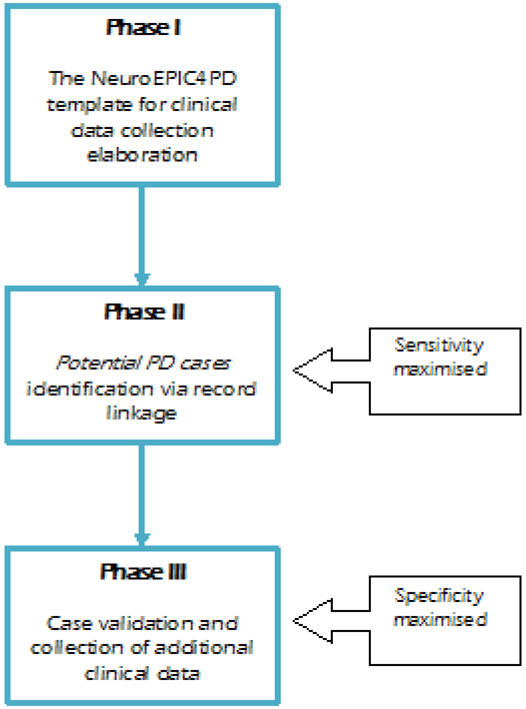
Online Table 1 Demographic characteristics of the NeuroEPIC4PD cohort

Centre	N	Men (%)	Mean age (yrs)	Mean years of follow-up	Person years (SD)	Last follow-up	Deaths
Umea (SW)	25,717	12,427 (48.3)	46.0 (10.3)	13.7 (2.7)	352,832 (432)	22/02/2009	1,391 (5.4)
Malmo (SW)	28,097	11,062 (39.4)	58.1 (7.6)	13.9 (3.0)	390,137 (495)	31/12/2008	4,421 (15.7)
Utrecht (NL)	17,031	0 (0.0)	57.7 (6.0)	12.9 (2.1)	220,190 (274)	01/01/2009	1,441 (8.5)
Cambridge (UK)	30,440	13,697 (45.0)	59.4 (9.4)	13.0 (3.1)	396,149 (534)	06/03/2010	5,969 (19.6)
Heidelberg (DE)	25,538	11,929 (46.7)	50.9 (8.1)	11.3 (2.0)	288,269 (316)	16/06/2010	1,407 (5.5)
San Sebastian (ES)	8,417	4,158 (49.4)	49.6 (7.7)	12.5 (1.5)	105,178 (139)	17/05/2008	450 (5.4)
Navarra (ES)	8,084	3,908 (48.3)	49.6 (7.6)	13.8 (1.7)	111,595 (152)	26/08/2009	410 (5.1)
Murcia (ES)	8,516	2,685 (31.5)	48.8 (8.2)	13.2 (1.5)	112,641 (137)	09/11/2008	339 (4.0)
Turin (IT)	10,587	6,040 (57.1)	50.1 (7.7)	13.4 (2.3)	141,770 (238)	21/05/2010	497 (4.7)
Varese (IT)	11,896	2,474 (20.8)	51.6 (8.2)	11.3 (1.8)	134,002 (193)	31/12/2006	415 (3.5)
Florence (IT)	13,596	3,513 (25.8)	51.5 (7.7)	12.0 (1.8)	163,572 (207)	31/12/2007	438 (3.2)
Naples (IT)	5,061	0 (0.0)	50.2 (7.7)	11.6 (1.2)	58,748 (85)	31/12/2006	140 (2.8)
Greece (GR)	27,514	11,427 (41.5)	53.1 (12.6)	9.7 (3.1)	266,396 (513)	23/12/2009	2,155 (7.8)
Total	220,494	83,320 (37.8)	53.1 (10.0)	12.4 (2.8)	2,741,780 (1,330)	16/06/2010	19,473 (8.8)

Online Table 2: Parkinson's disease cases divided by NeuroEPIC4PD label, and parkinsonian-related disorders, by centre in the NeuroEPIC4PD study

Centre	Cases reviewed in phase II	Definite PD	Very likely PD	Probable PD	Possible PD	Total PD	MSA	PSP	Vascular Parkinsonism	LBD	Essential tremor	PD + essential tremor	Unclassifiable parkinsonism
Umea (SW)	62	45	3	8	0	56	3	1	0	0	0	1	1
Malmo (SW)	299	24	41	49	52	166	8	10	17	15	4	3	22
Utrecht (NL)	17	8	5	0	1	14	1	0	0	0	1	0	1
Cambridge (UK)	416	7	130	13	157	307	2	1	4	4	4	1	2
Heidelberg (DE)	77	2	16	39	5	62	6	1	0	4	0	0	1
San Sebastian (ES)	39	11	21	0	0	32	0	2	1	2	1	0	0
Navarra (ES)	41	6	9	3	3	21	1	0	0	3	2	2	5
Murcia (ES)	99	7	17	16	13	53	1	5	5	2	5	2	4
Turin (IT)	81	0	10	2	17	29	2	1	2	1	6	0	32
Varese (IT)	13	2	1	2	2	7	1	0	0	1	1	0	3
Florence (IT)	50	11	14	1	0	26	0	1	1	0	1	0	0
Naples (IT)	22	8	5	0	0	13	0	0	0	0	0	0	0
Greece (GR)	120	68	3	13	11	95	1	0	4	2	5	0	4
Total	**1,336**	**199**	**275**	**146**	**261**	**881**	**26**	**22**	**34**	**34**	**30**	**9**	**75**

Online Figure 1: Framework of the phases for case ascertainment procedures in the NeuroEPIC4PD study



Online Figure 2: Matrix combining the two variables amount and quality of data available and the degree of confidence of the neurologist expert in movement disorders reviewing the evidence, resulting in the NeuroEPIC4PD label to be attached to each diagnosis

		Data amount and quality		
		Excellent	Good	Poor
Degree of confidence	High	Definite	Very likely	Very likely
	Medium	Probable	Probable	Possible
	Low	Possible	Possible	Possible

Online Figure 3: Sources of information used for Phase I, by centre

Centre	Drug registry	Hospital records	GP records	Self-reported	Mortality records	Other
Umea (SW)		√				
Malmo (SW)		√				
Utrecht (NL)				√		
Cambridge (UK)		√				
Heidelberg (DE)		√		√	√	
San Sebastian (ES)	√	√	√		√	
Navarra (ES)	√		√		√	
Murcia (ES)	√		√		√	
Turin (IT)	√	√			√	
Varese (IT)	√	√		√		
Florence (IT)	√	√		√	√	
Naples (IT)		√		√		
Greece (GR)				√		

Online Figure 4: Number of tremor-dominant, akinetic-rigid, and postural-instability/gait disturbance (PIGD) predominant forms of PD at onset, by NeuroEPIC4PD label. Data is not available for 14 definite PD, 146 very likely, 40 probable, and 223 possible PD.