SUPPLEMENTARY MATERIAL

Data source

We used longitudinal routine claims data from the years 2004–2015 collected by the largest German statutory health insurance company, the *Allgemeine Ortskrankenkasse* (AOK). In Germany, about 70 million people are insured through statutory programs, one third of whom are members of the AOK. The AOK covers more than 50% of the population of higher ages.[1] The data holder provided us with an age-stratified random sample of 250,000 individuals aged 50 years and older, which was about 2% of all individuals insured in the AOK. This sample was drawn in the first quarter of 2004 and enabled us to follow individuals through the end of 2015. The AOK sample is almost representative of the overall population in Germany in terms of gender but not of age, because individuals from the sample are somewhat older than average. The data contain types of diagnoses by the German modification of the International Classification of Diseases, Tenth Revision (ICD-10-GM), all treatments in the inpatient and outpatient sector which are relevant for billing, drug prescriptions filled in the outpatient sector according to the Anatomical Therapeutic Chemical Classification System, benefits from the long-term-care-insurance, and date of death.[2] This information is reported quarterly and covers every insured individual, regardless of actual utilization. Data access was legally approved by the Scientific Institute of the AOK. This study is based on anonymized administrative claims data which never involved patients directly, therefore individual patients cannot be identified and the analyses presented do not affect those patients whose anonymized records were used.

Analysis sample and outcome measure

The analysis sample consists of individuals aged 50 years and older who had not received any diagnosis of PD in the year 2004. PD was identified based on ICD-10-GM code G20. We developed internal validation strategies to rule out false positive diagnoses. First, we included only those diagnoses internally marked as "verified" in the outpatient sector and as "discharge diagnosis" or "secondary diagnosis" in the inpatient sector. Second, only valid PD diagnoses were considered; a diagnosis was considered to be valid, if a patient had also received a confirmative PD diagnosis (ICD-10 code for PD) during the observation period. If this was the case, the individual was identified as a PD patient starting at the time the first diagnosis was made. Third, PD cases in which the last diagnosis in our longitudinal data was atypical parkinsonism (G21, G22) or essential tremor (G25.0) were not considered valid.[2] After excluding prevalent PD cases we arrived at an analysis sample of 228,485 individuals, whom we followed from January 1, 2005 until one of the following occurred: a valid PD diagnosis, death, exit from the AOK insurance, or December 31, 2015 (end of the study period).

Exposure and comorbidities

In the primary analysis, we considered the impact of GIIs according to ICD-10-GM codes A00–A09 (intestinal infectious diseases). Diagnoses of GIIs had to be marked as "verified" in the outpatient sector and as "discharge diagnosis" or "secondary diagnosis" in the inpatient sector. We controlled for gender and age in five-year age groups (from 50 to 95+), and for common morbidities and risk factors for PD in old age: diabetes mellitus (E10–E14), cerebrovascular diseases (G45, G46, H34.0, I60, I69), hypertension (I10–

I15), ischemic heart diseases (I20–I25), hypercholesterolemia (E78.0), chronic obstructive pulmonary disease (COPD, J44), and intracranial injury (S06). In the bias analysis we controlled for smoking related cancers (SRC, C00–C14, C18–C20, C25, C30–C34, C64–C67) instead of COPD. Exposure and comorbidity variables were defined as time-varying dummy variables, receiving a value of one from the first time a specific diagnosis was made until the outcome or censoring. In the secondary analysis, we divided GIIs according to their origin: BIIs (A00-A05), VIIs (A08), and IGCUs (A09). We omitted amoebiasis (A06) and other protozoal intestinal diseases (A07) due the low number of cases.

Statistical analysis

We showed the cumulative risk of developing PD stratified by the presence of the diseases of interest.[3] Differences were assessed using a log-rank test. We used Cox proportional hazards models to examine whether the incidence of PD was associated with GIIs, adjusting for the covariates mentioned above. We examined the Schoenfeld residuals and found that the proportional hazards assumption was satisfied. Analysis time was defined by the number of months since January 1, 2005. For individuals transitioning to PD, the analysis time was set to the middle of the quarter. Deaths were assigned to the middle of the month of death. Furthermore, we performed subgroup analyses and calculated interaction effects for gender and age. To solidify our results we performed the following sensitivity analyses: (1) applying different disease-free periods prior to the first incident PD diagnosis (two, three, or four years), (2) PD identification based on one PD diagnosis (without the algorithm of repeated diagnoses), (3) exposure lagging by 4 and 5

3

years for each variable except for gender and age, (4) repeating the primary analysis

including appendicitis (K35–K37) as a negative exposure control variable, (5) excluding

individuals with prodromes of PD in 2004 (other sleep disorders, G47.8; constipation,

K59.0; anosmia, R43.0) in the year 2004, and (6) conducting a competing risk model as

proposed by Fine and Grey.[4] All analyses were conducted using the statistical software

Stata 12.1 (StataCorp LLC).

Limitations

We did not have information on the exposure to GIIs prior to the age of 50, the severity

of a GII, of individuals who did not seek medical attention, or of lifestyle (tobacco or

coffee consumption) and occupational (working in the agricultural sector or heavy

industry) risk factors known to be associated with PD and could not consider these

factors in our analysis. However, we considered the effect of smoking by taking

conditions into account which are strongly related to smoking. GIIs were still associated

with PD for individuals with or without COPD or smoking related cancers (table 2).

Supplementary references

1   Schulz A, Doblhammer G. Aktueller und zukünftiger Krankenbestand von Demenz in Deutschland auf Basis der Routinedaten der AOK. In: Günster C, Klose J, Schmacke N, eds. Versorgungs-Report 2012 Schwerpunkt: Gesundheit im Alter. Stuttgart: Schattauer, 2012:161-75.
2   Nerius M, Fink A, Doblhammer G. Parkinson's disease in Germany: prevalence and incidence based on health claims data. Acta Neurol Scand 2017;**136**:386-92.
3   Snapinn S, Jiang Q, Iglewicz B. Illustrating the Impact of a Time-Varying Covariate With an Extended Kaplan-Meier Estimator. The American Statistician 2005;**59**:301-7.
4   Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. J Am Stat Assoc 1999;**94**:496-509.

Supplementary table S1. Sensitivity analyses showing adjusted[a] hazard ratios for association between GIIs and risk of PD

Sensitivity Analysis	Exposure	HR	95% CI
Varying disease-free periods, in years			
2[b]	GIIs	1.40	1.31–1.50
3[c]	GIIs	1.43	1.33–1.53
4[d]	GIIs	1.42	1.32–1.52
PD incidence based on one PD diagnosis[e]	GIIs	1.41	1.33–1.50
Exposure lagging, in years			
4	GIIs	1.18	1.06–1.32
5	GIIs	1.18	1.04–1.34
Primary analysis including a negative exposure control	Appendicitis	1.18	0.93–1.50
Without patients with prodromes of PD in 2004[f]	GIIs	1.40	1.31–1.50
Competing risk analysis (taking the risk of death into account)[g]	GIIs	1.31	1.22–1.40
a Adjusted for gender, age, diabetes mellitus, cerebrovascular diseases, hypertension, ischemic heart diseases, hypercholesterolemia, chronic obstructive pulmonary disease, and intracranial injury			
b N = 217,980; PD cases = 5,370			
c N = 208,298; PD cases = 4,681			
d N = 197,786; PD cases = 4,105			
e N = 228,221; PD cases = 7,783			
f N = 221,682; PD cases = 5,878			
g N = 228,485; PD cases = 5,884; HR are given here as subdistribution HR			
GIIs, gastrointestinal infections; CI, confidence interval			

Supplementary table S2. Adjusted[a] hazard ratio for association between BII, VII, IGCU, and risk of PD

Risk Factor[b]	HR	95% CI
BIIs	1.30	1.12–1.50
VIIs	1.31	1.14–1.50
IGCUs	1.34	1.24–1.44
N = 228,485; PD cases = 6,195		
a Adjusted for gender, age, diabetes mellitus, cerebrovascular diseases, hypertension, ischemic heart diseases, hypercholesterolemia, chronic obstructive pulmonary disease, and intracranial injury		
b Reference group: not exposed to BIIs, VIIs, or IGCUs		
BIIs, bacterial intestinal infections; VIIs, viral intestinal infections; IGCU, infectious gastroenteritis and colitis of unspecified origin; HR, hazard ratio; CI, confidence interval		
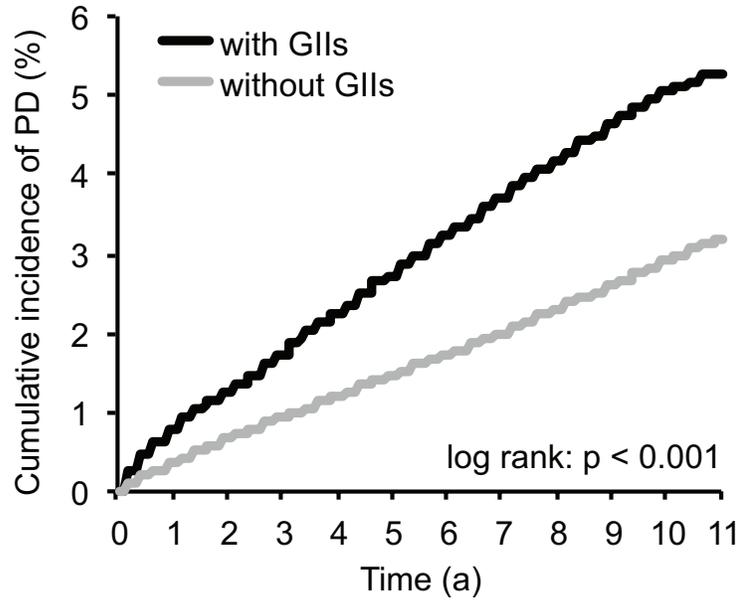
Supplementary figure S1 Cumulative incidence of PD for exposed and non-exposed individuals

The cumulative incidence of PD was increased in patients with GIIs in comparison to the control group not exposed to GIIs.