NEUROLOGY/2018/891457

Driving impairment and crash risk in Parkinson's disease: A systematic review and meta-analysis

Trevor Thompson, PhD[a], Damian Poulter, PhD[a], Clare Miles PhD[a], Marco Solmi MD[b], Nicola Veronese MD[c], André F. Carvalho MD PhD[de], Brendon Stubbs PhD[fg], Ergun Y. Uc MD[hi]

[a]Faculty of Education and Health, University of Greenwich, London SE9 2UG, UK

[b]Department of Neuroscience, University of Padova, Padova, Italy

[c]National Research Council, Neurosciences Department, Aging Branch, Padova, Italy

[d]Department of Psychiatry, University of Toronto, ON, Canada

[e]Centre for Addiction & Mental Health (CAMH), Toronto, ON, Canada

[f]Physiotherapy Department, South London and Maudsley NHS Foundation Trust, London SE5 8AZ, UK

[g]Department of Psychological Medicine, King's College, De Crespigny Park, London SE5 8AF, UK

[h]Department of Neurology, Carver College of Medicine, University of Iowa, Iowa City, IA, USA

[i]Neurology Service, Iowa City Veterans Affairs Medical Center, Iowa City, IA, USA

Title character count: 95

Number of references: 29

Number of tables: 0

Number of figures: 5

Word count abstract: 190

Word count manuscript (introduction, methods, results, discussion): 4844

Supplementary material: Reference e-1 (studies included in meta-analysis), Appendix e-1 (validity criteria), Table e-1 (study characteristics).

Corresponding Author

Dr Trevor Thompson, t.thompson@gre.ac.uk, +44 208 3319632

Statistical analysis conducted by corresponding author

Search Terms

Parkinson's disease/Parkinsonism; fitness to drive; meta-analysis

Author Contributions

Trevor Thompson – study design and conceptualization, analysis and interpretation, drafting manuscript

Damian Poulter – study selection, data extraction, quality assessment, drafting of manuscript

Clare Miles – study selection, data extraction, quality assessment

Marco Solmi - drafting the manuscript, interpretation of data

Nicola Veronese - critical revision of manuscript for intellectual content

André F. Carvalho - critical revision of manuscript for intellectual content

Brendon Stubbs - interpretation of data, critical revision of manuscript for intellectual content

Ergun Y. Uc – interpretation of data, critical revision of manuscript for intellectual content

Author Disclosures

All authors report no disclosures.

This study was funded by an internal grant awarded to the first and second authors by the University of Greenwich.

ABSTRACT

OBJECTIVES: To provide the best possible evidence base for guiding driving decisions in Parkinson's disease (PD), we performed a meta-analysis comparing PD patients to healthy controls (HCs) on naturalistic, on-the-road and simulator driving outcomes. **METHODS:** Seven major databases were systematically searched (to 01/2018) for studies comparing PD patients with HCs on overall driving performance, with data analysed using random-effects meta-analysis. **RESULTS:** Fifty studies comprising 5,410 participants (PD=1,955, HC=3,455) met eligibility criteria. Analysis found the odds of on-the-road test failure were 6.16 (95% CI, 3.79-10.03) times higher and the odds of simulator crashes 2.63 (95% CI,1.64-4.22) times higher for people with PD, with poorer overall driving ratings also observed (*SMDs*=0.50 to 0.67). However, self-reported real-life crash involvement did not differ between people with PD and HCs (Odds Ratio=0.84, 95% CI, 0.57-1.23, p=.38). Findings remained unchanged after accounting for any differences in age, sex and driving exposure, and no moderating influence of disease severity was found. **CONCLUSIONS:** Our findings provide persuasive evidence for substantive driving impairment in PD, but offer little support for mandated PD-specific relicensure based on self-reported crash data alone, and highlight the need for objective measures of crash involvement.

Keywords: Parkinson's disease; neurodegeneration; driving; fitness to drive; meta-analysis; crashes.

1 INTRODUCTION

Parkinson's disease (PD) symptoms can include motor instability, increased response time, attentional deficits, visual impairment, daytime sleepiness and medication-exacerbated sleep attacks.[4] As all of these factors may compromise capacity to drive, this has caused concerns over the driving safety of people with PD and led to recent policy debate over the need for mandatory based reevaluation of fitness to drive in those with degenerative diseases.[1]

As cessation of driving can compromise independence and lead to isolation, such decisions should be informed by the best possible evidence base. Unfortunately, research findings are somewhat inconsistent, likely due to considerable variation in study methodology, and there is currently no clear overall picture of driving safety risk in PD, or whether this risk is influenced by disease stage or medication use.

While narrative reviews[5,11] have been helpful in summarizing research evidence on driving impairment, there is a pressing need for a quantitative synthesis of all available empirical data that accounts for study heterogeneity, to provide the best possible evidence for guiding driving decisions in PD. The current study is the first meta-analysis to compare individuals with PD and healthy controls on driving performance measures from on-the-road, driving simulator and real-life crash studies and aims to establish: (1) precise estimates of the magnitude of any driving impairment; (2) the influence of severity stage and medication dosage on impairment; and (3) whether PD is linked to both poorer performance on driving assessments and an inflated real-life crash risk.

2 METHODS

This systematic review and meta-analysis was conducted in accordance with PRISMA guidelines,[14] and followed an *a priori* but unpublished protocol available upon request.

2.1 Eligibility Criteria

Inclusion criteria were the use of: (1) a group with a diagnosis of primary (idiopathic) Parkinson's disease (PD); (2) a neurologically healthy control (HC) group, (3) at least one quantitative measure of driving performance from real-world driving data, on-the-road tests or driving simulator assessments.

Three types of driving assessment were included to provide a comprehensive evaluation of driving performance. Real-world, naturalistic assessment typically provides crash data and thus a direct measure of safety risk. On-the-road assessments do not offer direct measures of crash risk, but do provide an accepted standard measure of fitness to drive, and expose both groups to the same driving challenges thus circumventing any effect of compensatory strategies (e.g. avoiding night driving) that can mask genuine driving deficits. High fidelity simulators are less naturalistic, but can use a range of driving conditions and assess driving response to situations with high crash potential.

Studies were excluded if outcomes were measures of driving strategy[13] only (e.g. avoiding rush hour), as the focus of this study is the involuntary degradation of functional driving ability in PD (rather than voluntary driving behaviours). We also excluded studies assessing performance

outside of a driving context (e.g. Stroop response time), given the inconsistent link between such neuropsychological assessments and actual road performance.[2]

2.2 Driving outcomes

Primary outcomes were (a) crashes (e.g. crash involvement (yes/no) or number of crashes), and (b) overall driving competence (e.g. test failure or overall performance ratings).

We examined only aggregate or overall driving competence outcomes, as separate analysis of all individual driving components that typically comprise overall outcome scores (e.g. lane deviation, road excursions) would produce excessive Type-I errors and multiple analyses varying in their sample sizes and methodological characteristics. We excluded (a) absolute speed as its relationship with driving competence is complex and dependent upon driving context[23] and (b) use of gears, route recall or pedal (e.g. brake) pressure, as these were not considered to be direct measures of driving competence.

2.3 Information Sources

PubMed, Embase, Transport Research International Documentation, CENTRAL, CINAHL Plus with Full Text, PsycINFO and Web of Science databases were searched from database inception until 25th January, 2018.

2.4 Search terms

Searches of database subject headings (where provided) and free text words of all fields were conducted to identify the largest possible pool of potentially eligible studies. PubMed search

terms were as follows, with equivalent strings constructed for other databases: ("automobile driving" [mh] OR driving[tw] or automobile[tw] or simulator[tw] or simulate[tw] or road-test[tw]) AND ("Parkinson disease"[mh] OR Parkinson*[tw]).

No language restrictions were imposed. Searches were filtered with the Cochrane sensitivity-maximising human filter. The search strategy was augmented through hand searching of relevant reviews and included articles.

2.5 Study selection

After removal of duplicates, two independent reviewers (CM, DP) screened titles/abstracts (stage 1), followed by full-text review of any resultant potentially eligible articles (stage 2). If multiple articles presented the same outcome data, only one article was retained (that with the clearest reporting or largest sample size), unless other articles also reported additional unique outcomes (in which case only unique outcomes were used). Ten corresponding authors of eligible articles were contacted across an eight-week period to clarify procedures or request additional data, with a single follow-up email sent in the event of non-response. Six authors (see acknowledgements section) replied and provided query resolutions. Any disagreements at any stage of the selection process were resolved through discussion with a third reviewer (TT) to reach a final list of retained articles.

2.6 Quality of evidence

Two reviewers (CM, DP) independently rated the quality of supporting evidence on an 18-item validity scale (Appendix e-1) based on Cochrane criteria and PRISMA recommendations and adapted from our previous work[19] for the current topic. Criteria were scored as unmet if their endorsement could not be explicitly demonstrated (i.e. criteria were unmet or unreported).

2.7 Standard protocol approvals, registrations, and patient consents

No additional ethics approval was required for this meta-analysis.

2.8 Data Extraction

Extraction and coding of study data was performed by one reviewer (CM) on a standardized form [19] with 100% of extracted data checked for accuracy by one reviewer (TT). The following data were extracted: (1) sociodemographic variables; (2) For PD: mean disease duration (years), symptom severity, functional state (ON/OFF) during driving/disease assessment, diagnostic criteria, cognitive impairment, usual treatment; (3) lifetime driving experience (e.g. years licence held) and current driving exposure (e.g. miles driven per week); (4) assessment model (real-life, on-the-road, driving simulator) and (5) outcomes. Where study data allowed computation of multiple effect sizes (e.g. across repeated trials), all such data were extracted.

Several extraction decisions were made: (1) where experimental manipulations were used during driving assessment (k=2), baseline data only was extracted; (2) for one research group reporting total and individual driving scores, we extracted the latter due to ambiguity in the former; (3) for one prospective study[25] that reported raw crash data and crash data adjusted for different group follow-up periods and other potential confounds, we included adjusted statistics only; (4) for one study reporting test outcome as safe/unsafe, we recoded this as pass/fail for consistency; (5) a few studies reported median outcomes which we used as an approximation of the mean. The impact of decisions 4-5 was examined in sensitivity analysis.

2.9 Effect size

For continuous outcomes, the standardized mean difference (*SMD*) for PD vs. HCs was computed using Hedges' *g* formula, where .20, .50 and .80 can be roughly translated as small, medium and large effects.[3] For categorical outcomes (e.g. crash history yes/no), odds ratios (*OR*) were computed (log *OR* were used in the meta-analysis and subsequently converted back to *OR*).

Effect sizes were coded so that positive values (for both *SMD* and log *OR*) indicated driving impairment in PD (e.g. more driving errors, higher crash rate).

2.10 Meta-analysis

Group differences in driving outcomes were estimated using a random-effects model given anticipated heterogeneity in effect size. Robust variance estimation (RVE)[20] was used to model dependency amongst multiple effect sizes from within a study (when total and subscore effect sizes were reported we included only the former). As the same patient groups were used across several studies (identified by contacting authors and examining article text/data), we also reran analysis including patient group as a second-order hierarchical variable to account for sample dependency.

Meta-analysis was performed on data collapsed across real-life, on-the-road and simulator models. Assessment model was also added as a dummy-coded moderator, and if significant, separate meta-analyses were performed for each model. Analyses were not performed if adjusted $df<4$ (reflecting insufficient independent studies) as such model estimates are unreliable.[20]

2.11 Meta-regression

If moderate or greater inconsistency (I^2>50%[10]) in effect size emerged, meta-regression analyses were performed to identify potential sources of effect size variation.

The primary moderator was disease severity (symptom severity/disease duration), as increased neurodegeneration causes greater impairment. Secondary moderators consisted of medication dosage (L-dopa equivalent), study sex ratio, age, and lifetime driving experience.

If PD and HC groups differed on potential confounds of driving experience, age, and sex, which could explain group differences in driving outcomes, we reran analyses entering standardized difference scores on these variables as covariates. For real-life crash data, where miles driven cannot be experimentally fixed to be equivalent across groups, we included current driving exposure (e.g. mileage) differences over the assessed crash period as a covariate.

2.12 Publication bias

Funnel plots of mean study effect sizes against standard errors were examined for Indicators of possible publication bias, with any asymmetry due to a lack of small studies with small effects tested statistically.[8,16]

2.13 Analysis software

All analyses were performed using the *metafor*[28] and *robumeta*[9] packages in R.

3 RESULTS

3.1 Study selection and data characteristics

3.1.1 *Study inclusion*

3,962 unique hits were identified through database searches, with initial screening of titles/abstracts identifying 226 potentially eligible articles, further reduced to 54 eligible articles following full text review (Figure 1). Of these 54 articles, 50 provided sufficient quantitative data for meta-analysis consisting of a total *N*=5,410 (PD *n*=1,955, HC *n*=3,455). The 50 articles represented 43 journal publications, 4 conference abstracts, 2 PhD theses and 1 transport report. Detailed individual study characteristics are presented in Table e-1 and summarized below.

3.1.2 *Participant characteristics*

37 unique participant samples were identified as being used across the 50 studies (often with different driving models/outcomes) with an aggregate *N*=4,149 (*n*=1,576 PDs, *n*=2,573 HCs). Mean age (reported for *k*=36 of 37 samples) was similar for PD (*M*=66.6 yrs, SD=4.7, range=54-75) and HCs (*M*=65.6 yrs, SD=6.8, range=48-77), although the mean proportion of males (*k*=31) was higher for PD (*M*=74%, SD=17.6, range=12-100) compared to HCs (*M*=62%, SD=18.1, range=14-100).

Where lifetime driving experience had been reported (*k*=24), PDs and HCs were similar in mean years of driving experience (*k*=15; PD *M*=44.2 yrs, HC *M*=42.9 yrs), and when other classifications (*k*=9) were used, such as the proportion of participants with >10 years' driving experience. Current driving exposure (*k*=24) was assessed in a variety of metrics with most studies reporting similar values for both groups, e.g. number of days driven per week (*k*=8; PD *M*=5.2, HC *M*=5.5).

Samples were recruited from the following locations: USA (k=15), Australia (k=7), Greece (k=3), Canada (k=2), France (k=2), Germany (k=2), Belgium (k=1), Denmark (k=1), Finland (k=1), Netherlands (k=1), Thailand (k=1) and UK (k=1).

3.1.3 *Parkinson's disease severity and medication*

Mean disease duration (reported for *k*=22 of 37 samples) was 6.7 yrs (*SD*=1.6, range=3.5-10.9) with symptom severity assessed during the ON state with the Hoehn-Yahr (*k*=24; *M*=2.1, *SD*=0.4, range=1.0-4.0) and/or the UPDRS-III (*k*=20; *M*=20.9, *SD*=5.6, range=11.7-30.1) scales. PD diagnosis was confirmed with clinician assessment (*k*=13) or no diagnostic detail was provided (*k*=24). All studies reported regular use of antiparkinson medication by varying sample proportions, with *k*=21 studies providing exact figures (*M*=96%, *SD*=9.0).

A minimum level of cognitive functioning in PD for study inclusion was specified by 14 studies, with MMSE>=24 being the most common criterion. The mean MMSE score (*k*=21) was 28.2. Driving assessments were administered during ON states (*k*=22), with 15 studies not reporting ON/OFF state.

3.2 Assessment outcomes

3.2.1 *Real-life driving data*

15 studies provided naturalistic driving data, with 13 providing self-reported crash statistics for 2,143 (PD *n*=873, HC *n*=1,270) participants, and 2 studies providing speed and response times using in-car monitoring technology. Crash statistics were (a) the proportion of each group involved

in a crash (*k*=8), and (b) number of crashes (*k*=5) experienced over a mean period of 3.0 years (SD=2.5, range=1-10 years). Crashes were typically treated as driving events causing injury or damage to the vehicle. Most assessments consisted of a generic self-report statement (e.g. “how many accidents have you been involved in over the past *X* years?)", with one study corroborating self-report with department of transport records.[25]

3.2.2 *On-the-road driving tests*

On-the-road tests were used in *k*=17 studies, comprised of *N*=2,079 (PD=724, HC=1,355) participants. Tests were generally based on standard certified driving tests and conducted and scored in a fairly uniform manner across studies using established, standardized protocols, e.g. Iowa Department of Transportation Driving Test Scoring Standards, 2005.[27] Scoring was carried out by a professional driving instructor, *k*=11, driving rehabilitation specialist, *k*=8, or trained experimenter, *k*=2 (4 studies used 2 types of assessor). Performance was evaluated during assessment (*k*=12), by post-assessment video review (*k*=4) or both (*k*=1). Driving tests were mostly of 45-60 minutes duration and performed in an instrumented vehicle in off-peak, daylight hours in good weather using a mixture of residential, suburban, urban and highway roads.

Main driving outcomes were (a) test pass/failure (*k*=8), and/or (b) overall performance score (*k*=15)[a] typically computed by summing scores on a broad range of individual driving elements (e.g. lane deviation, speed control, maneuvers).

---

[a] Three studies[22,23,26] of 18 originally identified were excluded from analysis of overall test performance, as they reported performance subscale scores used to compute overall scores in another included study[24]

3.2.3 *Driving simulator assessments*

Twenty-five studies with a total *N*=1,420 (PD=628, HC=792) assessed driving performance using a variety of simulators with real car or simulated cockpits. Sessions typically lasted 5-20 mins (*k*=12; *M*=11 mins), with multiple sessions sometimes used. Every day driving situations were commonly reproduced using different road and traffic conditions involving simple to hazardous driving (e.g. cars turning out of side streets). Performance was scored by a driving instructor (e.g. traffic sign compliance) or automatically by the simulator (e.g. lane deviation).

Outcomes included (a) crash rate/number of crashes (*k*=8), usually collisions with pedestrians, cars and objects in challenging driving scenarios, and/or (b) performance scores (*k*=22). Performance aspects commonly assessed included speed regulation, lane deviation, reaction time (e.g. to red lights), stop sign errors etc. Most (*k*=14) of the 22 test performance studies assessed performance on a relatively broad range of key components, while a few studies (*k*=8) were primarily restricted to a specific, circumscribed outcome, usually reaction time (see Table e-1).

Participant characteristics were similar across real-life, on-the-road and simulator studies (sections 3.1.2 and 3.1.3), with the possible exception of PD severity, with UPDRS-III scores 8 points higher for the on-road sample (*M*=24.6) compared to the simulator sample (*M*=16.4).

3.3 Quality of supporting evidence

Study validity ratings (Appendix e-1) largely appeared to support sound methodological practices, with driving assessed using well-established assessment protocols (96% of studies) and clearly defined performance measures (92%). Most studies reported similar group age (86%) and driving experience (64%). However, male/female ratio was similar for only 46% of included studies, and

thus we examined sex as a possible confounder (section 3.8). Most studies used performance measures adjudged to be sufficiently broad to reflect overall competence (74%), but with Table e-1 showing that some simulator studies used highly specific, circumscribed measures (e.g. reaction time to red lights). The potential influence of these studies was accordingly examined in sensitivity analysis (section 3.9). Finally, limited information was provided for selection of controls (42%) and many studies recruited PD participants from movement disorder clinics, and thus generalizability is difficult to ascertain. With respect to PD, movement clinics attract both complex referrals from neurologists and relatively non-complex patients from the local community, and thus there appears to be no obvious reason why study patients would be unrepresentative of the wider PD community.

3.4 Rater agreement

Inter-rater agreement for final study selection was 97%, with 86% overall agreement on study validity items. All disagreements were resolved following discussion with a third reviewer.

3.5 Meta-analysis

3.5.1 *All outcome data*

Meta-analysis of all outcome data (k=50) indicated moderate[3] overall driving impairment in PD (Figure 2), SMD=0.48, CI_{95}[0.35, 0.62], p<.001. However, a moderating effect of assessment type revealed greater impairment for on-the-road (ΔSMD=0.87, p<.001) and simulator (ΔSMD=0.56, p=.007) assessments compared to real-life driving and thus meta-analysis was conducted for each assessment method separately.

3.5.2 *Real life data*

Eight studies reporting crash rates (PD *M*=18.5%, HC *M*=21.0%) were combined with 5 studies reporting mean crashes. Meta-analysis of the 13 studies (*N*=2,143) revealed slightly lower odds of crash involvement for those with PD compared to HCs (Figure 3), but this did not reach significance, *OR*=0.84, CI_{95}[0.57, 1.23], *p*=.38. High inconsistency in the magnitude of these effects was observed (I^2=71% τ^2=0.40).

3.5.3 *On-the-road driving tests*

Driving test failure rates (*k*=8; *N*=821) and test performance scores (*k*=15; *N*=1,865) were analyzed separately, as test failure rate in particular represents an easily interpretable and meaningful index of driving competence.

Analysis revealed the odds of test failure were just over 6 times greater for PD patients (Figure 4), *OR*=6.16, CI_{95}[3.79, 10.03], *p*<.001, with a mean failure rate 46.5% in PD participants and 12.2% in HCs. Low heterogeneity was observed (I^2=24% τ^2=0.11), with all studies finding higher failure rates in PD. Analysis of test performance scores (Figure 5) also found poorer performance in PD (*SMD*=0.67, CI_{95}[0.53, 0.80], $p < .001$), suggesting moderate to high impairment.[3] All studies indicated greater driving impairment in PD, but with variation in the magnitude of impairment (I^2=62% τ^2=0.07).

3.5.4 *Driving simulators*

Crash data (*k*=8; *N*=598) and overall simulator performance scores (*k*=22; *N*=1,318) were analyzed separately, so that comparisons of crash rates across real-life (section 3.2.1) and simulator models can be made.

A greater likelihood of simulator crashes (*OR*=2.63, CI_{95}[1.64, 4.22], *p*=.008) and poorer overall driving performance (*SMD*=0.50, CI_{95}[0.33, 0.68], *p*<.001) was found in PD. The majority of studies found poorer performance in PD for crashes (*k*=8/8) and overall performance (*k*=19/22), although moderate-high variation in the magnitude of this impairment was found (I^2=60-71% τ^2=0.20-0.41).

3.6 Publication bias

Neither funnel plots or statistical tests (*p*=.27-.99) indicated effect size asymmetry for any outcomes, suggesting little detectable evidence of publication bias.

3.7 Meta-regression

As primary moderator data were only reported by around half of the included studies, meta-regression was performed for each moderator on the whole dataset (pooled across assessment models) to maximize power, with assessment model included as a covariate to control for its impact on effect size. *SMD* (transformed and untransformed) was used as the effect size, as this was the predominant statistic reported.

Meta-regression found group differences in driving performance to be unaffected by primary moderators of disease duration (*k*=27; *p*=.29) and symptom severity (H&Y: *k*=29; *p*=.36, UPDRS-III: *k*=24, *p*=.21), or secondary moderators of age (*k*=44; *p*=.39), sex (*k*=38; *p*=.38), L-dopa equivalent dosage (*k*=20; *p*=.14) and years of driving experience (*k*=16; *p*=.31).

3.8 Confounders

To control for the higher proportion of males in the PD samples, all analyses were rerun entering group differences in male/female proportion as a covariate, but revealed little change in effect size (max change: *SMD*=0.01, *OR*=0.37). For real-life crash data, we were not able to enter driving exposure (section 2.11) over the assessed crash period as a covariate due to limited studies and the variety of metrics used (e.g. mileage, trip frequency). Instead, we reran meta-analysis of real-life crash data (section 3.5.2) including 6 studies (*N*=943) from the original 13 that reported driving exposure in PD as matching or exceeding that of HCs. An identical effect size of OR=0.84, was found, demonstrating that lack of differences in self-reported crash rate was unlikely to be attributable to reduced driving exposure in PD.

3.9 Sensitivity analysis

Recomputing meta-analysis estimates after omitting studies where (a) extraction decisions were made (section 2.8), (b) highly circumscribed simulator outcomes (section 3.2.3) were used, and (c) SMDs were transformed to ORs and vice-versa (section 3.5.1-3.5.2), resulted in no substantive changes in effect size (max change: *SMD*=.05, *OR*=0.13). Finally, including patient group as a second-order hierarchical variable (section 2.10) resulted in negligible effect size changes, with the possible exception of overall performance scores in simulator studies which saw a reduction from *SMD*=.50 to *SMD*=.30.

3.9.1 *Missing studies*

Four driving simulator studies of varying sample sizes (N=9-53), described in conference abstracts by the same author (51-54 in References e-1), were excluded due to insufficient outcome data,

but reported results broadly consistent with analyzed data suggesting little evidence of bias from their omission.

4 DISCUSSION

Meta-analysis of 50 studies totalling 5,410 participants provided clear evidence of driving impairment in on-the-road (OTR) and driving simulator assessments, but found no evidence of increased risk of real-life crash involvement as assessed by self-report. More specifically: (1) In OTR tests, the odds of driving test failure were over 6 times higher for PD compared to HCs and overall test performance scores indicated moderate to strong impairment (*SMD*=.67); (2) in driving simulator assessments, the odds of crash involvement were over 2.5 times higher in PD and overall driving performance scores indicated moderate impairment (*SMD*=.50); (3) while considerable heterogeneity in the magnitude of performance degradation was generally observed, direction of effects in OTR and simulator assessments consistently indicated poorer performance; (4) no evidence was found that the degree of driving impairment was moderated by disease severity, disease duration or medication dosage.

Meta-analysis of OTR and simulator data provided convincing evidence of impaired driving competency in PD, which could not be attributed to group differences in age or sex. As OTR assessments were largely based on established, standardized driving tests used to determine licensure and fitness to drive, these results suggest that key skills required for typical every day driving challenges are compromised in PD. While simulator assessments may be less representative of real-life driving, especially when a single, molecular driving assessment is used

(e.g. reaction time to a red light), a consistent pattern of impairment to OTR assessments in independent studies was nevertheless observed. This would seem to suggest that, when confronted with equivalent driving challenges in real-life, those with PD are likely to perform more poorly. Furthermore, driving tests were conducted during optimal conditions (e.g. daylight, good weather, optimal medication states), suggesting further performance degradation is likely during more difficult conditions when already compromised cognitive, motor and perceptual capacities are further challenged.[21] It is also noteworthy that average Hoehn-Yahr severity stage ranged from 1-3 suggesting substantive driving impairment occurs even at the early mild to moderate disability levels, when the patient is usually considered to be fully independent in all activities of daily living.

However, despite strong evidence of impaired ability, PD was not associated with inflated real-life crash risk, based on self-report data from over 2,100 participants. This finding could not be attributed simply to reduced driving exposure in PD (e.g. fewer trips or miles covered). One possible explanation for the lack of increased crash risk is that people with PD self-regulate their driving behaviour in a number of ways. They may use compensatory strategies (e.g. driving more slowly, avoiding risky manoeuvres such as turning across traffic flow, using familiar roads, etc.) that minimise challenges to their impairments.[7,25] Additionally, evidence suggests that PD drivers with the most severe driving deficits, who are likely to be those with the highest crash vulnerability, may simply elect to discontinue driving altogether.[25,27] A second possibility is that the driving deficits observed in PD simply do not contribute substantially to crash risk, with established factors such as distracted driving (e.g. cell phone use), speeding, alcohol, driving attitude etc. being the primary risk factors for crashes.

Given that PD is a progressive disease, the lack of relationship between driving impairment and disease duration or severity based on 32 studies is somewhat surprising. This finding suggests that driving impairment may occur relatively early on in the disease state with little evidence of further deterioration. Nevertheless, several caveats should be applied to this interpretation. First, patient samples were generally restricted to those with a Hoehn-Yahr stage of 1-3, and it may be that substantial deterioration is seen only at stage 4 (severely disabling), when many PD patients may elect to discontinue driving. Second, disease severity measures used, typically Hoehn-Yahr staging and UPDRS-III, were based on motor dysfunction. However, it is well recognized that severity of non-motor symptoms such as cognitive and visual dysfunction in PD are key predictors for poor driving performance.[25,27] Finally, standardized road tests may fail to detect motor-based impairment in driving as they do not include hazardous circumstances that require immediate motor response. Nevertheless, real-life and simulator driving do incorporate such hazards, and so any existing relationship might still be expected to emerge in these data.[21] Collectively, there appears to be no persuasive evidence that driving progressively degrades with increasing severity in motor dysfunction up to Hoehn-Yahr stage 3, but it would be unwise to dismiss this possibility, or a role of non-motor symptoms, without further primary studies.

There is no persuasive evidence that drivers with PD show an increased real-life crash risk based on currently available self-report data. While the reliability of such data is uncertain, they do little to advance arguments for immediate legislative change in relicensing or current clinician recommendations for those with PD, pending availability of more objective measures. People with

PD often self-regulate their driving to limit safety risk,[5,25] and the most effective reassessment is likely to be that performed on an individual case-by-case basis rather than driven by diagnosis.[12]

At the same time, it is evident that PD is associated with some degree of operational deficit in driving ability, even if there is currently limited evidence that safety is compromised. Awareness of this in assessors may help the development of targeted interventions to tackle these deficits, improving competency when more challenging driving situations are encountered. PD deficits may be similar to those of the older driver, and may benefit from specific areas of retraining highlighted by assessment, such as awareness of road positioning and speed regulation.[17]

Limitations of the study should be noted. First, real-life crash data was based on self-report, and fear of licence revocation or encouraging negative perceptions may cause drivers with PD to under-report their crash involvement.[6] Inaccuracy of reporting may be also a product of a diminished higher-order ability to self-reflect on performance deficits exhibited by some PD patients,[15] and efforts to develop accurate assessment of such metacognitive deficits may also be an important factor in determining driving capacity. Interestingly, crash statistics from the one primary study that used state records[5,25] were only negligibly different from overall meta-analytic summary estimates that were otherwise based on self-report. Furthermore, a recent study found 'substantial' agreement between self-reported and state-recorded crashes in 2000 older adults[18] who may be susceptible to similar bias. Nevertheless, it would be unwise to draw firm conclusions without further corroborating evidence. Second, PD study samples in OTR and simulator studies may be biased towards more confident drivers less fearful of evaluation,[6] which would suggest driving ability might even be further impaired in the wider PD population. Third, while we found no evidence that the lack of inflated real-life crash risk in PD might be due to reduced driving

exposure, limited available information and the variety of different metrics used (e.g. trips, monthly mileage) restricted analytical power of our analysis and it is unwise to discount this as a possible explanatory factor. Similarly, factors such as daytime sleepiness/sudden sleep onset, non-use of medication, visuospatial deficits and severe global cognitive impairment (beyond the level commonly specified by study entry criteria), may also contribute to driving impairment but could not but be examined due to inadequate data for meta-analysis. Fourth, while meta-analytic summary data provides important information on 'average' impairment, PD is a heterogeneous condition and thorough assessment of each individual case is essential to reliably determine fitness to drive.

Further research examining real-life driving based on large-sample 'objective' crash statistics such as transport department records and insurance data, or empirical naturalistic driving data such as computerized multimodal quantitative assessments of driver behaviour in the driver's own vehicle,[29] are needed to corroborate self-report findings. Although such databases are not currently readily accessible, future availability would increase confidence in the current findings and potentially highlight in what situations crash vulnerability might be increased. Insights into whether overall driving impairment is primarily driven by selective impairment in individual motor, cognitive or visual deficits or for specific driving situations would also be gained by large sample OTR and simulator studies with carefully specified *a priori* comparison of outcomes. Such research could help target areas of driving impairment for retraining.

Meta-analysis of a total of 5,410 participants across 50 studies found persuasive evidence of driving performance deficits in those with PD. Participants with PD had an odds of on-the- test failure more than 6 times higher than controls, had more crashes in driving simulators and were

rated more poorly on overall driving performance. However, there is currently no evidence of an increased crash risk in real life driving based on self-report data, although corroborative objective indices are required. Driving is an important public safety issue and ensuring fitness to drive is paramount to minimising injuries and fatalities that occur on the road. However, these results do little to support mandated periodic driving reassessment for Parkinson's patients based on currently available evidence, and encourage thorough individualized assessment as the most appropriate method for determining fitness to drive.

ACKNOWLEDGEMENTS

We are especially grateful to Drs Alex Crizzle, Rene Stolwyk, Maud Ranchet, Sherrilene Classen, Judith Charlton and Carsten Buhmann for their rapid and very helpful responses to requests for additional study data.

REFERENCES

1. Adler G, Rottunda SJ. Mandatory testing of drivers on the basis of age and degenerative diseases: Stakeholder opinions. J Aging Soc Policy 2010;22:304-319
2. Bédard M, Riendeau J, Weaver B, Clarkson A. Roadwise Review has limited congruence with actual driving performance of aging drivers. Accid Anal Prev 2011;43:2209-2214
3. Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale, NJ: Lawrence Earlbaum Associates; 1988.
4. Comella CL. Daytime sleepiness, agonist therapy, and driving in Parkinson disease. JAMA 2002;287:509-511
5. Crizzle AM, Classen S, Uc EY. Parkinson disease and driving: an evidence-based review. Neurology 2012;79:2067-2074
6. Crizzle AM, Myers AM, Almeida QJ. Drivers with Parkinson's disease: who participates in research studies? Parkinsonism Relat Disord 2012;18:833-836
7. Devos H, Ranchet M, Emmanuel Akinwuntan A, Uc EY. Establishing an evidence-base framework for driving rehabilitation in Parkinson's disease: A systematic review of on-road driving studies. NeuroRehabilitation 2015;37:35-52
8. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997;315:629-634
9. Fisher Z, Tipton E. robumeta: Robust variance meta-regression. R package version 1.3. 2014
10. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003;327:557-560
11. Jacobs M, Hart EP, Roos RAC. Driving with a neurodegenerative disorder: an overview of the current literature. J Neurol 2017;264:1678-1696
12. Klimkeit EI, Bradshaw JL, Charlton JL, Stolwyk RJ, Georgiou-Karistianis N. Driving ability in Parkinson's disease: Current status of research. Neurosci Biobehav Rev 2009;33:223-231
13. Michon JA. A critical view of driver behavior models: What do we know, what should we do. In: Evans L, Schwing RC, eds: Human Behavior and Traffic Safety. Boston, MA, Springer US, 1986, pp 485-524
14. Moher D, Liberati A, Tetzlaff J, Altman DG, Prisma G. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 2009;339:b2535
15. Rizzo M, Uc EY, Dawson JD, Anderson SW, Rodnitzky RL. Driving difficulties in Parkinson's disease. Mov Disord 2010;25:pp S136-S140
16. Schwarzer G, Carpenter JR, Rucker G. Meta-analysis with R. Switzerland: Springer; 2015.
17. Singh R, Pentland B, Hunter J, Provan F. Parkinson's disease and driving ability. J Neurol Neurosurg Psychiatry 2007;78:363-366
18. Singletary BA et al. Self-reported vs state-recorded motor vehicle collisions among older community dwelling individuals. Accid Anal Prev 2017;101:22-27
19. Thompson T et al. Pain perception in Parkinson's disease: A systematic review and meta-analysis of experimental studies. Ageing Res Rev 2017;35:74-86
20. Tipton E. Small sample adjustments for robust variance estimation with meta-regression. Psychol Methods 2015;20:375-393
21. Uc EY et al. Driving under low-contrast visibility conditions in Parkinson disease. Neurology 2009;73:1103-1110
22. Uc EY, Rizzo M, Anderson SW, Sparks J, Rodnitzky RL, Dawson JD. Impaired visual search in drivers with Parkinson's disease. Ann Neurol 2006;60:407-413
23. Uc EY, Rizzo M, Anderson SW, Sparks JD, Rodnitzky RL, Dawson JD. Driving with distraction in

Parkinson disease. Neurology 2006;67:1774-1780
24. Uc EY et al. Road safety in drivers with Parkinson disease. Neurology 2009;73:2112-2119
25. Uc EY et al. Real-life driving outcomes in Parkinson disease. Neurology 2011;76:1894-1902
26. Uc EY, Rizzo M, Anderson SW, Sparks JD, Rodnitzky RL, Dawson JD. Impaired navigation in drivers with Parkinson's disease. Brain 2007;130:2433-2440
27. Uc EY, Rizzo M, O'shea AMJ, Anderson SW, Dawson JD. Longitudinal decline of driving safety in Parkinson disease. Neurology 2017;89:1951-1958
28. Viechtbauer W. Metafor: meta-analysis package for R. R package version 2010;2010:1-0
29. Yu L, Dawson JD, Aksan NS, Tippin J, Rizzo M. Effects of environmental factors on naturalistic driving in obstructive sleep apnea. Proc Int Driv Symp Hum Factors Driv Assess Train Veh Des 2013;2013:142-148

LIST OF FIGURES

LIST OF ONLINE SUPPLEMENTARY MATERIAL

Table e-1. Study characteristics

Study	N	PD Severity[a] (duration, severity scores)	PD Usual medication[b]	Driving Model	Outcome
Jitkritsadakul 2017[22]	41 (PD) 41 (HC)	9 yrs, HY=2.5	NR	Real-life Simulator	Crashes in last year Reaction time (e.g. to falling object), driving 'mistakes'
Uc et al 2017[45]	67 (PD) 110 (HC)	6.9 yrs, UPDRS-3=24, HY=2.2	NR (DA+other 67%)	Real-life On-road	Crash rate in preceding year Errors (speed regulation, lane changing, overtaking, parking, rail road crossing, stop signs, signals etc.)
Jitkritsadakul 2016[23]	60 (PD) 60 (HC)		NR	Real-life Simulator	Crash rate in preceding year Reaction time (e.g. to falling object), driving 'mistakes'
Pavlou, Beratis et al 2016[31]	25 (PD) 31 (HC)		NR	Simulator	Speed variability, headway, lane deviation, reaction time
Pavlou et al 2016[30]	25 (PD) 31 (HC)		NR	Simulator	Crash rate, Reaction time
Ranchet et al 2016[32]	16 (PD) 21 (HC)	6.9 yrs, UPDRS-3=17, HY=2.2	LD+DA+MAO (89%), MAO (5%), DA (5%)	Simulator	Speed variability
Vardaki et al 2016[46]	10 (PD) 10 (HC)	UPDRS-3=12.9, HY=1.9	NR	Real-life Simulator	Crashes in last 2 yrs Speed
Aksan et al 2015[2]	39 (PD) 77 (HC)		NR	On-road	7 error types (lane change, lane observance, speed control, traffic signs, stop signs, turns, pulling off from curb)
Chen et al 2015[7]	16 (PD) 18 (HC)	UPDRS-3=12.4, HY=1.8	NR	Simulator	Speed regulation, lane deviation, gap acceptance
Dotzauer et al 2015[16]	9 (PD) 9 (HC)	5.7 yrs	NR	Simulator	Speed regulation, headway

Buhmann et al 2014[4]	21 (PD) 21 (HC)	6 yrs, HY=1.9	*LD (57%), DA (76%), MAO-B (42%), COMT (23%), amandatine (33%), anticholinergics (10%)	Simulator	Driving test errors
Classen et al 2014[8]	101 (PD) 138 (HC)	UPDRS-3=25.9, HY=2 (median)	NR	Real-life On-road	Crashes (rate and mean number) in last 3 yrs Driving test result (pass/fail) 8 error categories (vehicle positioning, speed regulation, lane maintenance, yielding, signalling, visual scanning, adjustment to environment, gap acceptance)
Papadimitriou 2014[29]	7 (PD) 17 (HC)		NR	Simulator	Speed variability, headway, lane deviation
Crizzle et al 2013[14]	27 (PD) 20 (HC)	3.9 yrs, UPDRS-3=30.1	LD (85%), rasagiline (7%), ropinirole (4%), rasagiline + ropinirole (4%)	Real-life	Reaction time
Crizzle and Myers 2013[13]	27 (PD) 20 (HC)	3.9 yrs, UPDRS-3=30.1	LD (85%), rasagiline (7%), ropinirole (4%), rasagiline + ropinirole (4%)	Real-life	Speed
Chee et al 2013[5]	28 (PD) 30 (HC)	HY=1.7	NR	Simulator	Performance scores (speed regulation, stop signs, lane deviation, road edge excursions, stop at lights) Crashes
Dotzauer et al 2013[17]	9 (PD) 9 (HC)	5.7 yrs	NR	Simulator	Crash rate
Ranchet et al 2013[33]	19 (PD) 21 (HC)	7.5 yrs, UPDRS-3=16.4, HY=2.1	LD + DA + MAO (89%), MAO (5%), DA (5%)	On-road	Errors (lateral position control, following distance, speed regulation, visual information processing, traffic signals, overtaking behavior, traffic insight,

					interaction with traffic) Instructor interventions
Crizzle et al 2012[12]	20 (PD) 148 (HC)		NR	On-road	Driving test result (pass/fail)
Classen et al 2011[10]	41 (PD) 41 (HC)	UPDRS-3=27.4, HY=2.4	NR	On-road	Driving test result (pass/fail) 8 error categories from 91 manoeuvres (e.g. vehicle positioning, speed regulation, lane maintenance, yielding, signalling, visual scanning, adjustment to environment, gap acceptance)
Lee et al 2011[25]	53 (PD) 129 (HC)	5.3 yrs, UPDRS-3=18, HY=1.6	LD (NR%)	Simulator	Performance score (lane changing, traffic signs, T-junctions, speed regulation), Crashes
Scally 2011[34]	19 (PD) 19 (HC)	6.6 yrs, UPDRS-3=15.4	LD (47%), LD + DA (26%), LD + COMT (26%)	Simulator	Reaction time (approaching red light)
Uc et al 2011[41]	106 (PD) 130 (HC)	5.9 yrs, UPDRS-3=24.9, HY=2.2	NR	Real-life On-road	(Prospective) crash rate in two year follow-up period Driving test errors (e.g. starting, pulling away from curb, traffic signals and signs, turns, lane observations, overtaking , speed, regulation, reverse driving, parking manoeuvres etc.)
Barrash et al 2010[3]	33 (PD) 24 (HC)	HY=3	NR	On-road	76 error types (e.g. starting, pulling away from curb, traffic signals and signs, turns, lane observations, overtaking another vehicle, control of speed, reverse driving, parking manoeuvres)
Chee et al 2010[6]	7 (PD) 15 (HC)	HY=1.6	NR	Simulator	Performance scores (speed regulation, stop signs, lane deviation, road edge excursions, stop at lights) Crashes
Vaux et al 2010[47]	8 (PD) 18 (HC)		NR	Real-life Simulator	Crashes in last 2 yrs Collision detection score
Uc et al 2009[40]	67 (PD)	6.5 yrs, UPDRS-	NR	Simulator	Lane deviation, reaction time (to crash), Crash rate

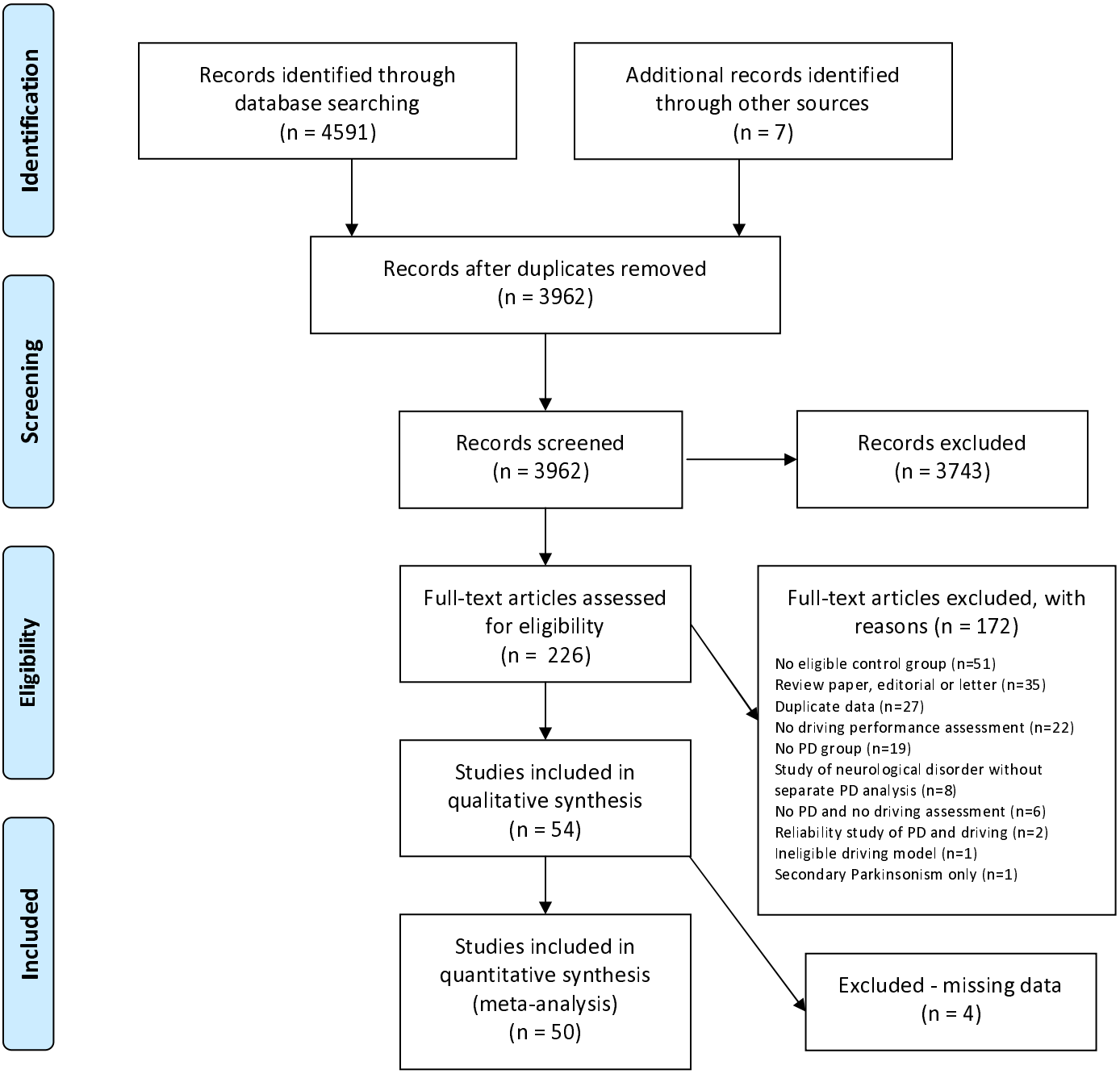
	51 (HC)	3=25.6, HY=2.3			
Uc et al 2009[37]	84 (PD) 182 (HC)	5.9 yrs, UPDRS-3=25.6, HY=2.2	NR	On-road	Errors (lane changing, steering, speed regulation etc.)
Classen et al 2009[9]	19 (PD) 104 (HC)	4.9 yrs, UPDRS-3=25.6	NR	On-road	Driving test result (pass/fail) 8 error categories (vehicle positioning, speed regulation, lane maintenance, yielding, signalling, visual scanning, adjustment to environment, gap acceptance)
Cordell et al 2008[11]	53 (PD) 129 (HC)	5.3 yrs, UPDRS-3=18, HY=1.6	NR	On-road	Performance scores (traffic signs, roundabouts, steering, braking, traffic lights, indicators)
Devos et al 2007[15]	40 (PD) 40 (HC)	6.7 yrs, UPDRS-3=20.4, HY=2 (median)	NR	Simulator	13 Performance items (e.g. speed regulation, traffic light faults)
Kaußner et al 2007[24]	24 (PD) 24 (HC)	6.7 yrs, UPDRS-3=18.5	LD + DA (79%), DA (21%)	Real-life Simulator	Crash rate in preceding 5 years Lane deviation, speed regulation
McCarthy et al 2007[28]	19 (PD) 62 (HC)	UPDRS-3=25.9	NR	On-road	Driving test result (pass/fail) Errors (lane deviation, speed variability)
Uc et al 2007[43]	77 (PD) 152 (HC)	5.7 yrs, UPDRS-3=23.7, HY=2.2	LD + DA (36%), LD (27%), DA (29%), other (7%)	On-road	Errors (e.g. steering, lane deviation, shoulder incursion, stopping/slowing in unsafe circumstances, unsafe intersection behavior)
Uc et al 2006[44]	31 (PD) 19 (HC)		NR	Simulator	Errors (e.g. steering, lane deviation, shoulder incursion, stopping/slowing in unsafe circumstances, unsafe intersection behavior)
Uc et al 2006[42]	71 (PD) 147 (HC)	5.3 yrs, UPDRS-3=23.5, HY=2.1	LD + DA (36%), LD (27%), DA (29%), other (7%)	On-road	Errors (e.g. steering, lane deviation, shoulder incursion, stopping/slowing in unsafe circumstances, unsafe intersection behavior)
Uc et al 2006[39]	79 (PD) 151 (HC)	5.6 yrs, UPDRS-3=24.1,	LD + DA (36%), LD (27%), DA (29%),	On-road	Errors (e.g. steering, lane deviation, shoulder incursion, stopping/slowing in unsafe circumstances,

		HY=2.2	other (7%)		unsafe intersection behavior)
Uc et al 2006[38]	71 (PD) 147 (HC)	HY=3	LD + DA (36%), LD (27%), DA (29%), other (7%)	On-road	Errors (e.g. steering, lane deviation, shoulder incursion, stopping/slowing in unsafe circumstances, unsafe intersection behavior)
Ferreira et al 2006[19]	176 (PD) 174 (HC)	10 yrs, UPDRS-3=19.8, HY=2.3	*LD (87%), DA (69%)	Real-life	Crashes in last 3 yrs
Stolwyk 2006[36]	18 (PD) 18 (HC)	6.7 yrs, UPDRS-3=11.7	LD + other (95%), other (5%)	Simulator	Speed approaching red light, speed variability, lane deviation
Worringham et al 2006[49]	25 (PD) 21 (HC)	6.2 yrs, UPDRS-3=27.4, HY=2.3	NR	On-road	Driving test result (pass/fail)
Grace et al 2005[20]	21 (PD) 21 (HC)	7.1 yrs, UPDRS-3=28.4, HY=2 (mode)	LD + DA (38%), LD (29%), DA (29%)	Real-life On-road	Crash rate in preceding 3 years Driving test result (safe/unsafe) Errors (e.g. lane changing, signalling, speed regulation)
Stolwyk et al 2005[35]	18 (PD) 18 (HC)	6.7 yrs, UPDRS-3=11.7	LD + other (95%), other (5%)	Simulator	Reaction time approaching red light, speed variability, lane deviation
Wood et al 2005[48]	25 (PD) 21 (HC)	6.2 yrs, UPDRS-3=27.4, HY=2.3	NR	Real-life On-road	Crash rate in preceding 10 years Instructor interventions Errors (e.g. lane deviation, braking, indicating, gap selection approach, blind spot)
Zesiewicz et al 2002[50]	39 (PD) 25 (HC)	UPDRS-3=17.8, HY=2.2	NR	Simulator	Crashes (rate and mean number)
Adler et al 2000[1]	89 (PD) 423 (HC)	5.8 yrs	NR	Real-life	Crash rate in preceding 3 years
Heikkila 1998[21]	20 (PD) 20 (HC)	5.6 yrs, HY=1.9	LD + MAO (65%), LD + MAO + DA (30%), LD + MAO + COMT (5%)	On-road	Driving test result (pass/fail) Errors (a broad range from Finnish driving test)

Lings et al 1992[26]	28 (PD) 109 (HC)	8.8 yrs, HY=2.2	*LD (89%), anticholinergic (37%), DA (29%)	Simulator	Reaction time (traffic lights)
Dubinsky 1991[18]	45 (PD) 100 (HC)	10.5 yrs, HY=3	NR	Real-life	Crashes in last 3 yrs
Madeley et al 1990[27]	10 (PD) 10 (HC)	HY=1.9	NR	Simulator	Steering errors, reaction time, red lights missed

[a]**PD Severity:** UPDRS-3=Unified Parkinson's Disease Rating Scale-3 (motor subscale), HY=Hoehn and Yahr scale,
[b]**PD Medication**: LD=Levodopa, DA=Dopamine agonist, MAO=Monoamine Oxidase inhibitor, COMT=Catechol-O-methyltransferase inhibitor.
*includes use of medication alone and in combination with other PD medications.

Figure 1. PRISMA 2009 Flow Diagram



From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). *P*referred *R*eporting *I*tems for *S*ystematic Reviews and *M*eta-*A*nalyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit www.prisma-statement.org.

Figure 2. Standardized Mean Differences for all assessment models

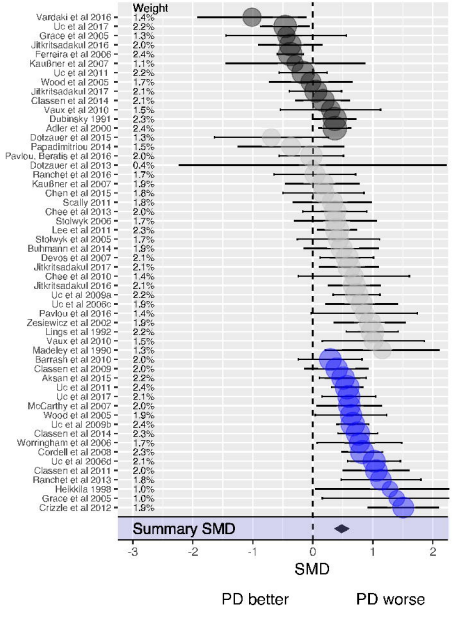


Figure 3. Odds ratios for real-life crash involvement

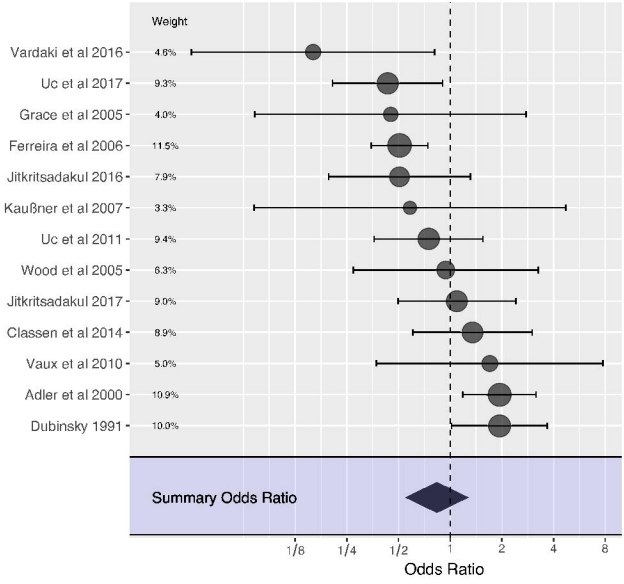


Figure 4. Odds ratios for on-the-road test failure

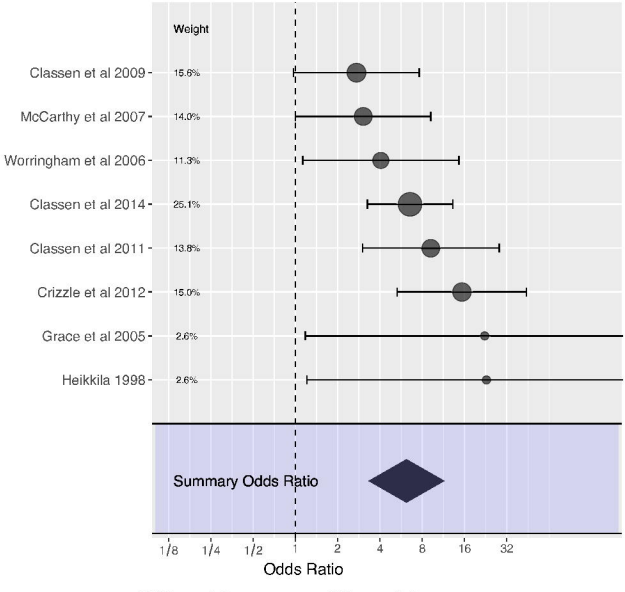


Figure 5. Standardized mean differences for on-the-road performance score