Parkinson disease (PD) is the second most common neurodegenerative disorder after Alzheimer disease. PD affects more than 4 million people worldwide and is projected to affect 9 million people by 2030 (1). The etiology for most sporadic cases remains unknown, as genetic variants account for only approximately 10% of cases (2). PD has a low incidence (8–18 per 100,000 person-years), and it is impractical to conduct a prospective, longitudinal epidemiologic study on a condition that relies on a clinical examination for a diagnosis because there is no biomarker or simple laboratory test that can confirm a PD diagnosis except autopsy. Registries for PD that cover large populations do not exist. Case-control studies are often limited by retrospective recall or ascertainment of exposures.

Large, population-based, prospective cohort studies designed to examine more common diseases might provide cost-effective ways to study PD within a longitudinal framework. Such studies have accumulated sufficient numbers of patients with incident PD to allow examinations of the associations with multiple potential risk factors and associations that develop over long periods of time. Indeed, the use of existing cohorts to study PD has resulted in numerous significant findings (2). However, most cohort studies have not implemented rigorous systematic screening for PD and have necessarily relied on secondary data sources to identify participants with PD. Common secondary sources of data used to identify PD include self-report, use of antiparkinsonian medications, and Medicare claims, respectively. The negative predictive values were all higher than 0.99. Our results highlight the limitations of using only Medicare claims data and suggest that population-based cohorts may be utilized for the study of PD determined via self-report or medication inventories while preserving a high degree of confidence in the validity of PD case identification.

Abbreviations: FHS, Framingham Heart Study; ICD-9, International Classification of Diseases, Ninth Revision; NPV, negative predictive value; PD, Parkinson disease; PPV, positive predictive value.
Ninth Revision (ICD-9) codes on forms for Medicare claims (3). Direct validation of these secondary data sources against clinical diagnostic criteria in prospective population-based cohorts is needed. Because of the probable misclassification of PD status, reported results may represent inaccurate estimates of the true underlying association between etiological factors and PD.

The Framingham Heart Study (FHS) is a longitudinal community-based cohort study of cardiovascular disease and other health conditions that was established in 1948. Since 1988, the FHS has prospectively screened and evaluated participants for PD using widely accepted clinical diagnostic criteria. Medication inventories and self-reporting of PD were also systematically implemented, and ICD-9 codes from Medicare claims were linked to participants in the FHS. In the present study, we assessed how well self-report, use of anti-parkinsonian medications, and ICD-9 codes from Medicare claims could identify PD that was confirmed by clinical diagnostic criteria in the FHS. Because efforts are underway to pool data from cohorts with self-reported or medication inventory data, we also assessed the performance of a combination of secondary data sources in identifying clinically confirmed PD.

METHODS

Study population

From 1948 to 1953, a total of 5,209 subjects (2,873 women and 2,336 men) who were 28–62 years of age and comprised two thirds of the adults in the town of Framingham, Massachusetts, were enrolled in the FHS. Subjects returned every 2 years to provide a detailed medical history and undergo a physical examination and laboratory tests. In 1971, the children of original cohort members and their spouses were invited to participate in the prospective Framingham Offspring Study. This cohort was examined 8 years after baseline and every 4 years thereafter. Descriptions of the FHS design and implementation have been published previously (4, 5). The Boston University Medical Center/Boston Medical Center Institutional Review Board approved the use of human subjects for this study.

Identification of PD in FHS

Since 1988, PD cases have been identified through medical records, hospital surveillance, clinic examinations conducted by an FHS physician, and review of medical charts. The final diagnosis of PD is made by a panel of neurologists using the UK Brain Bank Criteria (6). Before the establishment of the UK Brain Bank Criteria, PD monitoring was done using the Boston University Criteria, which were based on published criteria from the Deprenyl and Tocopherol Antioxidative Therapy of Parkinsonism (DATATOP) Study (7), the largest and longest prospective controlled study of therapeutic interventions for PD. Details of both criteria are in Web Appendices 1 and 2, which are available at http://aje.oxfordjournals.org/ (6, 7).

Secondary data sources for PD available in the FHS

We used data from 2001 to 2012 because this was the earliest time period during which information about the 3 secondary data sources for PD identification were available. Namely, the FHS specifically queried about PD diagnosis via self-report, performed a medication inventory that included the use of antiparkinsonian medications, and collected claims data from the Centers for Medicare and Medicaid Services. For the present study, we assessed how well self-report, use of antiparkinsonian medications, and ICD-9 codes from Medicare claims could identify PD that was confirmed by clinical diagnostic criteria in the FHS. Because efforts are underway to pool data from cohorts with self-reported or medication inventory data, we also assessed the performance of a combination of secondary data sources in identifying clinically confirmed PD.

Self-reported PD. We used health history updates from 2001–2012 to assess self-reported PD. Between in-person visits or if a participant did not attend in-person examinations, health history updates were completed by mail or telephone to capture major medical events. These updates included 2 questions that asked participants whether they had been diagnosed with PD and whether they had ever been hospitalized because of PD. If they responded yes to either question, they were considered to be a PD case.

Use of antiparkinsonian medications. We used examinations from 2004–2005 for the original cohort and 2005–2008 for the offspring cohort to assess the use of antiparkinsonian medications. Participants were asked to bring in all medications taken at home and that information was then recorded. Medications used to identify PD cases included those with an Anatomical Therapeutic Chemical Classification code category of N04, which includes levodopa-containing compounds, dopamine agonists, anti-cholinergics, and amantadine. We also screened for drug-induced parkinsonism by using Anatomical Therapeutic Chemical Classification codes N05A (antipsychotics), A03FA06 (clozapine), A03FA01 (metoclopramide), N06AA17 (amoxapine), R06AD03 (thiethylperazine), C01EB15 (trimeperazine), C02AA02 (reserpine), N07XX06 (tetrabenazine), N07CA02 (cinamazoline), N07CA03 (flunarizine), C01BD01 (amiodarone), C02AB (methylodopa), and N03AG01 (valproic acid). A person initially identified as a PD case who had in his or her medical history an Anatomical Therapeutic Chemical Classification code for potentially drug-induced parkinsonism was reclassified as a noncase.

Medicare claims. We used Medicare claims with ICD-9 code 332.0 to identify PD cases from records obtained from inpatient hospitalizations, outpatient visits, skilled nursing facilities, hospice care, and post-acute home health care. Participants were included if they participated in Medicare fee for service during any time from 2001 to 2008. We identified PD cases as participants with at least 1, at least 2, or at least 3 claims that included code 332.0.

Evaluating the performance of secondary data sources in identifying PD

We compared secondary data sources used to identify PD (self-report; use of antiparkinsonian medications; Medicare claims; people who both reported having PD and used antiparkinsonian medications; and a combination of people who either self-reported or used antiparkinsonian medications) against the Boston University or UK Brain Bank Criteria. Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated using a 2 × 2 table in which the rows were “yes” or “no” for PD according to the secondary data source being evaluated and the columns were “yes” or “no” for PD according to clinical diagnostic criteria.
RESULTS

Participant characteristics are summarized in Table 1. Self-reported data were available for 2,411 participants. Using clinical diagnostic criteria, 28 cases of PD were identified among these participants. We identified 26 PD cases using self-report. Against clinical diagnostic criteria, self-report had a sensitivity of 0.929, a specificity of 1.0, a PPV of 1.0, and an NPV of 0.999.

Medication inventory data were available for 3,180 participants. Using clinical diagnostic criteria, 23 cases of PD were identified among these participants. When subjects who took only antiparkinsonian medications were considered, 21 cases were accurately identified as PD cases, yielding a sensitivity of 0.913, a specificity of 1.0, a PPV of 1.0, and an NPV of 0.999. When we reclassified some PD cases as noncases because of exposure to medications that might cause drug-induced parkinsonism, we observed a sensitivity of 0.739, a specificity of 1.0, a PPV of 1.0, and an NPV of 0.998.

Claims data were available for 1,949 participants, of whom 22 were diagnosed with PD using clinical criteria. We evaluated how well the presence of 1, 2, or 3 Medicare claims with ICD-9 code 332.0 identified PD. The PPVs were 0.333, 0.483, and 0.50, respectively. Detailed results can be found in Table 2.

DISCUSSION

We observed that in the FHS, self-reporting or data on use of antiparkinsonian medications could identify PD cases with excellent accuracy when compared with clinical diagnostic criteria (PPV = 100%), whereas ICD-9 codes on Medicare claims forms did not perform well (PPV = 33%–50%). To our knowledge, this is the first report in which the identification of PD cases by secondary data sources has been compared with that by clinical diagnostic criteria in a longitudinal community-based cohort study of cardiovascular disease and other health conditions.
In epidemiologic studies of PD, a critical issue is the certainty of identified PD cases. Autopsy aside, the most accurate method of identification is to use clinical criteria, which encompass neurological signs and responses to medication after the exclusion of other causes of parkinsonism. These comprise the UK Brain Bank Criteria and Boston University Criteria, both of which were used in the FHS (6, 7). PD cases that are diagnosed using such criteria are confirmed at autopsy in 80%–90% of cases (8). We lacked a sufficient number of autopsies to pathologically validate PD diagnoses in this cohort. However, most longitudinal community-based cohort studies do not systematically apply PD clinical criteria, which necessitates the use of secondary data sources. These often include self-report, use of antiparkinsonian medication, or Medicare claims (3, 9, 10). We found that when compared with clinical diagnostic criteria, self-report and use of antiparkinsonian medications individually identified PD cases with more than 90% sensitivity and virtually 100% specificity, with a PPV of 100%. When we combined data on those who either self-reported or used antiparkinsonian medications, we observed similar results. Medicare claims data, although still highly specific, had just over 60% sensitivity and a PPV of 50% for participants with 3 or more claims with codes indicating PD.

We know of no other efforts to validate these secondary data sources against clinical diagnostic criteria in a longitudinal community-based cohort study. In a study utilizing an administrative database from the Veteran’s Health Administration, medication inventories and ICD-9 codes were validated against information extracted from medical records analogous to clinical diagnostic criteria. ICD-9 codes (332.0, 332.1, and 333.0) had a sensitivity of 18.7%, a specificity of 99.9%, and PPV of 81%. Use of antiparkinsonian medications (dopamine agonist or levodopa/carbidopa) had a sensitivity of 34.6%, a specificity of 99.6%, and a PPV of 60% (11). In another study in the United Kingdom, Meara et al. (12) assessed the accuracy of medication inventories in identifying PD cases using a prescription database from general practices. Using the UK Brain Bank criteria, PD was confirmed in 53% of subjects taking antiparkinsonian medications. Our findings of better performance of secondary data sources in identifying PD might be due to the fact that instead of relying on administrative databases in the general or veteran populations, we used data systematically gathered directly from participants as part of longitudinal cohort study. When using secondary data sources to identify PD cases, we were more stringent with ICD-9 codes (using only 332.0) and included more antiparkinsonian medications because more medication options for PD may have been available during the time of this cohort. Educational level may have contributed to the performance of self-reporting because the majority of participants in the present study had at least some college education (13). Interestingly, screening for drug-induced parkinsonism resulted in slightly worse performance of medications in identifying PD cases, possibly because medications that might result in parkinsonism may nonetheless be used to treat PD-related comorbidities (e.g., neuroleptics to treat psychosis). By attempting to exclude potentially drug-induced parkinsonism, we could have inadvertently excluded cases with PD. Therefore, we did not screen for drug-induced parkinsonism when combining use of antiparkinsonian medications with self-report to identify PD cases.

These methodological differences suggest limitations to the applicability of our results. Our results likely do not reflect the validity of secondary data sources in administrative databases for the general population (e.g., Center for Medicaid and Medicare Services, pharmacy inventories). Involvement in the FHS may bias participants to be more likely to recall historical medical information. Participants in the Boston metropolitan area may also receive more frequent medical care than subjects in other regions of the country, although

Table 2. Performance of Secondary Data Sources in Identifying Parkinson Disease Compared With Clinical Diagnostic Criteria, Framingham Heart Study, 2001–2012

<table>
<thead>
<tr>
<th>Source for Identification of Parkinson Disease</th>
<th>Sensitivity</th>
<th>95% CI</th>
<th>Specificity</th>
<th>95% CI</th>
<th>Negative Predictive Value</th>
<th>95% CI</th>
<th>Positive Predictive Value</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-report</td>
<td>0.929</td>
<td>0.765, 0.991</td>
<td>1.0</td>
<td>0.999, 1.0</td>
<td>0.999</td>
<td>0.997, 1.0</td>
<td>1.0</td>
<td>0.868, 1.0</td>
</tr>
<tr>
<td>Use of antiparkinsonian medication(s)</td>
<td>0.913</td>
<td>0.720, 0.989</td>
<td>1.0</td>
<td>0.999, 1.0</td>
<td>0.999</td>
<td>0.998, 1.0</td>
<td>1.0</td>
<td>0.839, 1.0</td>
</tr>
<tr>
<td>Use of antiparkinsonian medication(s) with exclusionsa</td>
<td>0.739</td>
<td>0.516, 0.898</td>
<td>1.0</td>
<td>0.999, 1.0</td>
<td>0.998</td>
<td>0.996, 0.999</td>
<td>1.0</td>
<td>0.805, 1.0</td>
</tr>
<tr>
<td>Self-report or use of antiparkinsonian medication(s)b</td>
<td>0.889</td>
<td>0.739, 0.969</td>
<td>1.0</td>
<td>0.999, 1.0</td>
<td>0.999</td>
<td>0.997, 1.0</td>
<td>1.0</td>
<td>0.891, 1.0</td>
</tr>
<tr>
<td>At least 1 Medicare claim</td>
<td>0.727</td>
<td>0.498, 0.893</td>
<td>0.983</td>
<td>0.977, 0.989</td>
<td>0.997</td>
<td>0.993, 0.999</td>
<td>0.333</td>
<td>0.204, 0.484</td>
</tr>
<tr>
<td>At least 2 Medicare claims</td>
<td>0.636</td>
<td>0.407, 0.828</td>
<td>0.992</td>
<td>0.987, 0.996</td>
<td>0.996</td>
<td>0.992, 0.998</td>
<td>0.483</td>
<td>0.295, 0.675</td>
</tr>
<tr>
<td>At least 3 Medicare claims</td>
<td>0.636</td>
<td>0.407, 0.828</td>
<td>0.993</td>
<td>0.988, 0.996</td>
<td>0.996</td>
<td>0.992, 0.998</td>
<td>0.50</td>
<td>0.307, 0.694</td>
</tr>
</tbody>
</table>

Abbreviation: CI, confidence interval.

a Cases with potentially drug-induced parkinsonism were excluded.
b Cases with potentially drug-induced parkinsonism were not excluded.
participants in the FHS offspring cohort are widespread throughout the country. However, our results are applicable to large population-based prospective cohort studies that, although not designed to examine PD, have nonetheless been used to examine multiple potential risk factors and their interactions. These include the Cardiovascular Health Study (14–16), the Atherosclerosis in Communities Study (17), the Honolulu Asia Aging Study (18–22), the Leisure World Cohort Study (23), the Nurses’ Health Study, and the Health Professionals Follow-up Study (24–27).

Many of the initial epidemiologic studies of PD were limited by sample size or methodology that compromised interpretation of results, and the present study is no exception. There are efforts underway to combine data from several prospective studies that have enough incident PD cases and follow-up time to confirm diagnoses. Our results highlight the limitations of using only Medicare claims data and suggest that population-based cohorts could be utilized for the study of PD in a prospective way by using self-reports or medication inventories while preserving a high degree of confidence in the validity of PD case identification, thereby opening new avenues of research.

ACKNOWLEDGMENTS

Author affiliations: Department of Neurology, University of Pittsburgh, Pittsburgh, Pennsylvania (Samay Jain); Department of Neurology, Boston University, Boston, Massachusetts (Jayandra Himali, Alexa Beiser, Sudha Seshadri, Samuel A. Frank); Precision Health Economics, San Francisco, California (Thanh G. N. Ton); The Framingham Study, Boston University, Framingham, Massachusetts (Margaret Kelly-Hayes), Department of Biostatistics, University of Washington, Seattle, Washington (Mary Lou Biggs), Department of Epidemiology, University of Washington, Seattle, Washington (Joseph A. C. Delaney); and Department of Epidemiology, University of Pittsburgh, Pittsburgh, Pennsylvania (Caterina Rosano).

All authors contributed equally to this work.

Funding for this study was provided by the American Parkinson Disease Association, Inc, The National Institute of Neurological Disorders and Stroke of the National Institutes of Health (grant 1K23 NS070867), the National Heart, Lung, and Blood Institute of the National Institutes of Health, and Boston University School of Medicine. This work was also supported by the National Heart, Lung and Blood Institute’s Framingham Heart Study (contract No. N01-HC-25195).

Conflict of interest: none declared.

REFERENCES


