Validation of a UPDRS-/MDS-UPDRS-based definition of functional dependency for Parkinson's disease

Neil Ramsay MRCP[a1], Angus D Macleod MRCP PhD[a*1], Guido Alves MD PhD[b,c], Marta Camacho MS[d], Lars Forsgren MD PhD[e], Rachael A Lawson PhD[f], Jodi Maple-Grødem PhD,[b,c] Ole-Bjørn Tysnes MD PhD[g], Caroline H Williams-Gray MRCP PhD[d], Alison J Yarnall MRCP PhD,[f] Carl E Counsell MRCP MD[a] on behalf of the Parkinson's Incidence Cohorts Collaboration

a. Institute of Applied Health Sciences, University of Aberdeen, Aberdeen, UK
b. The Norwegian Centre for Movement Disorders, Stavanger University Hospital, Norway
c. Department of Chemistry, Bioscience and Environmental Engineering, University of Stavanger, Norway
d. Department of Clinical Neurosciences, University of Cambridge, UK
e. Department of Clinical Science, Neurosciences, Umeå University, Sweden
f. Translational and Clinical Research Institute, Newcastle University, UK
g. Department of Neurology, Haukeland University Hospital, University of Bergen, Norway

[1]Joint first authors
*Corresponding Author: Room 1.015, Polwarth Building, Foresterhill Campus, University of Aberdeen, AB25 2ZD, Email: angus.macleod@abdn.ac.uk Phone: +44 (0)1224 437120

Email addresses of authors:
Dr Ramsay: neil.ramsay@nhs.net
Dr Macleod: angus.macleod@abdn.ac.uk
Prof Alves: guido.werner.alves@sus.no
Marta Camacho: msc72@cam.ac.uk
Prof Forsgren: Lars.Forsgren@umu.se
Dr Lawson: rachael.lawson@newcastle.ac.uk
Dr Maple-Grødem: jodi.maple@uis.no
Prof Tysnes: Ole-Bjorn.Tysnes@uib.no
Dr Williams-Gray: chm27@cam.ac.uk
Dr Yarnall: alison.yarnall@newcastle.ac.uk
Dr Counsell: Carl.Counsell@abdn.ac.uk

Word count: 1700
Abstract word count: 242

Key Words: Parkinson's disease, Dependency, measurement scales, validation

Declarations of interest: none

Abstract

Introduction:
Functional dependency in basic activities of daily living (ADLs) is a key outcome in Parkinson's disease (PD). We aimed to define dependency in PD, using the original and MDS versions of the Unified Parkinson's Disease Rating Scale (UPDRS).

Methods:
We developed two algorithms to define dependency from items of UPDRS Part 2 and MDS-UPDRS Part 2 relating to basic ADLs (feeding, dressing, hygiene and walking, and getting out of a chair). We validated both algorithms using data from 1110 patients from six community-based PD incidence cohorts, testing concurrent validity, convergent validity, and predictive validity.

Results:
Our optimal algorithm showed high specificity and moderate to high sensitivity versus Schwab & England <80% (specificity 95% [95% confidence interval (CI) 93-97] and sensitivity 65% [95% CI 55-73] at baseline; 88% [95% CI 85-91] and 85% [95% CI 79-97] respectively at five-years follow-up). Convergent validity was demonstrated by strong associations between dependency defined by the algorithm and cognition (MMSE), quality of life (PDQ39), and impairment (UPDRS part 3) (all p <0.001). Algorithm-defined dependency status also predicted mortality: HR for mortality in those dependent vs independent at baseline was 1.6 (95%CI 1.2-2.1) and in those dependent vs independent at five-years' follow-up was 2.2 (1.6-3.0).

Discussion:
We have demonstrated the concurrent validity, convergent validity, and predictive validity of a UPDRS-/MDS-UPDRS-based algorithm to define functional dependency in PD. This can be used for studying dependency in any study where UPDRS or MDS-UPDRS part 2 data have been collected.

Introduction

Functional dependency is an important patient-orientated outcome in Parkinson's disease (PD) [1]. Here we focus specifically on dependency in terms of needing help with basic activities of daily living (ADLs), such as washing, dressing, toileting, feeding, or walking. Previous population-based studies have shown that dependency is common, even early in the disease course, [2,3] but there has been little research into the factors that influence this aspect of PD [1].

Existing dependency or activity limitation scales for PD research have drawbacks. The Schwab & England scale has been used most frequently but it is inherently unclear (in particular, the descriptor "chores" is not defined). A Schwab & England score <80% has been used as a definition of dependency (80%=completely independent in most chores; 70%=not completely independent) [3]. The Barthel Index has been used widely in other diseases, but infrequently in PD. Part 2 "Activities of daily living" of the Unified Parkinson's Disease Rating Scale (UPDRS) and part 2 "Motor Aspects of experiences of daily living" of the Movement Disorders Society (MDS) revision have been used widely, but only a minority of the items in these sections relate to ADLs. If dependency could be identified from the UPDRS part 2, this would allow research into dependency in studies which have collected UPDRS data but not a specific activity limitation/dependency scale. We therefore aimed to develop an algorithm to identify functional dependency using the original UPDRS and MDS-UPDRS Part 2.

Methods

Dataset

We used data from the Parkinson's Incidence Cohorts Collaboration (PICC), a project to pool data from six PD incidence cohorts in Northern Europe (CamPaIGN[4], ICICLE-PD[5], NYPUM[6], ParkWest[7], PICNICS[8], and PINE[9]). These studies each collect demographic, clinical, and genetic data at time of diagnosis and at regular follow-up visits thereafter, with data on 1110 patients at baseline and 714 patients at approximately 5 years. Reasons for missing data at year 5 included (i) death (N=167); (ii) loss to follow-up (N=123); (iii) participants had not reached 5-years' follow-up or were not seen at about 5 years (N=106). Available data included the original UPDRS (CamPaIGN, NYPUM, ParkWest, PINE), MDS-UPDRS (ICICLE-PD, PICNICS), Schwab & England scale (all studies, except not collected at baseline in PICNICS or ICICLE-PD) and mini-mental state examination (MMSE) (all studies), PDQ-39 (all studies except ParkWest), and Barthel Index (only PINE). Ethical approval for each study was obtained from relevant ethics committees. All participants gave informed consent to participation in the respective studies.

Algorithm development

We developed two algorithms (Table 1) to define dependency using items from UPDRS Part 2 relating to basic ADLs (feeding, dressing, hygiene and walking in both versions and getting out of a chair in MDS-UPDRS). We did not include the turning in bed item from the MDS-UPDRS

because an individual could live independently even if they were unable to turn over in bed. We first assessed what cut-off score for each item would necessarily indicate dependency based on the descriptors, i.e., if the patient had that score or higher they must have been dependent. For instance, for the original UPDRS hygiene item, a score of 3 ("requires assistance for washing, brushing teeth...") was the lowest item which indicated definite evidence of dependency. The cut-offs for each item which indicated patients were clearly dependent are given in table 1. Algorithm 1 was defined as indicating dependency if a patient was clearly dependent in any one of the specified ADL items. Although a simple approach, this algorithm was expected to lead to under-ascertainment of dependency because of the ambiguity in certain descriptors for these items. For example, the original UPDRS hygiene score of 2 ("needs help to shower or bathe; or very slow in hygienic care") could indicate dependency ("needs help to shower or bathe") or independency ("very slow in hygienic care"). Algorithm 2 therefore defined dependency if either (i) any one of these ADL items clearly indicated dependency or (ii) the sum of these ADL items was greater than a certain threshold. To select the threshold we calculated specificity and sensitivity for several thresholds in the PINE study against a "gold standard" of dependency defined by a Barthel Index score of less than 16, excluding the items relating to continence as these are not specific ADLs. We selected the threshold with the maximal accuracy ($\frac{sensitivity+specificity}{2}$). We chose the Barthel Index as a gold standard because a score of less than 16 (after continence items excluded) means the patient needs help with at least one of the following ADLs: mobility, stairs, transfers, dressing, bathing, grooming, feeding, or toilet use. The threshold for the MDS-UPDRS was calculated as the same proportion of maximum score from the specified ADL items.

Algorithm validation

For both algorithms, we tested concurrent validity (validating against other instruments that measure the same concept) [10] by assessing specificity and sensitivity of the selected algorithm against dependency defined by Schwab & England scale <80% in each study at baseline (diagnosis) and at five years follow-up (4.5 years in ICICLE-PD). We chose to develop the algorithm 2 cut-off score using the Barthel index in PINE and validate it against the Schwab & England scale in all studies (rather than development using Schwab & England and validation using Barthel in PINE and Schwab & England in the other studies) because i) we wished to use an unambiguous measure of dependency for model development; ii) it allowed all studies to be used in the validation; and iii) it allowed a common and unbiased assessment of heterogeneity across all studies.

We then evaluated convergent validity (evidence of correlation with similar, but not identical, measures) [10] by assessing associations of each algorithm with cognition (MMSE), quality of life (PDQ39), and parkinsonian motor impairment (UPDRS Part 3) at baseline and at 4.5-5 years. For each association, MMSE, PDQ39 or UPDRS part 3 defined the dependent variable

in multivariable linear regression, and the independent variables were dependency as defined by the particular algorithm, age at baseline and sex.

We tested predictive validity by comparing mortality in those defined as dependent and independent by these algorithms in all six studies using Cox regression adjusted for age at baseline and sex. We developed Cox models with survival time measured from (i) baseline and (ii) the 4.5-5 year assessment until death. Patients who had not died were censored at the date last known to be alive.

Statistical analyses were performed using SPSS version Stata version 16 and SPSS version 25.

Results

The PICC cohort included 1110 patients at baseline. Mean age was 69.5 (standard deviation 10.0) and 61% were men. Further characteristics are found in supplementary table 1.

The two algorithms are detailed in table 1. The selected thresholds for algorithm 2 were ≥6 (original UPDRS) and ≥7 (MDS-UPDRS). Concurrent, convergent, and predictive validity are shown in Table 2, Supplementary Tables 2-4, and Supplementary Figures 1-3. The numbers included in each analysis are detailed in the supplementary tables. Algorithm 1 had only low to moderate sensitivities versus Schwab & England (35% [95% CI 26-44] and 58% [95% CI 51-64] at baseline and year 5), but very high specificity (99% [95% CI 98-100] at both time points). Algorithm 2 had higher sensitivities (65% [95% CI 55-73] and 85% [95% CI 79-89] at baseline and year 5), while specificities remained high (95% [95% CI 93-97] and 88% [95% CI 85-91] at baseline and year 5).

There were significant associations between dependency in all the constructs across both algorithms. Using both algorithms, those defined as dependent had significantly higher mortality than those independent at both time points with mortality ratios ranging from 1.6 to 2.6. The results were mostly consistent between studies, but there was some heterogeneity.

Discussion

We have demonstrated the concurrent validity, convergent validity, and predictive validity of two definitions of dependency derived simply from the UPDRS/MDS-UPDRS. Although both algorithms had good face validity, convergent validity and predictive validity, algorithm 2 had better sensitivity while maintaining high specificity. We propose that algorithm 2 is a useful measure of functional dependency in basic ADLs but algorithm 1 may be useful as measure of a more-severe level of dependency. Concurrent validity was better at year 5 than at baseline, probably because baseline dependency relates relatively more to comorbid diseases and later dependency relatively more to PD-related factors; dependency items in UPDRS may be more specific to dependency caused by PD.

This study has several key strengths. The algorithms have been validated in six prospective cohort studies with very low selection bias (each cohort was derived from an incidence study, therefore attempting to include all PD in the population). Therefore, we believe these results are generalizable to Caucasian populations with PD. The data were prospective with a large pooled sample size, and we used multiple methods in our validation.

Nevertheless, this study has some limitations. We lacked a true gold standard for functional dependency. However, we did use validated ADL scales. The Barthel Index has been widely validated in assessing ADLs but has had little validation in PD [11]. Schwab and England has some validation data relating to activity limitation in PD but one study has suggested a score <80% has suboptimal sensitivity as a dichotomous dependency measure [12] despite its face validity. Furthermore, the Barthel data in PINE was self-reported and other scales were reliant on self-report if relatives/carers did not come to study visits so there may be some lack of objectivity. Arguably, the most accurate dependency definition would be to observe patients performing ADLs in their home environment, but this would be resource-intensive and time-consuming and such data were not available. There was heterogeneity in the results between the individual studies included in PICC. There may several reasons for this, including small numbers in individual studies in some of the analyses and variability between studies in how some of the scales were scored. The overwhelming majority of participants were white European so further validation could be done in other ethnic groups.

In conclusion, we have demonstrated the validity of a measure derived from the UPDRS or MDS-UPDRS to simply assess functional dependency in PD. It can therefore be easily applied in most existing clinical studies without additional data collection. Further work is needed to establish which ADL/dependency measure should be the gold standard for use in PD.

Acknowledgements
We would like to thank all participants, study personnel from each study and funders of individual studies and of PICC.

Members of PICC Steering Group: Dr Angus D Macleod, Dr Carl E Counsell, University of Aberdeen, UK; Prof Ole-Bjørn Tysnes University of Bergen, Norway; Marta Camacho, Dr Caroline Williams-Gray, University of Cambridge, UK; Dr Rachael A Lawson, Newcastle University, UK; Dr Jodi Maple-Grødem, Prof Guido Alves, Stavanger University Hospital, Norway; Prof Lars Forgren, Umeå University, Sweden.

We acknowledge the contributions of members of the individual study groups:
PINE Study: Carl E Counsell, Angus D Macleod, Robert Caslake, Kate S M Taylor, David J M McGhee, Joanna Gordon, Clare Harris, Hazel Forbes

CamPaIGN study: Roger A. Barker, Thomas Foltynie, Sarah L. Mason, Caroline H. Williams-Gray.

PICNICS study: Roger A. Barker, Marta Camacho, Gemma Cummins, Jonathan R. Evans, David P. Breen, Ruwani S. Wijeyekoon, Caroline H. Williams-Gray.

NYPUM Study: Lars Forsgren, Jan Linder, Mona Edström, Jörgen Andersson, Linda Eriksson, David Bäckström, Gun-Marie Hariz, Magdalena Domellöf.

ParkWest Study: ParkWest Principal investigators: Guido Alves (Norwegian Centre for Movement Disorders, Stavanger University Hospital) and Ole-Bjørn Tysnes (Haukeland University Hospital). Study personnel: Michaela Dreetz Gjerstad, Kenn Freddy Pedersen, Elin Bjelland Forsaa, Veslemøy Hamre Frantzen, Anita Laugaland, Jodi Maple-Grødem, Johannes Lange, Karen Simonsen, Eldbjørg Fiske and Ingvild Dalen (Stavanger University Hospital); Bernd Müller, Geir Olve Skeie and Marit Renså (Haukeland University Hospital); Wenche Telstad, Aliaksei Labusau and Jane Kastet (Førde Hospital); Ineke HogenEsch, Marianne Kjerandsen and Liv Kari Håland (Haugesund Hospital); Karen Herlofson, Solgunn Ongre and Siri Bruun (Sørlandet Hospital Arendal).

ICICLE-PD Study: David Burn, Lynn Rochester, Alison J Yarnall, Rachael A Lawson, Gordon W Duncan, Tien K Khoo.

Authors' Roles

Dr Ramsay: 1B, 1C, 2C, 2A, 2B, 3A.
Dr Macleod: 1A, 1B, 1C, 2C, 2A, 2B, 2C, 3B.
Prof Alves: 1A, 1B, 1C, 2C, 3B.
Camacho: 1B, 1C, 2C, 3B.
Prof Forsgren: 1A, 1B, 1C, 2C, 3B.
Dr Lawson: 1A, 1B, 1C, 2C, 3B.
Dr Maple-Grodem: 1B, 1C, 2C,3B
Prof Tysnes: 1A, 1B, 1C,2C, 3B.
Dr Williams-Gray: 1A, 1B, 1C,2C, 3B.
Dr Yarnall: 1A, 1B, 1C, 2C, 3B.
Dr Counsell: 1A, 1B, 1C, 2C, 3B.

Financial support:

The PICC collaboration was funded by the Chief Scientist Office of the Scottish Government, NHS Education for Scotland, and the Academy of Medical Sciences.

The CamPaIGN study has received funding from the Wellcome Trust, the Medical Research Council, the Patrick Berthoud Trust, Parkinson's UK, and the National Institute for Health

Research (NIHR) Cambridge Biomedical Research Centre Dementia and Neurodegeneration Theme (ref. 146281).

The ICICLE-PD study was funded by Parkinson's UK (J-0802, G-1301, G-1507) and the, Lockhart Parkinson's Disease Research Fund. The research was supported by the NIHR Newcastle Biomedical Research Unit and Centre based at Newcastle upon Tyne Hospitals NHS Foundation Trust and Newcastle University, and the National Institute for Health Research (NIHR) Cambridge Biomedical Research Centre Dementia and Neurodegeneration Theme (ref. 146281).

The NYPUM study was funded by Swedish Medical Research Council, Parkinson Foundation in Sweden, the Swedish Parkinson Disease Association, University of Umeå, Foundation for Clinical Neuroscience at Umeå University Hospital, Västerbotten County Council (ALF) and King Gustaf V's and Queen Victoria's foundation

The Norwegian ParkWest study was funded by the Western Norway Regional Health Authority (grant No 911218), the Research Council of Norway (grant No 177966 and 287842) and the Norwegian Parkinson Research Foundation.

The PICNICS study was funded by the Cure Parkinson's Trust, the Van Geest Foundation, the MRC and Parkinson's UK, and the National Institute for Health Research (NIHR) Cambridge Biomedical Research Centre Dementia and Neurodegeneration Theme (ref. 146281).

The PINE study was funded by Parkinson's UK, the Scottish Chief Scientist Office, NHS Grampian endowments, the BMA Doris Hillier award, RS Macdonald Trust, the BUPA Foundation and SPRING.

C H Williams-Gray holds a RCUK/UKRI Research Innovation Fellowship awarded by the Medical Research Council (MR/R007446/1) and receives support from the Cambridge Centre for Parkinson-Plus.

References

[1] A.D. Macleod, J.W. Grieve, C.E. Counsell, A systematic review of loss of independence in Parkinson's disease, J. Neurol. 263 (2016) 1-10.

[2] A. Bjornestad, K.F. Pedersen, O.-B. Tysnes, G. Alves, Clinical milestones in Parkinson's disease: A 7-year population-based incident cohort study, Parkinsonism Relat. Disord. 42 (2017) 28-33.

[3] A.D. Macleod, C.E. Counsell, Predictors of functional dependency in Parkinson's disease, Mov. Disord. 31 (2016) 1482-1488.

[4] C.H. Williams-Gray, S.L. Mason, J.R. Evans, T. Foltynie, C. Brayne, T.W. Robbins, R.A. Barker, The CamPaIGN study of Parkinson's disease: 10-year outlook in an incident population-based cohort, J. Neurol. Neurosurg. Psychiatry 84 (2013) 1258-1264.

[5] A.J. Yarnall, D.P. Breen, G.W. Duncan, T.K. Khoo, S.Y. Coleman, M.J. Firbank, C. Nombela, S. Winder-Rhodes, J.R. Evans, J.B. Rowe, B. Mollenhauer, N. Kruse, G. Hudson, P.F. Chinnery, J.T. O'Brien, T.W. Robbins, K. Wesnes, D.J. Brooks, R.A. Barker, D.J. Burn, Characterizing mild cognitive impairment in incident Parkinson disease: the ICICLE-PD study, Neurology 82 (2014) 308-316.
[6] J. Linder, H. Stenlund, L. Forsgren, Incidence of Parkinson's disease and parkinsonism in northern Sweden: a population-based study, Mov Disord 25 (2010) 341-8.
[7] G. Alves, B. Muller, K. Herlofson, I. HogenEsch, W. Telstad, D. Aarsland, O.B. Tysnes, J.P. Larsen, Incidence of Parkinson's disease in Norway: the Norwegian ParkWest study, J. Neurol. Neurosurg. Psychiatry 80 (2009) 851-7.
[8] D.P. Breen, J.R. Evans, K. Farrell, C. Brayne, R.A. Barker, Determinants of delayed diagnosis in Parkinson's disease, J. Neurol. 260 (2013) 1978-81.
[9] R. Caslake, K. Taylor, N. Scott, J. Gordon, C. Harris, K. Wilde, A. Murray, C. Counsell, Age-, gender-, and socioeconomic status-specific incidence of Parkinson's disease and parkinsonism in northeast Scotland: the PINE study, Parkinsonism Relat. Disord. 19 (2013) 515-21.
[10] I. McDowell, Measuring heath: A Guide to Rating Scales and Questionnaires, Oxford University Press, New York, 2006.
[11] D. Morley, C. Selai, A. Thompson, The self-report Barthel Index: preliminary validation in people with Parkinson's disease, Eur. J. Neurol 19 (2012) 927-9.
[12] A. Bjornestad, O.-B. Tysnes, J.P. Larsen, G. Alves, Reliability of Three Disability Scales for Detection of Independence Loss in Parkinson’s Disease, Parkinsons Dis 2016 (2016) 1941034 .

Scale version	Item	Individual item threshold	Threshold for sum of items for additional use in Algorithm 2
Original UPDRS	2.9 – Cutting food	≥3	≥6 (out of 16)
	2.10 – Dressing	≥3	
	2.11 – Hygiene	≥3	
	2.15 – Walking	≥3	
MDS-UPDRS	2.4 – Cutting food	≥3	≥7 (out of 20)
	2.5 – Dressing	≥3	
	2.6 – Hygiene	≥2	
	2.11 – Getting out of a chair	≥3	
	2.12 – Walking	≥4	

Table 1 – Thresholds for UPDRS ADL items used to develop the algorithms. Algorithm 1 defines dependency if a patient is necessarily dependent in any of the individual items, defined by the stated threshold for each item. Algorithm 2 defines dependency if a patient is necessarily dependent in any of the individual items (third column) OR if the sum of the individual items meets the stated threshold (fourth column).

	Algorithm 1				Algorithm 2			
	Baseline		Year 5[a]		Baseline		Year 5[a]	
Concurrent validity	TP/TP+FN Sens, % (CI)	TN/TN+FP Spec, %, (CI)	TP/TP+FN Sens % (CI)	TN/TN+FP Spec, % (CI)	TP/TP+FN Sens, % (CI)	TN/TN+FP Spec, % (CI)	TP/TP+FN Sens, % (CI)	TN/TN+FP Spec, % (CI)
PINE study only vs Barthel Index	18/60 30 (19-43)	94/95 99 (94-100)	28/69 41 (29-53)	37/37 100 (91-100)	38/60 63 (50-75)	95/95 100 (96-100)	51/69 74 (62-84)	37/37 100 (91-100)
All studies vs S&E <80	38/110 35 (26-44)	517/522 99 (98-100)	122/212 58 (51-64)	409/413 99 (98-100)	71/110 65 (55-73)	497/522 95 (93-97)	180/212 85 (79-89)	365/413 88 (85-91)
Convergent validity	Adjusted difference[b] (95% CI)	P-value	Adjusted difference[b] (95% CI)	P-value	Adjusted difference[b] (95% CI)	P-value	Adjusted difference[b] (95% CI)	P-value
PDQ-39	16.1 (11.3 – 20.9)	<0.001	24.7 (20.8 –28.7)	<0.001	17.4 (14.3 – 20.4)	<0.001	22.7 (19.6 – 25.7)	<0.001
MMSE	-1.5 (-1.9 – -1.0)	<0.001	-4.5 (-5.3 – -3.8)	<0.001	-0.8 (-1.1 – -0.5)	<0.001	-2.9 (-3.5 – -2.2)	<0.001
UPDRS/MDS-UPDRS part 3	13.4 (10.5 – 16.3)	<0.001	15.9 (13.5 – 18.3)	<0.001	12.7(10.7 – 14.6)	<0.001	16.4 (14.5 – 18.3)	<0.001
Predictive validity	HR for mortality, dependent vs independent (95% CI)	P value	HR for mortality, dependent vs independent (95% CI)	P value	HR for mortality, dependent vs independent (95% CI)	P value	HR for mortality, dependent vs independent (95% CI)	P value
All studies	1.7 (1.2 – 2.3)	0.002	2.6 (1.9 – 3.5)	<0.001	1.6 (1.2 – 2.1)	<0.001	2.2 (1.6 – 3.0)	<0.001

Table 2: Concurrent validity, convergent validity, and predictive validity of the two dependency algorithms. [a]4.5 years in ICICLE-PD. [b]Adjusted difference represents the difference in score between dependent and independent as defined by the algorithm, adjusted for age and sex. A positive difference in PDQ-39 score indicates worse quality of life, a negative difference in MMSE indicates poorer cognition and a positive difference in UPDRS part 3 indicates a more severe motor impairment in those who are dependent. Abbreviations: CI = confidence interval; FN = false negative; FP = false positive; HR = hazard ratio; MDS-UPDRS = Movement Disorder Society revision of the Unified Parkinson's disease rating scale; MMSE = mini-mental state examination; PDQ-39 = Parkinson's disease questionnaire – 39 item; S&E = Schwab & England; TN = true negative; TP = true positive; UPDRS = Unified Parkinson's disease rating scale.

Supplementary table 1: Characteristics of incidence cohorts included in PICC

Study	Location	Years of recruitment	Number of patients at baseline/5 years	Mean age in years of patients at baseline (SD)	Percentage male at baseline (%)	Maximum duration of follow-up data (years)
CamPaiGN	Cambridgeshire, UK	2000-2	142/109	70.4 (9.6)	56.3	10
ICICLE-PD	Newcastle and Gateshead, UK	2009-11	154/91	66.4 (10.4)	54.9	8
NYPUM	Umeå, Sweden	2004-9	144/109	71.2 (9.9)	60.3	13
ParkWest	Western Norway	2004-6	191/160	68.1 (9.3)	61.3	8
PICNICS	Cambridgeshire, UK	2008-13	280/120	68.8 (9.7)	62.1	9
PINE	Aberdeen, UK	2002-4;06-9	199/127	70.4 (9.6)	56.3	16
All Studies			1110/716	69.5 (10.0)	61.0	16

Supplementary table 2: Concurrent validity of both algorithms at baseline and year 5.

	Algorithm 1				Algorithm 2			
	Baseline		Year 5		Baseline		Year 5	
Study & comparison (N baseline/N year 5)	TP/TP+FN Sens, % (CI)	TN/TN+FP Spec, %, (CI)	TP/TP+FN Sens % (CI)	TN/TN+FP Spec, % (CI)	TP/TP+FN Sens, % (CI)	TN/TN+FP Spec, % (CI)	TP/TP+FN Sens, % (CI)	TN/TN+FP Spec, % (CI)
PINE study only vs Barthel Index (155/106)	18/60 30 (19-43)	94/95 99 (94-100)	28/69 41 (29-53)	37/37 100 (91-100)	38/60 63 (50-75)	95/95 100 (96-100)	37/69 74 (62-84)	51/69 100 (91-100)
Campaign vs S&E <80 (141/99)	44/37 16 (7-30)	97/97 100 (96-100)	32/56 57 (43-70)	43/43 100 (92-100)	24/44 55 (39-70)	95/97 98 (93-100)	48/56 86 (74-94)	32/43 74 (59-87)
ICICLE-PD vs S&E <80 (NA/ 90)	No data	No data	11/20 55 (32-77)	69/70 99 (92-100)	No data	No data	16/20 80 (56-94)	53/70 76 (64-85)
NYPUM vs S&E <80 (141/ 108	4/13 31 (9-61)	126/128 98 (95-100)	15/22 68(45-86)	85/86 99 (94-100)	11/13 85 (55-98)	125/128 98 (93-100)	20/22 91 (71-99)	83/86 97 (90-99)
ParkWest vs S&E <80 (190/159)	6/13 46 (19-75)	2/177 99 (96-100)	20/37 54 (37-71)	120/122 98 (94-100)	8/13 62 (32-86)	171/177 97 (93-99)	31/37 84 (68-94)	117/122 96 (91-99)
PICNICS vs S&E <80 (NA/ 42)	No data	No data	6/10 60 (26-88)	32/32 100 (89-100)	No data	No data	9/10 90 (56-100)	28/32 88 (71-97)
PINE vs S&E <80 (160/127)	21/40 53 (36-69)	118/119 99 (95-100)	38/67 57 (44-69)	60/60 100 (94-100	28/40 70 (54-83)	106/120 88 (81-94)	56/67 84 (73-92)	52/60 87 (75-94)
All studies vs S&E <80 (632/ 625)	38/110 35 (26-44)	517/522 99 (98-100)	122/212 58 (51-64)	409/413 99 (98-100)	71/110 65 (55-73)	497/522 95 (93-97)	180/212 85 (79-89)	365/413 88 (85-91)

Abbreviations: CI=95% confidence interval; FN = false negative; FP = false positive; N=number; NA=not applicable; S&E Schwab & England scale; sens=sensitivity; spec=specificity; TN = true negative; TP = true positive.

Supplementary table 3: Convergent validity of dependency algorithms versus PDQ-39, MMSE, and UPDRS/MDS-UPDRS part 3.

	Algorithm 1				Algorithm 2			
	Baseline		Year 5		Baseline		Year 5	
Outcome Study (N at baseline/N at 5 years)	Adjusted difference (95% CI)	P-value	Adjusted difference (95% CI)	P-value	Adjusted difference (95% CI)	P-value	Adjusted difference (95% CI)	P-value
PDQ-39								
Campaign (125/103)	40.3 (26.0 – 54.5)	<0.001	42.0 (30.7 – 53.2)	<0.001	34.7 (28.0 – 41.5)	<0.001	39.4 (29.2 – 49.6)	<0.001
ICICLE-PD (147/88)	-1.3 (-10.9 – 8.2)	0.78	30.4 (21.8 – 38.9)	<0.001	0.8 (-5.2 – 6.9)	0.79	20.8 (14.3 – 27.3)	<0.001
NYPUM (109/92)	9.5 (-8.2 – 27.2)	0.29	23.2 (13.8 – 32.5)	<0.001	17.7 (8.6 – 26.7)	<0.001	22.0 (15.1 – 28.9)	<0.001
ParkWest	No data		No data		No data		No data	
PICNICS (223/82)	25.0 (13.5 – 36.5)	<0.001	28.9 (15.6 – 42.2)	<0.001	27.4 (19.4 – 35.4)	<0.001	34.9 (25.2 – 44.6)	<0.001
PINE (153/102)	15.1 (9.5 – 20.7)	<0.001	23.2 (13.9 – 32.5)	<0.001	14.7 (10.7 – 18.7)	<0.001	15.1 (9.7 – 20.5)	<0.001
All Studies (757/467)	16.1 (11.3 – 20.9)	<0.001	24.7 (20.8 –28.7)	<0.001	17.4 (14.3 – 20.4)	<0.001	22.7 (19.6 – 25.7)	<0.001
MMSE								
Campaign (142 /48)	-0.6 (-1.6 – 0.5)	0.30	-5.1 (-7.0 – -3.3)	<0.001	-0.5 (-1.1 – 0.1)	0.11	-2.9 (-4.8 – -1.0)	<0.001
ICICLE-PD (154/90)	-0.7 (-1.5 – 0.1)	0.10	-3.4 (-5.0 – -1.7)	<0.001	-0.3 (-0.9 – 0.1)	0.19	-1.4 (-2.7 – -0.1)	0.03
NYPUM (134 /102)	-3.8 (-5.1 – -2.4)	<0.001	-7.0 (-9.3 – -4.8)	<0.001	-2.2(-3.2 – -1.3)	<0.001	-6.3 (-8.0 – -4.6)	<0.001
ParkWest (191/153)	-2.6 (-4.3 – -0.9)	0.002	-5.5 (-7.3 – -3.6)	<0.001	-1.2 (-2.5 – 0.1)	0.07	-4.4 (-5.9 – -3.0)	<0.001
PICNICS (279/123)	-0.9 (-1.6 – -0.2)	0.01	-1.8 (- 3.0 – -0.6	0.03	-0.5 (-1.0 – 0.01)	0.05	-0.9 (-1.8 – 0.1)	0.08
PINE (165/128)	-1.8 (-2.7 – -0.8)	<0.001	-4.7 (-6.7 – -2.8)	<0.001	-1.1 (-1.9 – -0.4)	0.004	-3.5 (-5.4 – -1.6)	<0.001
All Studies (1065/644)	-1.5 (-1.9 – -1.0)	<0.001	-4.5 (-5.3 – -3.8)	<0.001	-0.8 (-1.1 – -0.5)	<0.001	-2.9 (-3.5 – -2.2)	<0.001
UPDRS/MDS-UPDRS part 3								
Campaign (139/108)	11.2 (2.1 – 20.2)	0.02	15.8 (10.3 – 21.3)	<0.001	12.6 (7.7 – 17.4)	<0.001	16.1 (10.7 – 21.4)	<0.001
ICICLE-PD (153/91)	10.4 (2.7 – 18.2)	0.01	15.7 (8.8 – 22.6)	<0.001	14.7 (10.3 – 19.0)	<0.001	13.1 (8.3 – 18.0)	<0.001
NYPUM (144/106)	11.1 (2.7 – 19.5)	0.01	15.7 (8.7 – 22.8)	<0.001	10.8 (4.9 – 16.7)	<0.001	19.4 (14.3 – 24.6)	<0.001
ParkWest (191/159)	19.5 (12.3 – 26.8)	<0.001	23.1 (18.6 – 27.6)	<0.001	14.3 (8.8 – 19.9)	<0.001	18.1 (14.2 – 22.0)	<0.001
PICNICS (273/117)	11.2 (4.8 – 17.6)	0.001	11.1 (4.1 – 18.2)	0.002	10.7 (6.2 – 15.1)	<0.001	16.2 (11.1 – 21.2)	<0.001
PINE (189/130)	16.2 (11.4 – 21.0)	<0.001	14.0 (10.0 – 18.0)	<0.001	14.3 (10.7 – 17.9)	<0.001	12.7 (8.9 – 16.4)	<0.001
All Studies (1089/711)	13.4 (10.5 – 16.3)	<0.001	15.9 (13.5 – 18.3)	<0.001	12.7(10.7 – 14.6)	<0.001	16.4 (14.5 – 18.3)	<0.001

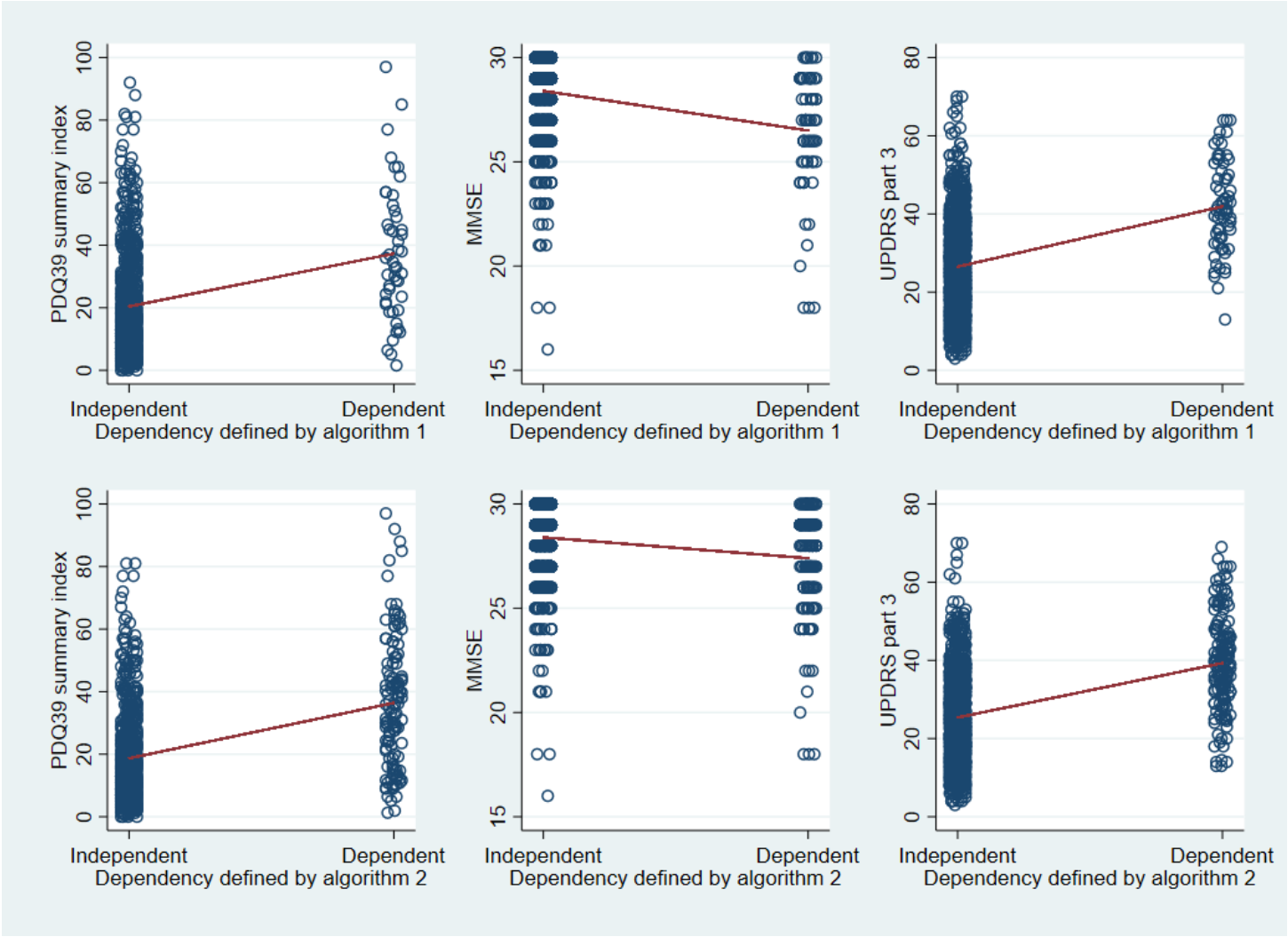
Abbreviations: CI=confidence interval; MDS-UPDRS=Movement Disorders Society revision of UPDRS; MMSE=mini-mental state examination; N=number; PDQ-39=Parkinson's disease questionnaire 39 item; UPDRS=unified Parkinson's disease rating scale.

Supplementary table 4: Predictive validity: Cox regression, hazard ratios for mortality comparing dependent with not dependent as defined by each algorithm.

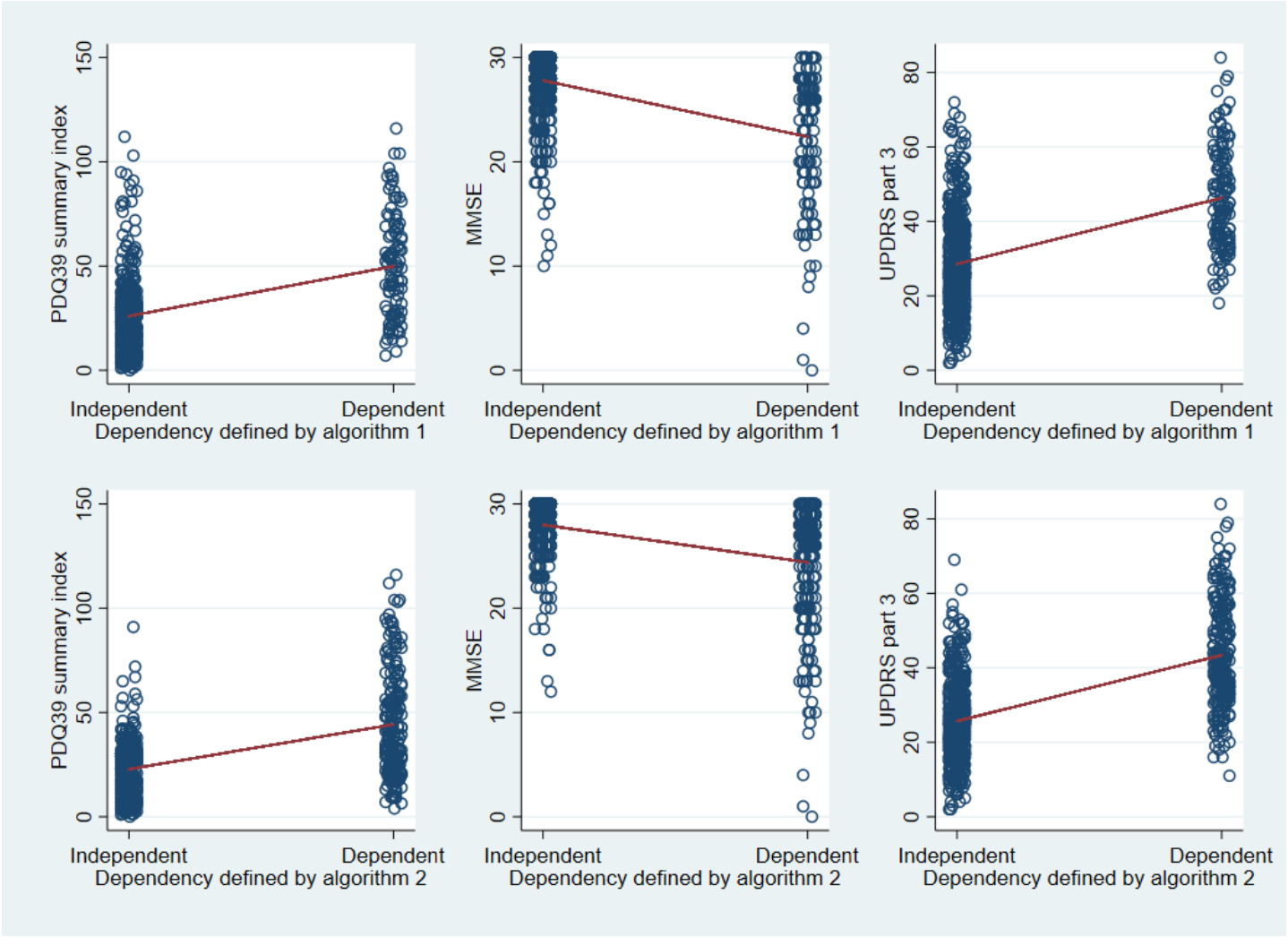
	Algorithm 1				Algorithm 2			
	Baseline		Year 5		Baseline		Year 5	
Study (N baseline/N alive at 5 years/N total deaths/N deaths after year 5)	HR (95% CI)	P value	HR (95% CI)	P value	HR (95% CI)	P value	HR (95% CI)	P value
Campaign (142/91/63/39)	1.46 (0.5 – 4.1)	0.48	2.5 (1.2 – 5.1)	0.01	2.0 (1.1 – 3.4)	0.02	2.6 (1.1 – 5.9)	0.03
ICICLE-PD (154/92/28/6)	3.0 (1.1 – 8.0)	0.03	1.8 (0.2 – 14.8)	0.61	4.6 (0.8 – 27.5)	0.09	3.7 (0.4 – 38.4)	0.27
NYPUM (144/106/77/42)	2.18 (0.87 – 5.4)	0.09	3.4 (1.6 – 7.0)	0.01	2.6 (1.3 – 5.1)	0.01	2.7 (1.4 – 5.4)	0.01
ParkWest (191/90/34/9)	0.63 (0.1 – 2.8)	0.54	10.1 (2.2 – 45.4)	0.003	1.02 (0.3 – 3.1)	0.97	6.3 (1.4 – 28.4)	0.002
PICNICS (280/79/50/10)	1.31 (0.5 – 3.8)	0.62	2.4 (0.5 – 12.6)	0.31	1.8 (0.8 – 4.3)	0.17	2.0 (0.5 – 7.7)	0.31
PINE (199/140/149/90)	1.47 (1.01 – 2.8)	0.05	1.8 (1.1 – 2.9)	0.02	1.2 (0.8 – 1.8)	0.45	1.6 (1.0 – 2.6)	0.06
All Studies (1110/598/401/196)	1.7 (1.2 – 2.3)	0.002	2.6 (1.9 – 3.5)	<0.001	1.6 (1.2 – 2.1)	<0.001	2.2(1.6 – 3.0)	<0.001

Abbreviations: CI=confidence interval; HR=hazard ratio; N=number.

Supplemental figure 1: Scatterplots showing association between dependency status (defined by each algorithm) and the three factors used to assess convergent validity at baseline. Data from all studies were included, where available. The red line illustrates the predicted values from the model (i.e. the average adjusted values from the model according to dependency status).



Supplemental figure 2: Scatterplots showing association between dependency status (defined by each algorithm) and the three factors used to assess convergent validity at year 5. Data from all studies were included, where available. The red line illustrates the predicted values from the model (i.e. the average adjusted values from the model according to dependency status).



Supplemental figure 3: Kaplan-Meier plots of survival from baseline in those defined as independent or dependent by algorithm 1 (A) and algorithm 2 (B) and survival from year 5 in those defined as independent or dependent by algorithm 1 (C) and algorithm 2 (D).