

Post Hoc Parkinson's Disease: Identifying an Uncommon Disease in the Cardiovascular Health Study

T.G. Ton^a S. Jain^d R. Boudreau^e E.L. Thacker^b E.S. Strotmeyer^e
A.B. Newman^e W.T. Longstreth^{a, b} H. Checkoway^{b, c}

^aDepartment of Neurology, School of Medicine, and Departments of ^bEpidemiology and ^cEnvironmental and Occupational Health Sciences, School of Public Health, University of Washington, Seattle, Wash., and

^dDepartment of Neurology, University of Pittsburgh Medical Center, and ^eDepartment of Epidemiology, Center for Aging and Population Health, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pa., USA

Key Words

Cardiovascular Health Study • Cohort study • Epidemiology • International Classification of Diseases, 9th edition • Smoking • Parkinson's disease

Abstract

Background: Although ongoing cohort studies offer a unique opportunity to apply existing information collected prospectively to further the scientific understanding of Parkinson's disease (PD), they typically have limited information for clinical diagnosis. **Methods:** We used combinations of self-report, International Classification of Diseases – 9th edition codes and antiparkinsonian medications to identify PD in the Cardiovascular Health Study. To determine whether the expected inverse association between smoking and PD is evident using our outcome definitions, we assessed baseline smoking characteristics for various definitions of PD. **Results:** We identified 60 cases with prevalent PD (1.0%; 95% confidence interval, CI = 0.8–1.3%) and 154 with incident PD by year 14. Clear associations were observed for current smokers (odds ratio, OR = 0.50; 95% CI = 0.26–0.95) and for those who smoked ≥ 50 pack-years (OR = 0.53; 95% CI = 0.29–0.96). Estimates for smoking were similar when ≥ 2 data sources were required. Estimates for self-report alone

were attenuated towards null. **Conclusions:** Using multiple data sources to identify PD represents an alternative method of outcome identification in a cohort that would otherwise not be possible for PD research. Ongoing cohort studies can provide settings in which rapid replication and explorations of new hypotheses for PD are possible.

Copyright © 2010 S. Karger AG, Basel

Introduction

Although Parkinson's disease (PD) was first described in 1817, genetic mutations account for only a minority of cases [1], and the etiology of most sporadic PD still remains obscure with the exception of a consistently observed inverse association with cigarette smoking [2]. The causes of PD remain poorly understood, in large part because of the difficulty in conducting epidemiologic studies of a condition in which the diagnosis is often not straightforward and for which registries covering large populations do not exist. Case-control studies are particularly suited for investigations of less common conditions but are often limited by the retrospective recall or ascertainment of exposures. On the other hand, prospective cohort studies provide an invaluable longitudinal

component allowing for the ability to assess changes in exposure or clinical disease over time but are time-consuming and costly to assemble and thus impractical for less common diseases such as PD. However, large ongoing cohort studies of more common outcomes such as cardiovascular events offer unique opportunities to apply a wealth of existing information to further the scientific understanding of PD but are often underutilized, in large part, due to the difficulty in defining PD, and because PD is often not considered a major risk factor or outcome in these studies and therefore not rigorously assessed. Because complete screening and evaluation for PD by a movement specialist in ongoing cohort studies are not typically feasible, identification of PD necessarily relies on self-report or medical records. Often, however, medical records are incomplete, and direct contact with participants is either impossible due to mortality over many years of follow-up, or prohibitive due to cost and participant burden. Despite these challenges, we attempted to identify PD in an ongoing cohort study of elderly participants, the Cardiovascular Health Study (CHS), by leveraging the overlapping nature of multiple data sources that are readily available after 14 years of follow-up. We assessed whether we would observe the expected inverse association between smoking and PD using various outcome definitions. We reasoned that observing significant associations in the correct direction between smoking and PD would provide support for exploring other unknown risk factors for PD in the CHS using our outcome definitions; conversely, the lack of an association would undermine efforts to conduct PD research in the CHS with the current definitions proposed here.

Materials and Methods

The CHS is an NHLBI-funded prospective cohort study of coronary heart disease and stroke in adults aged 65 years and older and in whom data were collected prospectively through year 14 of the study. The CHS recruited a total of 5,888 elderly men and women from 4 communities in the USA. Details about recruitment are described elsewhere [3]. In brief, 5,201 men and women were recruited into the original cohort, and an additional 687 African Americans were recruited after the initial baseline survey. Extensive physical and laboratory evaluations were performed at baseline to identify the presence and severity of cardiovascular risk factors such as hypertension, hypercholesterolemia and glucose intolerance, subclinical disease such as carotid artery atherosclerosis, left ventricular enlargement and transient ischemia, and clinically overt cardiovascular disease. Participants were contacted twice per year for new diagnoses, hospitalizations and medical procedures, and examined annually from 1989 to 1999, after which time telephone follow-up was continued. Reports of hospi-

talizations were cross-examined in the Health Care Finance Administration records to help enhance surveillance of events and deaths.

We used the following sources of data to screen for potential PD:

- (1) *Self-Report*. Participants in the original cohort were asked if they had a physician diagnosis of PD in the first year of follow-up and in year 11. Subjects in the African American cohort were asked to report a physician diagnosis of PD at their baseline and year 11.
- (2) *Medication Inventory*. Prescription medications were collected from the onset of the CHS in 1989. At annual visits, subjects were asked to report on and bring in all vials for medications that were taken within 14 days of the visit. All medications were visually inspected, and information on medication name, class, dose and frequency was recorded for each medication. A movement disorders specialist (S.J.) reviewed antiparkinsonian medications (APM) from the annual medication inventory to screen for potential PD: ethopropazine HCl (Parisdol®); levodopa-containing compounds (Sinemet®, Sinemet CR/ER®, Larodopa®, Dopar®, Stalevo®); pergolide (Permax®); pramipexole (Mirapex); ropinirole (Requip); bromocriptine (Parlodel®); selegiline (Eldepryl®, Deprenyl®, Emsam®, Zelapar); rasagiline (Azilect®); entacapone (Comtan®); tolcapone (Tasmar®); trihexyphenidyl HCl (Artane, Tremin); procyclidine (Kemadrin®); biperiden (Akineton®); benzotropine (Cogentin®); amantadine (Symmetrel®).
- (3) *Hospitalization Records*. Hospitalization discharge records included information on admission date, discharge date, vital status and International Classification of Diseases, 9th edition (ICD-9), discharge codes for all hospitalizations of CHS subjects. The discharge abstraction form used by the CHS allowed for a list of 10 different ICD-9 codes, with instructions to append additional pages for subjects with additional conditions. We used ICD-9 code 332.0 to identify PD patients. Up to June 2006, 5,326 of 5,888 (90%) subjects in the CHS had been hospitalized.

We created several definitions for PD based on data available in the CHS and evidence from published literature regarding the increased accuracy of multiple data sources over a single data source in identifying PD [4]. Participants were classified with PD according to the following definitions:

Definition 1. Participants who have at least 1 data source providing evidence of PD. This definition included participants with evidence of PD from only a single data source such as self-report or APM or ICD-9 as well as those with evidence of PD from 2 or more corroborating sources. Those with no evidence of PD were categorized as non-PD.

Definition 2. Participants who have evidence of PD from at least 2 data sources. Possible combinations included self-report + APM or self-report + ICD-9 or APM + ICD-9, as well as participants with evidence from all 3 data sources. Those with no evidence of PD were categorized as non-PD.

Definition 3. Participants who have evidence of PD from all 3 data sources: self-report + APM + ICD-9 code. Those with no evidence of PD were categorized as non-PD.

For each definition, we assigned to each participant with PD a date corresponding to the earliest evidence of PD from any data source. Those whose 'date of first evidence' occurred at baseline or in the first year of follow-up were defined as having prevalent

Table 1. Prevalent and incident PD in the CHS according to various definitions of PD

Definition	Sources	Total	Prevalent	Prevalence at baseline n/100	Incidence
One or more data sources	ICD-9 or APM or self-report, or 2 or more data sources	214	60	1.02 (0.78–1.31)	154
Two or more data sources	ICD-9 and APM or [CD-9 and self-report or APM and self-report or all 3 data sources	109	46	0.78 (0.57–1.04)	63
All 3 data sources	ICD-9 and APM and self-report	58	34	0.58 (0.40–0.81)	24
Self-report only	self-report only	122	47	0.80 (0.57–1.06)	75

Figures in parentheses indicate 95% CI.

PD. Participants who were identified as having PD during the subsequent follow-up were considered as incident PD cases. We reviewed the medication inventory for drugs that could potentially induce parkinsonism, for example neuroleptics and related compounds (see Appendix 1). We reclassified participants who took drugs that could potentially induce parkinsonism as non-PD if evidence of such drugs occurred in the medication inventory at the same time or before the 'date of first evidence'. By year 14, 39 participants had taken parkinsonism-inducing medications before their assigned date. Those who took parkinsonism-inducing drugs after this date remained classified as PD, since such agents can be used to treat symptoms which occur during PD (e.g. psychosis, nausea).

Statistical Analyses

In the absence of a clinical diagnosis and the ability to directly validate our definitions of PD, we sought to compare results obtained by our definitions to those reported by published studies. Specifically, we calculated the prevalence of PD at baseline and assessed the association of smoking with PD as classified by the different definitions of PD. Prevalence of PD was calculated as the proportion of CHS participants who had PD at the beginning of the study, defined at baseline or first year of follow-up. A Poisson distribution was used to calculate 95% confidence intervals (CI). Because hospitalizations and ICD-9 codes are not expected to occur at baseline, we used capture-recapture techniques for 2-source models [5] to estimate the number of participants with PD potentially missed at baseline by the medication inventory and self-report, and adjusted baseline prevalence estimates accordingly.

In all analyses relating to baseline smoking characteristics, we excluded participants with prevalent PD at baseline. We performed 3 sets of analyses in which results from definitions 1 and 2 were compared to those obtained from the definition using self-report alone. Because of difficulty in assigning the exact time of onset, we used unconditional logistic regression to obtain odds ratios (OR) and 95% CI associated with PD risk, and adjusted all models for age, sex and African American race. Baseline smoking variables were characterized in the following ways: smoking status (never smoked, ever smoked, current smoker), quartiles of total pack-years (never smokers, <13, 13–27, 28–49, ≥50), and years

since quitting (never smokers, ≥30, 20–29, 10–19, <10, current smoker). All analyses were conducted in Stata 10.1 (College Station, Tex., USA).

Results

Characteristics of CHS participants have been described elsewhere [6]. A summary of prevalent and incident PD by various definitions is provided in table 1. Using 1 or more data sources to screen for PD, we identified 214 CHS participants with evidence of PD. Among these, 60 had PD at baseline for an overall prevalence estimate of 1.0% (95% CI = 0.8–1.3%) within the CHS. Age-specific prevalence estimates of PD at baseline are 0.70% for ages 65–74, 1.67% for ages 75–84 and 1.81% for ages 85 and older. Using the maximum likelihood estimator in capture-recapture techniques, we estimated that an additional 5 participants may have been missed at baseline, corresponding to an overall baseline prevalence of 1.1% (95% CI = 0.9–1.4%) in the CHS. Characteristics of participants with and without prevalent PD at baseline are presented in table 2. Using 1 or more data sources to identify PD at baseline, those with PD were significantly older ($p = 0.002$), more likely to be men ($p = 0.002$) and less likely to be African American ($p = 0.02$). Those with and without PD at baseline were not different in relation to smoking status. We observed a suggestive trend in decreasing odds for decreasing number of years since quitting smoking ($p = 0.08$).

A total of 154 (2.6% of 5,828) participants developed PD during 14 years of follow-up. ICD-9 codes were the most frequent source ($n = 96$) in identifying those with PD, followed by self-report ($n = 74$) and APM ($n = 71$). Sixty-three participants (40.9%) had evidence of PD from

Table 2. Demographic and baseline smoking characteristics in relation to baseline prevalent PD as defined by at least 1 data source

Characteristic	Baseline PD				p value
	PD (n = 60)		no PD (n = 5,828)		
	n	%	n	%	
Age at baseline					0.002
65–74 years	27	45.0	3,867	66.4	
75–85 years	29	48.3	1,740	29.9	
>85 years	4	6.7	221	3.8	
Male sex	37	61.7	2,458	42.2	0.002
Education					0.26
<HS	14	23.3	1,718	29.6	
HS	14	23.3	1,606	27.6	
>HS	32	53.3	2,487	42.8	
African American	3	5.0	921	15.8	0.02
Smoking status					0.12
Never	34	56.7	2,704	46.4	
Former	21	35.0	2,423	41.6	
Current	5	8.3	695	11.9	
Total pack-years					0.21
Never smoker	34	56.7	2,704	48.0	
1st quartile (0–13)	9	15.0	856	15.2	
2nd quartile (14–27)	3	5.0	660	11.7	
3rd quartile (28–49)	8	13.3	515	9.1	
4th quartile (≥50)	6	10.0	898	15.9	
Years since quitting					0.08
Never smoker	34	56.7	2,704	47.0	
≥30	7	11.7	626	10.9	
20–29	4	6.7	579	10.1	
10–19	8	13.3	595	10.4	
<10	3	3.3	549	9.6	
Current smoker	5	8.3	695	12.1	

HS = High school. Numbers may not add up to column total due to missing values; p values assessed by Pearson's χ^2 test for dichotomous categories, Fisher's exact test for categorical variables with small numbers, score test for linear trend for ordinal variables; pack-years: quartiles of total pack-years determined among non-PD smokers.

at least 2 different sources, 24 of whom had evidence of PD by all 3 sources (fig. 1). Univariate associations between baseline smoking characteristics and incident PD for various definitions are summarized in table 3. We compared demographic and baseline smoking characteristics between those with and without incident PD as defined by having evidence from at least 1 data source (definition 1), 2 or more corroborating data sources (definition 2) and by self-report alone. Those with PD were

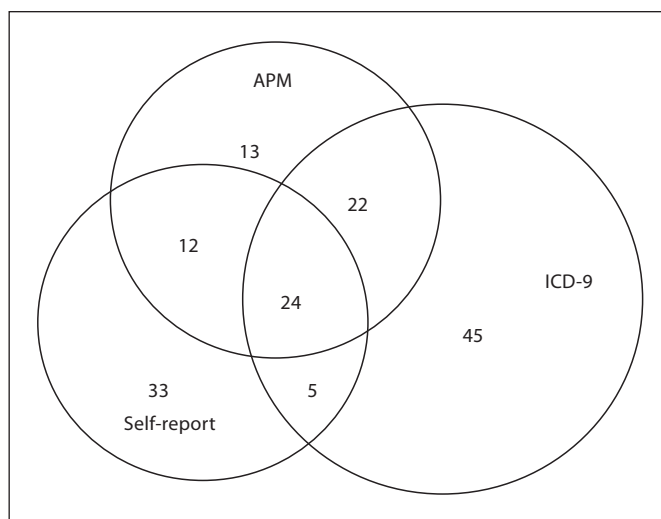


Fig. 1. Venn diagram of sources used to identify PD in the CHS.

significantly more likely to be men for definitions 1 and 2. This association was attenuated to nonsignificant levels when we used self-report as the sole data source to classify PD. Age at baseline was different only for the definition using self-report in which participants with PD were slightly younger than those without PD ($p = 0.06$). Those with PD were significantly less likely to be African American for definition 2 ($p = 0.04$). Statistically significant inverse associations between smoking and PD were not obvious in univariate analyses for any of these 3 definitions of PD. An inverse association was suggestive for a decreasing trend according to years since quitting ($p = 0.08$) for PD defined by definition 1.

OR obtained from models adjusting for age, sex and African American race are presented in table 4. When 1 or more data sources were used to define PD (definition 1), current smokers have a 50% reduced likelihood of PD relative to never smokers (OR = 0.50; 95% CI = 0.26–0.95). A significant linear trend in odds for smoking (i.e. never, former, current) was evident for definition 1 ($p = 0.04$) but not for definition 2 or for self-report alone. Most estimates of association for levels of smoking were less than 1.0, with a clear association observed for those who smoked a total of ≥ 50 pack-years (OR = 0.53; 95% CI = 0.29–0.96) relative to never smokers. Estimates, in general, decreased for increasing number of total pack-years and years since quitting. Linear trends in the OR achieved statistical significance only for total pack-years ($p = 0.03$). When 2 or more data sources were required for a PD classification (definition 2), estimates for current, former and

Table 3. Demographic and baseline smoking characteristics of subjects identified with incident PD in the CHS according to various outcome definitions

Characteristic	Definition 1 (at least 1 data source)				p value	Definition 2 (at least 2 data sources)				p value	Self-report only				p value
	PD (n = 154)		no PD (n = 5,674)			PD (n = 63)		no PD (5,674)			PD (n = 75)		no PD (n = 5,766)		
	n	%	n	%		n	%	n	%		n	%	n	%	
<i>Age at baseline</i>					0.26					0.94					0.06
65–74 years	108	70.1	3,759	66.3		43	68.3	3,759	66.3		58	77.3	3,815	77.3	
75–85 years	42	27.3	1,698	29.9		16	25.4	1,698	29.9		15	20.0	1,731	30.0	
>85 years	4	2.6	217	3.8		4	6.4	217	3.8		2	2.7	220	3.8	
Male sex	84	54.6	2,374	41.8	0.002	40	63.5	2,374	41.8	0.001	39	52.0	2,427	42.1	0.08
<i>Education</i>					0.58					0.58					0.48
<HS	42	27.3	1,676	29.6		17	27.0	1,676	29.6		19	25.3	1,703	29.6	
HS	44	28.6	1,562	27.6		17	27.0	1,562	27.6		22	29.3	1,589	27.6	
>HS	68	44.2	2,419	42.8		29	46.0	2,419	42.8		33	44.3	2,457	42.7	
African American	24	15.6	897	15.8	0.94	4	6.4	897	15.8	0.04	14	18.7	908	15.8	0.50
<i>Smoking status</i>					0.19					0.42					0.49
Never	75	48.7	2,629	46.4		30	47.6	2,629	46.4		35	46.7	2,681	46.6	
Former	68	44.2	2,355	41.5		29	46.0	2,355	41.6		35	46.7	2,388	41.5	
Current	11	7.1	684	12.1		4	6.4	684	12.0		5	6.7	691	12.0	
<i>Total pack-years</i>					0.26					0.53					0.66
Never smoker	75	50.0	2,630	48.0		30	48.4	2,630	48.0		35	47.3	2,682	48.1	
1st quartile (0–13)	22	14.7	743	13.6		10	16.1	743	13.6		12	16.2	755	13.6	
2nd quartile (14–27)	21	14.0	675	12.3		8	12.9	675	12.3		10	13.5	685	12.3	
3rd quartile (28–49)	18	12.0	720	13.1		9	14.5	720	13.1		10	13.5	728	13.1	
4th quartile (≥50)	14	9.3	716	13.0		5	8.1	716	13.0		7	9.5	723	13.0	
<i>Years since quitting</i>					0.08					0.24					0.33
Never smoker	75	49.3	2,629	47.0		30	47.6	2,629	47.0		35	46.7	2,681	47.2	
30	21	13.8	605	10.8		9	14.3	605	10.8		11	14.7	615	10.8	
20–29	18	11.8	561	10.0		9	14.3	561	10.0		9	12.0	570	10.0	
10–19	15	9.9	580	10.4		7	11.1	580	10.4		9	12.0	587	10.4	
<10	12	7.9	537	9.6		4	6.4	537	9.6		6	8.0	542	9.5	
Current smoker	11	7.2	684	12.2		4	6.4	684	12.2		5	6.7	691	12.2	

HS = High school. Numbers may not add up to column total due to missing values. p values assessed by Pearson's χ^2 test for dichotomous categories, Fisher's exact test for categorical variables with small numbers, score test for linear trend for ordinal variables; pack-years: quartiles of total pack-years determined among non-PD smokers.

ever smokers remained similar to their counterparts in definition 1. The corresponding confidence intervals, however, widened and associations previously observed using definition 1 lost statistical significance with the use of definition 2. Linear trends in the OR were weakened for increasing number of pack-years and became only suggestive of a linear trend ($p = 0.12$). When only self-report was used to identify PD, nearly all estimates were attenuated towards the null compared to those in definition 1. None of the baseline smoking characteristics was significantly associated with PD when self-report was the single source used to classify outcome.

Discussion

Using multiple available data sources to identify PD, we obtained an estimate of baseline prevalence within the CHS similar to those reported in other populations with similar age ranges. Prevalence estimates have ranged between 0.8% [7] among men of Japanese ancestry and 1.5% among people of European and Caucasian ancestry [8–12]. As expected, baseline prevalence of PD increased with increasing age, and those with prevalent PD at baseline were significantly older and more likely to be men. However, we did not observe clear associations between baseline smoking status and prevalent PD at baseline, ex-

Table 4. Association between baseline smoking variables and incident PD in the CHS according to several outcome definitions

Baseline smoking characteristics	Definition 1 (≥1 data source)		Linear trend p value	Definition 2 (≥2 data sources)		Linear trend p value	Self-report only		Linear trend p value
	OR	95% CI		OR	95% CI		OR	95% CI	
<i>Smoking status</i>			0.04			0.17			0.17
Never	1.00	reference		1.00	reference		1.00	reference	
Former	0.84	0.59–1.19		0.82	0.48–1.40		0.94	0.57–1.55	
Current	0.50	0.26–0.95		0.49	0.17–1.41		0.45	0.17–1.12	
<i>Total pack-years</i>			0.03			0.12			0.22
Never smoker	1.00	reference		1.00	reference		1.00	reference	
1st quartile (0–13)	0.82	0.57–1.50		1.01	0.49–2.11		1.06	0.54–2.07	
2nd quartile (14–27)	0.95	0.58–1.56		0.87	0.39–1.93		0.95	0.47–1.95	
3rd quartile (28–49)	0.72	0.42–1.23		0.86	0.40–1.84		0.84	0.41–1.73	
4th quartile (≥50)	0.53	0.29–0.96		0.44	0.16–1.14		0.57	0.25–1.32	
<i>Years since quitting</i>			0.33			0.38			0.34
Never smoker	1.00	reference		1.00	reference		1.00	reference	
30	1.01	0.61–1.68		0.93	0.43–2.01		1.21	0.59–2.42	
20–29	0.92	0.54–1.58		1.03	0.47–2.22		1.02	0.47–2.14	
10–19	0.76	0.43–1.34		0.83	0.36–1.91		0.97	0.46–2.05	
<10	0.67	0.36–1.25		0.54	0.19–1.57		0.69	0.29–1.68	
Current smoker	0.50	0.26–0.95		0.49	0.17–1.41		0.45	0.18–1.18	

Adjusted for age, sex and African American race; pack-years: quartiles of total pack-years determined among non-PD smokers.

cept for a suggestion of decreasing trend in risk for decreasing years since smoking.

Our estimates of association for incident PD with baseline smoking characteristics were in the direction we expected for all smoking characteristics when using definition 1 to classify PD, and for most estimates when using definition 2. When self-report was used as the sole source of data to identify PD, estimates were attenuated towards null.

We encountered a number of methodologic issues in our attempt to identify PD within the CHS. First, we did not attempt to calculate an incidence rate of PD within the CHS due to difficulties in accurately defining a date of diagnosis. Such difficulties stem from the fact that the earliest evidence of PD came from 1 of 3 different data sources for each individual, and that each data source could be associated with a different average time from diagnosis. In the CHS, for instance, ICD-9 codes from discharge records became available only after participants had been hospitalized although not necessarily for PD. Although discharge records were available for almost all participants by year 14, they offer little information allowing for an accurate extrapolation of time since PD diagnosis. Similarly, extrapolating an accurate date of diagnosis from the time APM was initiated is difficult because the issue of when and how to start treatment is still

actively debated [13–15]. Nonetheless, we did assign a ‘date of earliest evidence’ to each participant with PD and used it only in an attempt to identify patients who may have previously used parkinsonism-inducing drugs. Given these difficulties, we did not rely on the person-years allocated for each individual by the ‘date of earliest evidence’ to calculate an incidence rate or to conduct a time-to-event analysis.

Our primary limitation was disease misclassification. We attempted to reduce disease misclassification by excluding those identified as having PD who had been taking parkinsonism-inducing drugs before the ‘date of earliest evidence’ of PD. We identified a proportion of potential drug-induced parkinsonism (18%) in our PD group that is similar to what Bower et al. [16] described among subjects in Olmsted County with parkinsonism. Ultimately, without the ability to validate our group of PD against a clinical gold standard, the degree of misclassification is not precisely quantifiable. Nonetheless, we can infer that some misclassification exists in our definitions because estimates for smoking become attenuated with definitions that are expected to have lower specificity [17]. Point estimates from several meta-analyses pooling over 50 studies report relative risks in the range of 0.5–0.7 for ever smokers, 0.6–0.8 for former smokers and 0.2–0.5 for current smokers [2, 18]. Our estimates are within

these same ranges. Use of self-report alone produced slightly attenuated estimates for former smokers and ever smokers compared to those in definition 1. Similar attenuation of estimates between definition 1 and self-report can generally be seen for highest levels of cigarettes per day, total pack-years and years since quitting, suggesting that greater misclassification or other sources of bias may exist in the definition that relies solely on self-report. The associations between smoking and PD appear strongest when 2 or more data sources are used to classify PD, supporting the notion that using multiple data sources can help reduce misclassification [4, 17].

Unfortunately, reducing disease misclassification by requiring 2 or more data sources to provide evidence of PD comes at the cost of statistical power. For PD, self-report alone does not provide a sensitive screen for purposes of assembling a group of cases for epidemiologic investigations. Up to 30% of clinically confirmed cases of PD may be missed by relying solely on self-report [10]. In our study, although self-report identified more outcomes than definition 2, self-report accounted for 48% of the 154 incident PD cases identified by using all 3 available data sources; relying solely on self-report would have reduced the total number of incident events to 74. All associations seen for definition 1 lost statistical significance when other definitions were used. Other reasons exist for considering the use of definition 1 over using self-report alone to identify incident PD. Because self-report of PD was asked in year 11 of the study, those who did not survive until then would have been classified as non-PD by the definition relying on self-report alone. This explanation likely accounts for the observation in our study that incident PD by self-report is significantly more common among younger participants. This is in contrast to the highly significant trend in odds associated with increasing age categories when assessed in relation to baseline PD, as defined by definition 1. In contrast to using self-report alone, both definitions 1 and 2 incorporated data sources that were used frequently over time to assess outcome; the medication inventory was carried out annually, and records with ICD-9 codes appeared more frequently as hospitalizations among participants increased over the follow-up period. Nonetheless, some bias still exists in definitions 1 and 2 since risk of PD was not strongly associated with increasing age, as would be expected for an age-related neurodegenerative condition. The effects of survival bias, however, appear less for definitions 1 and 2 than what is observed when self-report alone is used to define PD.

When clinical diagnosis of PD has been unattainable, studies have relied on self-report or medications to iden-

tify PD. In particular, Chen et al. [19] have recently conducted an epidemiologic study of PD within the Atherosclerosis Risk in Communities Study, a population-based ongoing cohort study of cardiovascular outcomes among nearly 16,000 participants 45–65 years old. The authors identified 95 incident cases of PD using self-reported use of APM, ICD-9 code (332.0) and self-report of disease at the fourth study visit. They identified as potential PD those who had evidence from any of the 3 data sources, which is equivalent to definition 1 in our study. To reduce disease misclassification, they also excluded those with potential drug-induced parkinsonism as well as those who used only amantadine and/or anticholinergic drugs without additional supporting evidence of PD. Without a clinical diagnosis, the authors indirectly assessed the validity of their disease classification by evaluating the association between PD risk and baseline smoking data, and used logistic regression as their primary method of analysis. This study conducted by Chen et al. [19] as well as others that have based disease classification on self-report or medication data alone have made valuable contributions to the literature [20, 21].

Within the CHS, using 1 or more data sources to identify PD maximizes statistical power over other alternative definitions without resulting in much attenuation in risk estimates due to disease misclassification. In addition, given the study design of the CHS, using a definition of PD that incorporates a protocol to screen for disease multiple times during follow-up minimizes bias seen in a definition that relies solely on self-report. Given that disease misclassification likely still exists when using 1 or more data sources, restricting to 2 or more data sources might be considered a reasonable secondary analysis to assess the effect of disease misclassification on the estimates of association.

The CHS provides important opportunities to contribute to the PD literature given the vast amount of data already collected and immediately available. Although using multiple data sources to identify PD cannot achieve a similar degree of accuracy as clinical examination, it represents an alternative method of outcome identification in a cohort that would otherwise not be possible for PD research. Important ongoing cohort studies like the CHS can provide the setting in which rapid replication of current findings and explorations of new hypotheses for PD are possible.

Appendix 1. Medications used to screen for possible drug-induced parkinsonism

Generic name	Brand	Type	Generic name	Brand	Type
α-Methyl dopa	Aldomet, Apo-Methyl dopa, Dopamet, Novomedopa	Hypotensive	Pimozide	Orap	Neuroleptic
Amiodarone	Cordarone, Pacerone	Antiarrhythmic	Pipotiazine	Piportil	Neuroleptic
Amisulpride	Solian	Neuroleptic (antipsychotic)	Prochlorperazine	Compazine, Stemizine, Buccastem, Stemetil, Phenotil	Antiemetic
Amoxapine	Asendin	Neuroleptic	Promazine	Sparine	Neuroleptic
Aripiprazole	Abilify	Antidepressant	Quetiapine	Seroquel	Neuroleptic
Chlorpromazine	Thorazine	Neuroleptic	Remoxipride		Neuroleptic
Chlorprothixene	Tarctan	Neuroleptic	Reserpine	Harmony	Hypotensive
Cinnarizine	Stugeron	Ca channel blocker	Risperidone	Risperdal	Neuroleptic
Cleopride		Neuroleptic	Sulpride	Meresa, Sulpirid Ratiopharm, Sulpirid Neuraxpharm, Bosnyl, Dogmatil, Eglonyl, Sulpiryd	Neuroleptic
Droperidol	Inapsine	Neuroleptic, antiemetic	Tetrabenazine	Xenazine, Nitoman	Antidyskinetic
Flunarizine	Sibelium	Ca channel blocker	Thiethylperazine	Torecan	Antiemetic
Flupentixol	Fluanxol	Neuroleptic	Thioridazine	Mellaril	Neuroleptic
Fluphenazine	Prolixin	Neuroleptic	Thiothixine	Navane	Neuroleptic
Haloperidol	Haldol	Neuroleptic	Tiapride		Antianxiety
Iloperidone	Zomaril	Neuroleptic	Trifluoperazine	Eskazinyl, Eskazine, Jatroneural, Modalina, Stelazine, Terfluzine, Trifluoperaz	Neuroleptic
Levomepromazine	Nosinan, Nozinan, Levoprome	Neuroleptic	Triflupromazine	Vesprin	Neuroleptic
Lithium	Eskalith, Lithobid, Lithonate, Lithotabs	Mood stabilizer	Trimetazidine	Vastarel MR	Antianginal
Loxepine	Loxitane	Neuroleptic	Valproate, valproic acid	Depakote, Depakene	Antiepileptic
Mesoridazine	Serentil	Neuroleptic	Veralipride	Agreal, Agradil	Neuroleptic
Metaclopramide	Reglan	Antiemetic	Ziprasidone	Geodon	Neuroleptic
Molindone	Moban	Neuroleptic	Zotepine	Nipolept, Losizopilon, Lodopin, Setous	Neuroleptic
Olanzapine	Zyprexa	Neuroleptic	Zuclopenthixol	Cisordinol, Clopixol, Acuphase	Neuroleptic
Oxipertine	Integrin	Neuroleptic			
Pericyazine	Neulactil	Neuroleptic			
Perphenazine	Trilafon	Neuroleptic			

Acknowledgements

The research reported in this article was supported by the National Institute of Neurological Disorders and Stroke (R03NS057257A), the National Institute on Aging (R01 AG-023629, R01 AG-15928, R01 AG-20098 and R01 AG-027058) and the University of Pittsburgh Claude D. Pepper Older Americans Independence Center (P30 AG-024827), with additional contri-

butions from contract No. KL2 RR024154, N01-HC-85079 to N01-HC-85086, N01-HC-35129, N01 HC-15103, N01 HC-55222, N01-HC-75150, N01-HC-45133, grant No. U01 HL080295 from the National Heart, Lung and Blood Institute. E.L.T. was supported by National Heart, Lung and Blood Institute training grant T32 HL007902. A full list of principal CHS investigators and institutions can be found at <http://www.chs-nhlbi.org/pi.htm>. S.J. is a site investigator for Novartis.

References

- Gandhi S, Wood NW: Genome-wide association studies: the key to unlocking neurodegeneration? *Nat Neurosci* 2010;13:789–794.
- 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;52:276–284.
- Fried LP, Borhani NO, Enright P, Furberg CD, Gardin JM, Kronmal RA, Kuller LH, Manolio TA, Mittelmark MB, Newman A, et al: The Cardiovascular Health Study: design and rationale. *Ann Epidemiol* 1991;1:263–276.
- Swarztrauber K, Anau J, Peters D: Identifying and distinguishing cases of parkinsonism and Parkinson's disease using ICD-9 CM codes and pharmacy data. *Mov Disord* 2005;20:964–970.
- Hook EB, Regal RR: Capture-recapture methods in epidemiology: methods and limitations. *Epidemiol Rev* 1995;17:243–264.
- Bild DE, Fitzpatrick A, Fried LP, Wong ND, Haan MN, Lyles M, Bovill E, Polak JF, Schulz R: Age-related trends in cardiovascular morbidity and physical functioning in the elderly: the Cardiovascular Health Study. *J Am Geriatr Soc* 1993;41:1047–1056.

- 7 Morens DM, Davis JW, Grandinetti A, Ross GW, Popper JS, White LR: Epidemiologic observations on Parkinson's disease: incidence and mortality in a prospective study of middle-aged men. *Neurology* 1996;46:1044–1050.
- 8 Anderson DW, Rocca WA, de Rijk MC, Grigoletto F, Melcon MO, Breteler MM, Maraganore DM: Case ascertainment uncertainties in prevalence surveys of Parkinson's disease. *Mov Disord* 1998;13:626–632.
- 9 De Rijk MC, Breteler MM, Graveland GA, Ott A, Grobbee DE, van der Meche FG, Hofman A: Prevalence of Parkinson's disease in the elderly: the Rotterdam Study. *Neurology* 1995;45:2143–2146.
- 10 Wirdefeldt K, Gatz M, Bakaysa SL, Fiske A, Flensburg M, Petzinger GM, Widner H, Lew MF, Welsh M, Pedersen NL: Complete ascertainment of Parkinson disease in the Swedish twin registry. *Neurobiol Aging* 2008;29:1765–1773.
- 11 Von Campenhausen S, Bornschein B, Wick R, Botzel K, Sampaio C, Poewe W, Oertel W, Siebert U, Berger K, Dodel R: Prevalence and incidence of Parkinson's disease in Europe. *Eur Neuropsychopharmacol* 2005;15:473–490.
- 12 Mayeux R, Denaro J, Hemenegildo N, Marder K, Tang MX, Cote LJ, Stern Y: A population-based investigation of Parkinson's disease with and without dementia: relationship to age and gender. *Arch Neurol* 1992;49:492–497.
- 13 Aminoff MJ: Treatment should not be initiated too soon in Parkinson's disease. *Ann Neurol* 2006;59:562–564, discussion 564–565.
- 14 Schapira AH, Obeso J: Timing of treatment initiation in Parkinson's disease: a need for reappraisal? *Ann Neurol* 2006;59:559–562.
- 15 Lang A: When and how should treatment be started in Parkinson's disease? *Neurology* 2009;72:S39–S43.
- 16 Bower JH, Maraganore DM, McDonnell SK, Rocca WA: Incidence and distribution of parkinsonism in Olmsted County, Minnesota, 1976–1990. *Neurology* 1999;52:1214–1220.
- 17 Koepsell TD, Weiss NS: *Epidemiologic Methods: Studying the Occurrence of Illness*. New York, Oxford University Press, 2003.
- 18 Ritz B, Ascherio A, Checkoway H, Marder KS, Nelson LM, Rocca WA, Ross GW, Strickland D, Van Den Eeden SK, Gorell J: Pooled analysis of tobacco use and risk of Parkinson disease. *Arch Neurol* 2007;64:990–997.
- 19 Chen H, Mosley TH, Alonso A, Huang X: Plasma urate and Parkinson's disease in the Atherosclerosis Risk in Communities (ARIC) Study. *Am J Epidemiol* 2009;169:1064–1069.
- 20 Kamel F, Tanner C, Umbach D, Hoppin J, Alavanja M, Blair A, Comyns K, Goldman S, Korell M, Langston J, Ross G, Sandler D: Pesticide exposure and self-reported Parkinson's disease in the Agricultural Health Study. *Am J Epidemiol* 2007;165:364–374.
- 21 Hu G, Jousilahti P, Bidel S, Antikainen R, Tuomilehto J: Type 2 diabetes and the risk of Parkinson's disease. *Diabetes Care* 2007;30:842–847.