Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Shortened smell tests to predict alpha-synuclein seed amplification assay status

Rafaela Ravazio
Pontifícia Universidade Católica do Rio Grande do Sul
Eleanor Mitchell
Queen Mary University of London
Christian Mattjie
Pontifícia Universidade Católica do Rio Grande do Sul
Sophie I. Meyer
Queen Mary University of London
Rodrigo C. Barros
Pontifícia Universidade Católica do Rio Grande do Sul
Alastair Noyce
a.noyce@qmul.ac.uk

Queen Mary University of London

Article

Keywords:

Posted Date: November 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11704712/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations:
Competing interest reported. Authors RR, EM, CM, SIM & RCB declare no financial or non-financial competing interests. AJN reports grants from Parkinson's UK, Barts Charity, Cure Parkinson's, National Institute for Health and Care Research, Innovate UK, UKRI, HORIZON 2020, the Medical College of Saint Bartholomew's Hospital Trust, Alchemab, and the Michael J Fox Foundation. AJN reports consultancy fees from AbbVie, Bial, and Roche. AJN receives a travel stipend from TNC. AJN reports honoraria from the International Parkinson and Movement Disorders Society.

Figure 1 and 2 are available in the Supplementary Files section

Shortened smell tests to predict alpha-synuclein seed amplification assay status

Authors: Rafaela Ravazio[1]*, Eleanor Mitchell[2]*, Christian Mattjie[1]*, Sophie I. Meyer[2], Rodrigo C. Barros[1] Alastair J Noyce[2,3].

*These authors contributed equally to the work and share first authorship.

Affiliations: [1]Machine Learning Theory and Applications Lab, School of Technology, Pontifical Catholic University of Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil. [2]Centre for Preventive Neurology, Wolfson Institute of Population Health, Queen Mary University of London, London, UK. [3]Department of Neurology, The Royal London Hospital, Barts Health NHS Trust, London, UK.

Corresponding author:

Prof. Alastair J Noyce,

Professor of Neurology and Neuroepidemiology

Faculty of Medicine and Dentistry, Queen Mary University of London, Charterhouse Square, London, UK, EC1M 6BQ.

Email: a.noyce@qmul.ac.uk

Telephone: +44 (0) 207 882 5841

Word count: 4046; Introduction (335), Methods (1573), Results (874), Discussion (1,264).

Abstract

α-synuclein seed amplification assay (αsyn-SAA) positivity is an early biomarker of Parkinson's disease (PD), but reliable detection currently requires a lumbar puncture. Smell loss (hyposmia) appears to be a useful predictor of αsyn-SAA status in PD, however the association is less clear in prodromal phases. Large-scale administration of the 40-item University of Pennsylvania Smell Identification Test (UPSIT) is also resource intensive. Leveraging data from the Parkinson's Precision Medicine Initiative (PPMI) we used artificial intelligence to develop novel abbreviated UPSITs for αsyn-SAA status prediction in 1,117 'hyposmia-enriched' controls, free of neurological disease but with a high prevalence of smell loss. High-performing candidate tests were then evaluated in isolated REM-sleep behaviour disorder (iRBD; n=289) and PD (n=442). We demonstrate that the UPSIT can be abbreviated to 9 items with no reduction in specificity. In hyposmia-enriched controls, the 9-item test (cut-off ≤5 and ≤6 in males and females, respectively) was 14 percentage points less sensitive (80% vs 94%) but 17 points more specific (60% vs 43%) than the 40-item version. For iRBD, 6 percentage points were lost in sensitivity (68% vs 74%) and in PD the sensitivities were similar (72% vs 76%). Specificity was identical in iRBD and PD. AUC was 0.75 vs 0.77 (hyposmia-enriched controls), 0.83 vs 0.87 (iRBD) and 0.87 vs 0.90 (PD). The difference in AUC reached significance only in iRBD ($p<0.001$). Balanced accuracy differed by at most 3 points in any group, and performance was similar across 30 re-splits. For screening initiatives, abbreviated UPSITs could generate substantial cost savings, at the cost of more false-negatives. Different odour choices, test lengths or cut-offs might be preferable, depending on the intended application. For example, repeating the analysis with six North-American odours removed from the candidate pool produced an alternative, internationally 'portable' 9-item test (Cherry, Motor Oil, Mint, Banana, Clove, Leather, Apple, Soap, Rose) with the same cut-offs. For this test, balanced accuracy was 0.75 in Discovery, 0.69 in Validation, 0.84 in iRBD and 0.81 in PD.

Word count: 325

Introduction

Parkinson's disease (PD) is characterised by nigral neuronal loss and intraneuronal aggregation of α-synuclein[1]. Cross-sectional post-mortem data suggest this may occur sequentially, starting in the olfactory bulb or gastrointestinal tract before involving the nigrostriatum[2,3]. Until recently, this process could only be confirmed post-mortem. Over the last 10 years, α-synuclein seed amplification assays (αsyn-SAA)[1,4] have been developed, allowing for misfolded α-synuclein to be identified in cerebral spinal fluid (CSF), in both manifest PD and prodromal states[5,6]. Two recently proposed staging systems aim to biologically reconceptualise PD, with incorporation of αsyn-SAA status as a central component[7,8]. Though a promising research tool, the current gold standard necessitates an invasive lumbar puncture (spinal tap) and laboratory infrastructure, potentially widening global inequality in diagnosis and care.

Hyposmia (smell loss) is an early and frequent clinical biomarker of PD, occurring in 90% of confirmed cases and manifesting years before overt motor deficits[9]. Hyposmia also appears to be somewhat predictive of αsyn-SAA status[10]. For example, Siderowf et al. recently demonstrated that, in PD, hyposmia is more predictive of αsyn-SAA status than any common clinical or genetic factor[6,10] . Smell tests are safe, relatively quick and easy to administer (including remotely). Accordingly, they are increasingly being used in clinical studies to identify 'at risk' hyposmic individuals, suitable for αsyn-SAA testing[10,11]. Where a lumbar puncture is unavailable or inappropriate, smell testing may plausibly serve as a proxy. However, administering the thousands of tests required for large-scale initiatives is resource-intensive. The value of abbreviated smell tests as a singular predictor of αsyn-SAA status is worth exploring.

Leveraging data from the Parkinson's Precision Medicine Initiative (PPMI), we analysed Revised University of Pennsylvania Smell Identification Test (UPSIT) responses in a large, 'hyposmia-enriched' control group, free from overt neurological disease but with a high prevalence of smell loss[12]. We used artificial intelligence to develop novel abbreviated smell tests, optimised for predicting αsyn-SAA status

in this group. Candidate tests were then applied to participants with isolated REM-sleep behaviour disorder (iRBD) and PD, to assess performance across the disease spectrum.

Methods

Study aim

We aimed to develop an abbreviated smell test that accurately predicts αsyn-SAA status across three groups: hyposmia-enriched controls, iRBD and PD.

Participant source

The Parkinson's Precision Medicine Initiative (formally Parkinson's Progression Markers Initiative) is a longitudinal, observational, multi-centre study investigating clinical, biological and neuroimaging biomarkers of PD. Protocol information can be found at https://dx.doi.org/10.17504/protocols.io.n92ldmw6ol5b/v2. Briefly, participants with PD are recruited from movement disorder clinics at >50 international sites (majority USA-based). Individuals must be recently diagnosed (≤2 years) and PD drug-naïve at baseline. Healthy controls are recruited through various routes, including traditional & social media, commercials and charities. The prodromal cohort consists of people with hyposmia (determined via 'pre-screening' UPSIT), polysomnography-confirmed iRBD or a genetic predisposition to PD (e.g. *LRRK2* carriers). **Figure 1** summarises the inclusion and exclusion criteria; available in full at www.ppmi-info.org. All PPMI Clinical participants undergo extensive screening & baseline assessments, including smell testing and CSF collection. For our study, we excluded participants aged <50 years at the time UPSIT administration (n=49) - as the normative scoring percentiles are intended for adults aged ≥50 years - or those enrolled via the monogenic PD risk pathway (n=27).



Figure 1: Overview of participant selection.

Flowchart showing inclusion/exclusion criteria for all participants (left) and subgroups (middle). The healthy control and hyposmic cohort were combined to create a novel 'hyposmia-enriched' control group, split into Discovery & Validation sets (60%/40%). Candidate tests were developed on a class-balanced Discovery subset (n = 219 / 219). High performing tests were then evaluated in iRBD and PD cohorts.

*Depending on protocol version. PPMI = Parkinson's Precision Medicine Initiative; DaTSCAN™ = Dopamine transporter scan; PD = Parkinson's disease; iRBD = Isolated REM sleep behaviour disorder; UPSIT = University of Pennsylvania Smell Identification Test; αsyn-SAA = α-synuclein seed amplification assay.

Biomarker selection

Smell testing

The University of Pennsylvania Smell Identification Test (UPSIT) is a validated and internationally recognised assessment of olfactory function. It consists of 40 "scratch-and-sniff" microencapsulated strips, each accompanied by a four-option forced multiple-choice question. In 2020, a Revised UPSIT[12] was introduced, featuring several updated odorants and answer choices to improve odour recognisability. Both the Original and Revised versions have been administered in PPMI, with the Revised being increasingly used. Because the two versions differ significantly, we could not combine the data. We therefore excluded 356 participants whose first UPSIT was the Original version. 13 participants were excluded for missing >2 UPSIT item responses. Tests missing one or two item responses (131 PD and 29 hyposmia-enriched controls) were retained, with the missing items imputed by a deterministic k-nearest-neighbour rule (k=5), drawing donors from the Discovery cohort only. The 40-item comparator uses PPMI's own total and percentile, in which a missing item is scored as correct. A complete-case sensitivity analysis is reported in Supplementary Table 1.

αsyn-SAA status determination

αsyn-SAA testing protocols are detailed elsewhere[4]. Briefly, the Amprion α-synuclein seed amplification assay involves cyclic amplification of endogenous α-synuclein aggregates *in vitro*, allowing detection and quantification via fluorescent dyes. A sample is subsequently deemed αsyn-SAA positive or negative according to a probabilistic algorithm. For each participant, triplicate samples are tested. The algorithm combines the three replicates, including their amplification kinetics, into a positive, negative or inconclusive result[6]. Participants with only inconclusive results were excluded (n=4). For our cross-sectional design, we used each participant's baseline αsyn-SAA result. For 21 participants without a definitive baseline result, the first later definitive result (year 1 or 2) was used. For three participants

with both a positive and a negative baseline result, the first reported result was used; excluding the former or reversing the latter changed no cut-off and no conclusion (Supplementary Table 2).

Group selection

We developed candidate smell tests using UPSIT/αsyn-SAA data from two PPMI enrolment groups – healthy controls and 'hyposmics'. Healthy controls are defined as those free from clinically significant disease; historically UPSIT results were not factored in at recruitment. Since 2020, a separate hyposmic group has been established, either by distributing UPSITs in population-level screening initiatives or direct recruitment from taste-and-smell clinics. An UPSIT score ≤15th percentile (≤10th since 2023) was considered the threshold for hyposmia. Given the phenotypic overlap between these groups - i.e. a portion of those enrolled as healthy controls are hyposmic and participants enrolled as hyposmic are otherwise healthy - we combined them to create a hybrid 'hyposmia-enriched' group. Inclusion of a disproportionately high percentage of hyposmics was necessary for statistical power given the relatively low prevalence of αsyn-SAA positivity (2-6%[6,11,13,14]) in the general population. The discussion section explores the implications of this. High performing abbreviated UPSITs were then tested in two other comparator groups i) iRBD (polysomnography-confirmed) and ii) PD. See Figure 1 for a summary of the participant selection process.

Statistical analysis

The hyposmia-enriched control group was split into Discovery (60%) and Validation (40%) cohorts, stratified by αsyn-SAA status, sex, age band and UPSIT administration mode. Candidate tests were developed on a class-balanced subset of the Discovery cohort (438 participants; 219 αsyn-SAA positive and 219 negative), formed by randomly under-sampling the larger αsyn-SAA group within each sex and age band. Throughout, the 40-item comparator, a was considered positive at the ≤15th percentile[12], as this is the threshold used in PPMI until 2023 (later changed to ≤10th percentile)[10].

The UPSIT features 40 different odours, each of which could be included in a shortened smell test of varying lengths. As such, there are over one trillion possible odour groupings. We used a genetic algorithm[15,16]; an artificial intelligence method that mimics biological evolution to find solutions to complex problems. The following steps were taken:

- Initialisation: We generated a population of 1,000 candidate smell tests ('solutions') by randomly selecting each of the 40 odours independently with probability 0.5. We imposed no length constraint.
- Evaluation: We scored each test on two objectives, optimised together (NSGA-II): balanced accuracy for predicting αsyn-SAA status (maximised) and number of odours (minimised).
- Crossover: Tests ranked highly on both objectives were combined to generate new 'offspring'; parents and offspring then competed for places in the next generation, so the best tests were carried forward unchanged.
- Mutation: We slightly altered a random subset of offspring by adding or removing odours.
- Repeat: We repeated the cycle of selection, crossover, and mutation for up to 80 generations, stopping early once 20 generations passed without improvement in the highest balanced accuracy. Each run returns its Pareto front: the tests for which no other test is both at least as short or accurate.

During evaluation, we calculated the balanced accuracy (mean of sensitivity and specificity) separately in males and females. For each test, we determined the optimal sex-based diagnostic cut-off; a test was positive when the number of correctly identified odours was at or below the cut-off. Cut-offs were fitted in the class-balanced Discovery subset alone and applied unchanged to the Validation, iRBD and PD cohorts. The overall balanced accuracy for a test was the mean of its best accuracies for males and females.

To ensure reproducibility, we ran the genetic algorithm ten times, each with a different 'seed'. A seed sets the starting point for the random number generator used throughout the run. Each seed generates 1000 new random odour subsets for the initialisation stage. Once the algorithm had completed ten runs, 95 candidate UPSITs were identified on the ten final Pareto fronts. After removing duplicates, 23 remained, spanning lengths of 1-10 and 13-17 odours. The test with the highest balanced accuracy at each length in the class-balanced Discovery subset was then applied to the Validation, iRBD and PD groups. We report the 9-item test as the headline test. As a longer alternative, the 13-odour test represents the shortest test within 0.001 of the highest Discovery balanced accuracy (0.78 at 17 odours).

Confidence intervals and significance testing.

Sensitivity and specificity are reported with Wilson score intervals. Balanced accuracy, predictive values and paired differences in sensitivity and specificity are reported with a stratified bootstrap (2,000 replicates, resampled within αsyn-SAA status and sex). We compared areas under the curve (AUC) using DeLong's method for correlated curves, which also gave the confidence interval of their difference, and discordant classifications using McNemar's exact test. Predictive values hold only at each cohort's own αsyn-SAA prevalence and are not screening values. Sensitivity and specificity are calculated over all participants in a cohort, whereas balanced accuracy is the mean of the male and female balanced accuracies, so it need not equal the mean of the reported sensitivity and specificity.

Robustness

To confirm that the selected test lengths were not an artefact of one partition, the Discovery/Validation split was repeated 30 times under the same stratified rule, with the genetic algorithm run twice on each Discovery half (60 runs in total; Supplementary Table 3). Post hoc sensitivity analyses covered a 40-item comparator with a cut-off fitted on Discovery (Supplementary Table 4), complete cases only

(Supplementary Table 1), normosmic and hyposmic participants separately (Supplementary Table 5), clinic versus remote administration (Supplementary Table 6), cut-offs fitted on the full rather than the class-balanced Discovery cohort (Supplementary Table 7), age-adjusted and sensitivity-targeted cut-offs (Supplementary Table 8) and strictly baseline αsyn-SAA results (Supplementary Table 2). A wider local search at every length, started independently of the genetic algorithm, checked whether it had missed better tests (Supplementary Table 9). A further analysis was prompted by concerns about the cultural applicability of some odours. We repeated the genetic-algorithm selection with six North-American confectionery and food odours removed from the candidate pool. Stability was checked in the same 30 re-splits, using a local search at 9 odours in place of the genetic algorithm (Supplementary Tables 10–14).

Consent and ethical approval

All PPMI Clinical participants provided written informed consent to participate in the study. The PPMI study is conducted in accordance with the Declaration of Helsinki and the Good Clinical Practice guidelines, following approval of the local ethics committees of participating sites. The current PPMI Clinical study protocol (#002) received initial WCG approval (IRB Tracking #20200597) on April 20, 2020. The previous PPMI Clinical protocol (#001) received initial IRB approval on May 7, 2010, by the University of Rochester Research Subjects Review Board (RSRB #00031629) and was closed by the RSRB on March 9, 2021. Any questions pertaining to study compliance can be directed to The PPMI Data and Publications Committee (DPC): PPMI.Publication@indd.org.

Results

Baseline characteristics

See Table 1 for a summary of baseline characteristics. We included 1,117 hyposmia-enriched controls, 289 participants with iRBD, and 442 with PD; 63%, 70%, and 90% were αsyn-SAA-positive, respectively. Educational attainment was similar across cohorts, and participants with PD were on average about two years younger (mean 65.9 vs 68.2 and 67.7 years). The hyposmia-enriched control group featured a higher proportion of females (58%) than the iRBD (25%) or PD (36%) cohort.

Characteristic	Hyposmia-enriched controls (n=1,117)	iRBD (n=289)	PD (n=442)
Age			
Mean (SD)	68.2 (5.6)	67.7 (6.1)	65.9 (7.7)
Median (IQR)	68 (64–72)	68 (64–72)	66 (61–71)
Range	51–90	51–83	50–84
Age at PD diagnosis, mean (SD)	—	—	65.1 (7.8)
Sex			
Female, n (%)	651 (58.3)	71 (24.6)	161 (36.4)
Male, n (%)	466 (41.7)	218 (75.4)	281 (63.6)
Race*			
White, n (%)	1,031 (92.3)	267 (92.4)	407 (92.1)
Black, n (%)	17 (1.5)	5 (1.7)	9 (2.0)
Unknown, n (%)	14 (1.3)	4 (1.4)	4 (0.9)
Other, n (%)	9 (0.8)	2 (0.7)	8 (1.8)
Asian, n (%)	5 (0.4)	7 (2.4)	6 (1.4)
American Indian / Alaska Native, n (%)	4 (0.4)	1 (0.3)	1 (0.2)
≥1 race, n (%)	37 (3.3)	3 (1.0)	7 (1.6)
Ethnicity			
Not Hispanic or Latino, n (%)	1,078 (96.5)	271 (93.8)	417 (94.3)
Unknown or declined, n (%)	22 (2.0)	12 (4.2)	9 (2.0)
Hispanic or Latino, n (%)	17 (1.5)	6 (2.1)	16 (3.6)
Ancestry (self-reported, optional)			
Ashkenazi Jewish descent, n (%)	85 (7.6)	31 (10.7)	38 (8.6)
Basque descent, n (%)	3 (0.3)	1 (0.3)	3 (0.7)
African Berber descent, n (%)	2 (0.2)	0 (0.0)	0 (0.0)
Cognitive			
Education years, mean (SD)	16.8 (2.7)	16.7 (3.3)	16.7 (3.1)
MoCA total, mean (SD)	26.7 (2.3)	26.4 (2.6)	26.7 (2.3)
PD-associated features			
MDS-UPDRS III, mean (SD)	4.8 (6.2)	4.7 (5.4)	24.3 (10.0)
UPSIT score, mean (SD)	22.5 (7.6)	25.0 (8.7)	22.6 (7.8)
UPSIT score ≤15th percentile, n (%)	904 (80.9)	162 (56.1)	307 (69.5)
αsyn-SAA positive, n (%)	708 (63.4)	201 (69.6)	397 (89.8)

Table 1: Baseline characteristics of study participants. *Racial/ethnic/ancestry categories listed as they appear in study questionnaires. PD = Parkinson's disease; iRBD = Isolated REM sleep behaviour disorder; SD = Standard deviation; IQR = Interquartile range; MoCA = Montreal Cognitive Assessment; MDS-UPDRS = The Movement Disorder Society-Unified Parkinson's Disease Rating Scale; UPSIT = University of Pennsylvania Smell Identification Test; αsyn-SAA = α-synuclein seed amplification assay.

Test length vs performance

Table 2 shows the highest balanced accuracy achieved for candidate smell tests of 1–17 items in the hyposmia-enriched Discovery group. Performance dropped when the same tests were applied to the hyposmia-enriched Validation group, but was comparable in iRBD and PD.

For shorter tests, the number of odourants was clearly associated with balanced accuracy (see Figure 2 for a visual summary, odours are also listed in Supplementary Table 15). However, above 6 items a performance ceiling was evident, with additional odours yielding minimal improvement (≤2.5 percentage points). Starting independently of the genetic algorithm, the wider local search recovered the same tests at 1–9 odours and found no better 13-odour test; at some longer lengths it found tests marginally better on Discovery (by at most 0.5 points at the same length), with no consistent gain in the held-out cohorts (Supplementary Table S9).

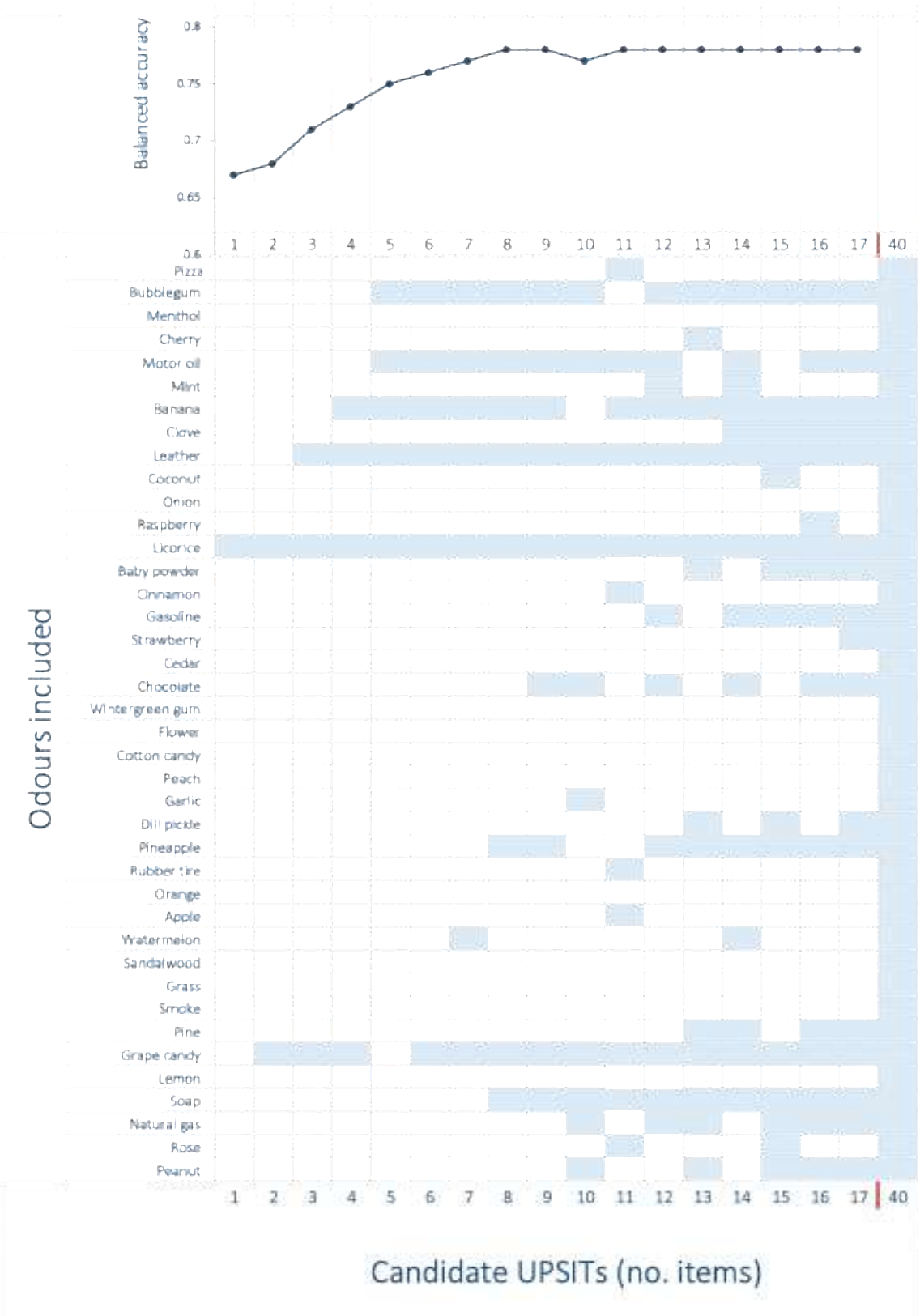

Figure 2: Length vs balanced accuracy in the Discovery cohort. Line graph shows the highest balanced accuracy achieved for candidate UPSITs of 1-17 items. In the grid, specific odours included are represented by a blue box. 11 and 12-item tests were found by local search as the algorithm produced none of these lengths that outperformed 10 items (Supplementary Table 9). UPSIT = University of Pennsylvania Smell Identification Test.

Abbreviated vs 40-item UPSIT

There exists no singular 'best' measure of test performance. Instead, we considered quantitative metrics (balanced accuracy, sensitivity, and specificity) alongside practical factors. For example, longer tests are more robust to quality-control issues or odour-familiarity differences, but shorter tests are cheaper and less burdensome. One 9-item combination (Bubble Gum, Motor Oil, Banana, Leather, Licorice, Chocolate, Pineapple, Grape Candy, Soap) represented a reasonable compromise. We will highlight and evaluate this as an illustrative example. However, other test lengths or cut-offs might be preferable, depending on the intended application.

In the Validation group - at a cut-off of ≤5 and ≤6 in males and females, respectively - the test identified 225 of 282 αsyn-SAA positive individuals (80% sensitivity), with comparatively low specificity (60%). When applied to the iRBD cohort, sensitivity dropped (68%), but specificity improved to 84%, giving false positive and negative rates of 16% and 32%, respectively. Sensitivity was higher than the 40-item test in PD (72% vs 67%), with a similar specificity (87% vs 89%). Likewise, given that the vast majority (90%) of the PD cohort were αsyn-SAA positive, pre-test probability was very high. This resulted in a favourable positive predictive value (98%). Conversely, about 3 in 4 of PD cases with a 'negative' smell test were αsyn-SAA positive (negative predictive value 26%).

See **Table 3** and **Figure 3** for a comparison of 9-item and 40-item test performance. The full UPSIT was 13.8, 6.0, and 4.3 percentage points more sensitive in hyposmia-enriched controls, iRBD, and PD, respectively, but 17.1 points less specific in hyposmia-enriched controls and equally specific in iRBD and PD. Balanced accuracy was similar; 0.69 vs 0.68 in hyposmia-enriched controls, 0.78 vs 0.81 in iRBD and 0.79 vs 0.82 in PD. AUC was 0.75 vs 0.77, 0.83 vs 0.87 and 0.87 vs 0.90, respectively. The difference in AUC reached significance only in iRBD ($p<0.001$). Estimates with 95% confidence intervals and the paired comparisons are provided in **Supplementary Table 16**. Against a 40-item cut-off fitted on

Discovery in the same way as the 9-item test's (≤22 men, ≤28 women), the 9-item Validation balanced accuracy was lower (0.69 vs 0.72). A 13-item alternative test returned 0.01, 0.03 and 0.01 of the full UPSIT's AUC in Validation, iRBD and PD (Supplementary Table 4).

Across 30 re-splits of the hyposmia-enriched control cohort (Supplementary Table 3), 9-item tests reached a median Validation balanced accuracy of 0.69 (interquartile range 0.68–0.70); the main split's 0.69 is therefore typical. Their median loss in AUC against the full UPSIT was 0.02 in Validation, 0.03 in iRBD and 0.03 in PD. The main split is unfavourable to the headline test in iRBD (loss 0.05, against a median of 0.03 across re-splits). This partly explains the apparent iRBD advantage of the shorter test described below.

UPSIT items (n)	Cut-off (≤)		Hyposmia-enriched controls				iRBD			PD		
			Discovery	Validation								
	Male	Female	Balanced acc.	Balanced acc.	Sensitivity	Specificity	Balanced acc.	Sensitivity	Specificity	Balanced acc.	Sensitivity	Specificity
1	0	0	0.67	0.66	0.72	0.65	0.69	0.60	0.82	0.75	0.68	0.80
2	1	1	0.68	0.65	0.83	0.55	0.71	0.70	0.76	0.78	0.80	0.76
3	1	2	0.71	0.69	0.79	0.56	0.71	0.55	0.86	0.77	0.66	0.84
4	2	2	0.73	0.66	0.78	0.60	0.73	0.63	0.81	0.77	0.73	0.80
5	2	2	0.75	0.65	0.61	0.72	0.72	0.52	0.89	0.73	0.57	0.89
6	3	3	0.76	0.66	0.72	0.66	0.74	0.59	0.86	0.76	0.66	0.87
7	3	4	0.77	0.70	0.76	0.65	0.75	0.55	0.90	0.78	0.65	0.89
8	4	5	0.78	0.69	0.77	0.63	0.76	0.60	0.86	0.78	0.68	0.87
9	5	6	0.78	0.69	0.80	0.60	0.78	0.68	0.84	0.79	0.72	0.87
10	6	7	0.77	0.69	0.78	0.61	0.77	0.61	0.88	0.76	0.66	0.87
11	6	7	0.78	0.67	0.87	0.52	0.78	0.71	0.83	0.80	0.73	0.84
12	6	8	0.78	0.68	0.77	0.60	0.79	0.60	0.90	0.77	0.65	0.87
13	7	9	0.78	0.70	0.82	0.60	0.78	0.62	0.88	0.81	0.71	0.89
14	8	10	0.78	0.70	0.84	0.58	0.83	0.71	0.89	0.81	0.73	0.87
15	8	10	0.78	0.71	0.84	0.59	0.80	0.66	0.89	0.80	0.72	0.87
16	9	11	0.78	0.69	0.78	0.60	0.80	0.65	0.88	0.79	0.69	0.89
17	10	12	0.78	0.69	0.82	0.58	0.82	0.69	0.88	0.80	0.72	0.89

Bubble Gum, Motor Oil, Banana, Leather, Licorice, Chocolate, Pineapple, Grape Candy, Soap

Table 2: Test length vs. performance. Candidate UPSITs were developed in the hyposmia-enriched Discovery cohort. For each test length, the highest-performing test by balanced accuracy was evaluated in the Validation, iRBD and PD cohorts. Balanced accuracy, sensitivity and specificity were calculated using optimal sex-specific cut-offs. A test is positive when the number of correctly identified odours is at or below the sex-specific cut-off. The 9-item (red border) represents a reasonable compromise between length and balanced accuracy. 11 and 12-item tests were found by local search as the algorithm produced none of these lengths that outperformed 10 items (Supplementary Table 9). Balanced Acc. = balanced accuracy (mean of the male and female balanced accuracies); PD = Parkinson's disease; iRBD = Isolated REM sleep behaviour disorder; UPSIT = University of Pennsylvania Smell Identification Test.

UPSIT version	Cut-off (≤)		Hyposmia-enriched controls				iRBD			PD		
			Discovery	Validation								
	Male	Female	Balanced acc.	Balanced acc.	Sensitivity	Specificity	Balanced acc.	Sensitivity	Specificity	Balanced acc.	Sensitivity	Specificity
9-item 'headline'	5	6	0.78	0.69	0.80	0.60	0.78	0.68	0.84	0.79	0.72	0.87
9-item 'internationally portable'	5	6	0.75	0.69	0.83	0.59	0.84	0.75	0.85	0.81	0.71	0.89
13-item	7	9	0.78	0.70	0.82	0.60	0.78	0.62	0.88	0.81	0.71	0.89
Revised 40-item	22	28	0.70	0.71	0.83	0.60	0.79	0.63	0.90	0.78	0.67	0.89
	≤15th percentile		0.67	0.68	0.94	0.43	0.81	0.74	0.84	0.82	0.76	0.87

Table 3: Novel abbreviated UPSITs vs. 40-item Revised version. 9-item 'headline' test was selected as a reasonable compromise between length and test performance. The 'internationally portable' test was generated by repeating the analysis with six North-American odours removed from the candidate pool. 13-odour test is the shortest test within 0.001 of the highest Discovery balanced accuracy (0.781, against 0.782 for the 17-odour test). The first three rows use optimal sex-specific cut-offs calculated in discovery cohort. 40-item performance is also measured using age & sex specific cut-offs derived from normative data (Brumm et al. 2023)[12], where ≤15th percentile is considered hyposmic. PD = Parkinson's disease; iRBD = Isolated REM sleep behaviour disorder; UPSIT = University of Pennsylvania Smell Identification Test; B-SIT = Brief Smell Identification Test.

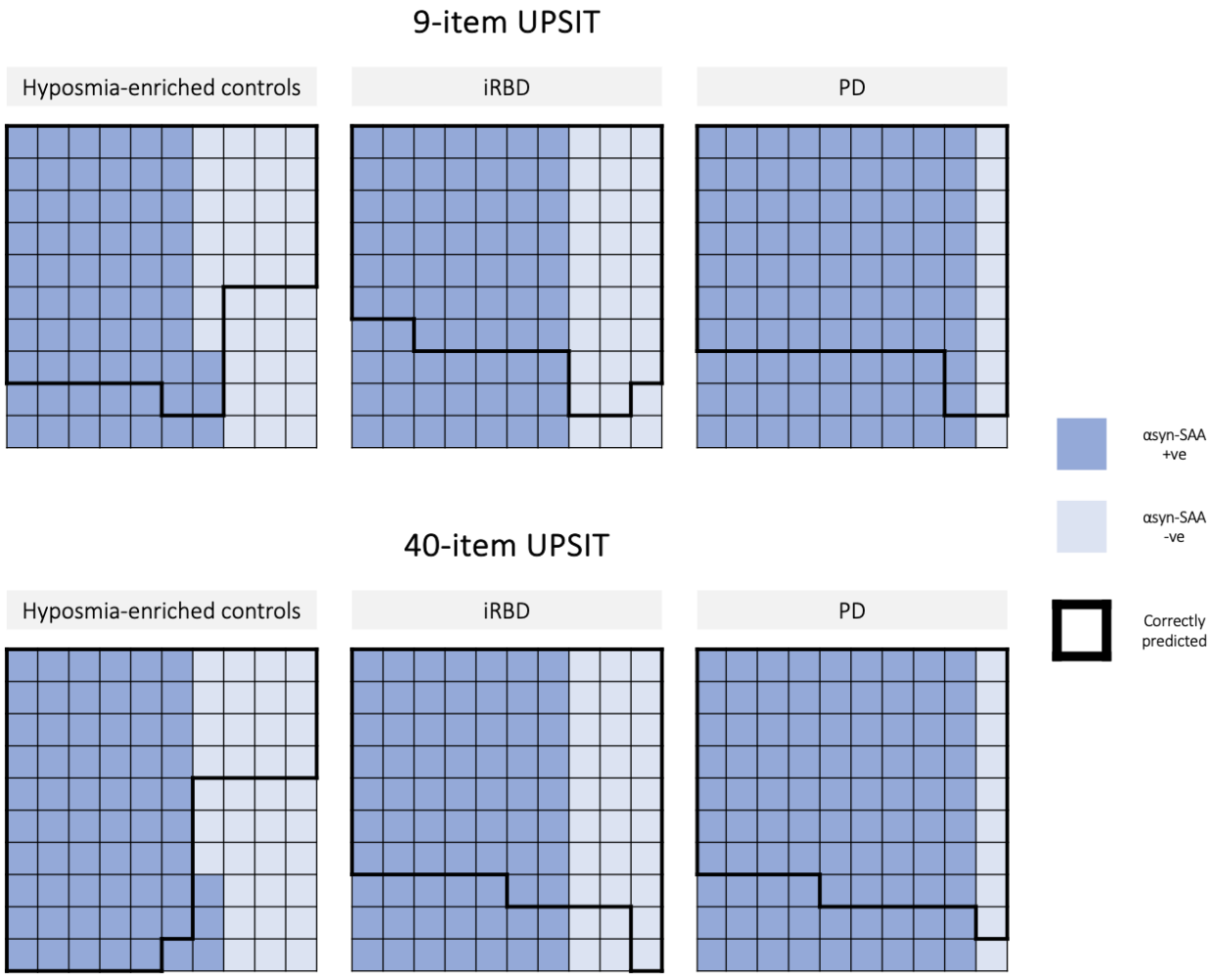

Figure 3: Comparison of 40-item and 9-item UPSIT performance for predicting αsyn-SAA status. Each 10x10 grid represents a study group; hyposmia-enriched control validation cohort (n=446), iRBD (n=289) and PD (442). One small square = 1% of participants in that group. Dark and light blue shading indicate αsyn-SAA positive and negative participants, respectively. The black border contains the proportion correctly classified by each test. PD = Parkinson's disease; iRBD = Isolated REM sleep behaviour disorder; UPSIT = University of Pennsylvania Smell Identification Test.

Odour choice

Licorice and Grape Candy were chosen most consistently (82% and 77%, respectively for 8–10-odour tests; no other odour exceeded 60%; Supplementary Table 3). The remaining odours varied between runs that performed similarly, so the specific selection beyond these two should not be over-interpreted.

Odour recognisability

Three of the nine odours — Bubble Gum, Licorice and Grape Candy — are North-American confectionery flavours. Repeating the selection with these (plus three further North-American odours) removed from the candidate pool produced an alternative, internationally 'portable' 9-item test (Cherry, Motor Oil, Mint, Banana, Clove, Leather, Apple, Soap, Rose) with the same cut-offs. Across 30 re-splits of the hyposmia-enriched control cohort, with the selection repeated by local search, the constraint cost a median of 1.1 percentage points of Discovery balanced accuracy and made no material difference out of sample (median paired difference in AUC −0.0004 in held-out hyposmia-enriched controls, −0.0006 in iRBD and +0.007 in PD). On the main split, balanced accuracy was 0.75 in Discovery, 0.69 in Validation, 0.84 in iRBD and 0.81 in PD. The apparent iRBD advantage over the headline test (difference in AUC +0.037) did not replicate; across re-splits the 'portable' test was better than iRBD in 13 of 30 and never by as much (Supplementary Tables 10–14).

Discussion

To our knowledge, this is the first study evaluating the use of shortened smell tests to predict αsyn-SAA status; previous analyses have focused on differentiating people with PD from healthy controls[17]. Our results suggest that, for αsyn-SAA prediction, the 40-item Revised UPSIT can be abbreviated to 9 items without losing specificity. The cost is sensitivity, most clearly in hyposmia-enriched controls, where it is offset by higher specificity. Balanced accuracy and discrimination remained close to those of the full test in all three groups.

Performance across clinical cohorts

Our selected groups are, to an extent, reflective of the disease course, where some individuals progress from non-specific abnormalities (smell loss), through high-risk prodromal stages (iRBD), eventually converting to manifest PD[7]. At each stage, the association between hyposmia and αsyn-SAA status is strengthened. Poor smell is a common complaint, affecting roughly 13.5% of adults above 40 years and rising sharply thereafter[18]. By contrast, αsyn-SAA positivity is uncommon in the general population (2-6%[6,11,13]). Therefore, most cases of idiopathic hyposmia are presumably unrelated to synucleinopathy. In other words, for neurologically healthy people, smell tests will generate numerous false-positives if used to predict αsyn-SAA positivity. Instead, they function more appropriately as a 'rule out' test: at the 2-6% general-population prevalence cited above, the 9-item test would carry a negative predictive value of 98-99% (against 94-98% without testing) but a positive predictive value of only 4-11%. When applied to groups with higher rates of αsyn-SAA positivity (iRBD → PD), the likelihood of detected hyposmia being synucleinopathy-associated increases and specificity rises, from 60% in hyposmia-enriched controls to 84% in iRBD and 87% in PD. This reflects differences in group composition rather than prevalence itself, as specificity does not depend on prevalence. Additionally, Siderowf et al. recently showed that the majority of *LRRK2* carriers with normal olfaction are αsyn-SAA negative[6]. Conversely, sporadic PD cases with concurrent hyposmia are significantly more likely to test αsyn-SAA

positive than any other subgroup[6,11]. Our results are consistent with smell testing being a useful proxy for characterising clinical phenotypes, which may contribute to its high specificity in manifest disease, albeit specificity in PD rests on only 45 αsyn-SAA-negative participants.

Though the 9-item test was reasonably sensitive in hyposmia-enriched controls, drops of 12 and 8 percentage points were observed in iRBD and PD, respectively. This is presumably a function of clinicopathological variation. PD is a heterogenous disease and α-synuclein-associated neurodegeneration does not always involve olfactory structures[9]. Accordingly, a substantial minority of participants in the iRBD or PD cohorts were both normosmic (>15th percentile) and αsyn-SAA positive; 18% and 22%, respectively (Supplementary Table 5). When smell loss is absent or mild, these participants are misclassified and sensitivity declines.

Future applications

PD researchers are increasingly using multi-step recruitment initiatives to define & enrich study cohorts[19]. Typically, protocols entail (i) scalable, non-invasive assessments (i.e. questionnaires, smell tests) followed by (ii) invasive, resource-intensive investigations (i.e. lumbar puncture or neuroimaging)[10,20,21]. Recently, the PPMI Clinical protocol has been revised such that a 'pre-screening' UPSIT constitutes the first stage of recruitment. Hyposmic participants then undergo CSF collection and those deemed αsyn-SAA positive are considered eligible for enrolment.

We propose this could be streamlined by abbreviating the 40-item UPSIT to 9 odours. Sensonics currently market similar abbreviated tests at a lower unit cost; $9.95 vs $29.50 for the full UPSIT[22]. For a hypothetical study in which 1,000 participants undertake a shortened smell test, we can assume cost savings of approximately $19,550, in addition to reduced import fees and postage expenses. In terms of performance, a 9-item test incurs up to ~14 percentage points lower sensitivity than the full UPSIT (depending on cohort), with an equal or lower false positive rate. Moreover, shorter tests are

considerably less burdensome for participants. Previous studies have concluded as few as 4-6 items are satisfactory for hyposmia detection[21,23,24]. However, slightly longer tests might provide necessary redundancy if certain odours are not universally recognisable. Three of our nine odours are derived from North-American confectionery flavours. Whether they are equally recognisable elsewhere is unknown; PPMI does not report country of residence in the data released for this analysis and the test-language field is recorded for only 48 of 1,117 hyposmia-enriched controls. We therefore report an internationally 'portable' alternative that excludes them, at a cost of roughly one percentage point of balanced accuracy and no measurable loss out of sample. Future analyses should attempt to validate this portable set (Cherry, Motor Oil, Mint, Banana, Clove, Leather, Apple, Soap, Rose) in non-North-American populations.

The value of detecting hyposmia in healthy people depends upon the intended application. For large-scale observational studies, smell tests can effectively 'narrow down' a study population to those at increased risk of synucleinopathy. However, poor olfaction alone is not sufficiently specific for synucleinopathy to justify potentially harmful preventative interventions such as pharmaceuticals, especially given the proportion of α-syn-SAA positive people that ultimately develop PD is not well established. Even for low-risk interventions, the ethical implications are significant. Overstatement of PD risk based solely on hyposmia can cause unnecessary psychological distress and mitigation strategies should be incorporated into the study design. For example, Slow-SPEED-UK[25] - an exercise-based PD prevention trial - has employed an innovative approach in which participants receive PD risk counselling before they consent to smell testing.

In iRBD and PD, abbreviated UPSITs reliably identify individuals with a high probability of αsyn-SAA positivity; 91% and 98% positive predictive value, respectively. This largely reflects the high prevalence in these cohorts (70% and 90%), and the full UPSIT performs similarly (91% and 98%). Such performance

is compelling if the test were used to enrich a cohort for αsyn-SAA positivity. The capacity for smell testing to predict genetic subtype warrants further exploration.

Limitations

There are methodological reasons why the reported magnitude of the association between hyposmia and αsyn-SAA positivity may not generalise to an unselected population. αsyn-SAA negative individuals are under-represented in our study. The PPMI Clinical protocol was revised in 2020 to include a prodromal cohort (hyposmia, iRBD and/or genetic risk factors). Initially, prospective hyposmic participants were excluded at screening if they demonstrated a normal or near-normal DAT-SPECT scan - a factor correlated to αsyn-SAA negativity. A lowest putamen binding ratio of <80% of the age -and sex -expected value was required for inclusion; in May 2022 this was broadened to ≤100%. The most recent Amendment 4 has removed DAT-SPECT scanning from the screening process entirely[10]. Instead, CSF collection is now performed early, allowing for preferential selection of αsyn-SAA positive individuals for full enrolment[26]. Moreover, in 2023, the threshold for categorising participants as 'hyposmic' was lowered from ≤15th to ≤10th age-and sex-expected percentile, as severe hyposmia was found to be strongly associated with αsyn-SAA positivity[10].

The precise impact of these measures is hard to quantify as they have not been applied universally or consistently. PPMI protocols are flexible and have shifted over the 16-year study period to enable recruitment of a heterogenous – but not necessarily representative – population. Nonetheless, inevitable selection bias will have inflated the association between poor smell test performance and αsyn-SAA positivity.

Ideally, future analyses would include an unselected cohort, where hyposmia and αsyn-SAA positivity occur in proportion to the general population. The low prevalence of these biomarkers necessitates a large sample size. Such studies are in the pipeline; notably AccessBrainHealth[27] recently commenced

recruitment of approximately 1,000 healthy volunteers >60 years old. Participants will undergo annual assessments including smell testing and CSF collection. It will be valuable to assess the performance of abbreviated tests in this 'real-world' context, especially on a longitudinal scale. Finally, alongside the sex-specific cut-offs, we also tested an age-adjusted threshold. These did not improve performance, but nevertheless confirmation in independent normative datasets would be valuable.

Data availability

Data used in the preparation of this article were obtained from the 19 May 2026 release of the Parkinson's Precision Medicine Initiative (PPMI) database (https://www.ppmi-info.org/access-data-specimens/download-data), RRID:SCR_006431. For up-to-date information on the study, visit http://www.ppmi-info.org.

Code availability

All analysis code is available at https://github.com/Malta-Lab/short-upsit-saa. It reproduces every result in this paper from the PPMI data, which approved researchers can access.

Acknowledgments

PPMI – a public-private partnership – is funded by the Michael J. Fox Foundation for Parkinson's Research and funding partners, including AbbVie, Alamar Biosciences, Aligning Science Across Parkinson's (ASAP), Arrowhead Pharma, Arvinas, AskBio, BIAL, BioArctic, Biohaven, BlueRock Therapeutics, Bristol Myers Squibb, Calico Labs, Capsida Biotherapeutics, Critical Path Institute, DaCapo Brainscience, Denali, Edmond J. Safra Foundation, Eli Lilly, Gain Therapeutics, GE Healthcare, Genentech, GSK, Insitro, Johnson & Johnson Innovative Medicine, Lundbeck, Merck, Neumora, Neuron23, Novartis, Olink, Regeneron, Roche, Sanofi, Tenvie, UCB, Vanqua Bio, Voyager Therapeutics, The Weston Family Foundation. The Centre for Preventive Neurology is funded by the Bart's Charity.

The Centre for Preventive Neurology receives income from the Michael J Fox Foundation for its role as a PPMI study site. The funder played no role in study design, data collection, analysis and interpretation of data, or the writing of this manuscript.

Author contributions

RR, EM and CM developed the research question and study design together. EM contributed the background research, diagrams and write-up. RR & CM performed all statistical analysis. RR, CM, SIM, RCB & AJN provided feedback on multiple drafts. All authors reviewed and edited the final manuscript prior to submission.

Competing interests

Authors RR, EM, CM, SIM & RCB declare no financial or non-financial competing interests. AJN reports grants from Parkinson's UK, Barts Charity, Cure Parkinson's, National Institute for Health and Care Research, Innovate UK, UKRI, HORIZON 2020, the Medical College of Saint Bartholomew's Hospital Trust, Alchemab, and the Michael J Fox Foundation. AJN reports consultancy fees from AbbVie, Bial, and Roche. AJN receives a travel stipend from TNC. AJN reports honoraria from the International Parkinson and Movement Disorders Society.

Abbreviations

αsyn-SAA	α-synuclein seed amplification assay
AUC	Area under the curve
CSF	Cerebrospinal fluid
iRBD	Isolated rapid eye movement sleep behaviour disorder
PD	Parkinson's disease
PPMI	Parkinson's Precision Medicine Initiative

References

1. Kuang, Y.*, et al.* α-Synuclein seeding amplification assays for diagnosing synucleinopathies: an innovative tool in clinical implementation. *Transl Neurodegener* **13**, 56 (2024).
2. Mastenbroek, S.E.*, et al.* Disease progression modelling reveals heterogeneity in trajectories of Lewy-type α-synuclein pathology. *Nat. Commun.* **15**, 5133 (2024).
3. Braak, H.*, et al.* Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiol Aging* **24**, 197-211 (2003).
4. Concha-Marambio, L., Pritzkow, S., Shahnawaz, M., Farris, C.M. & Soto, C. Seed amplification assay for the detection of pathologic alpha-synuclein aggregates in cerebrospinal fluid. *Nat Protoc* **18**, 1179-1196 (2023).
5. Fairfoul, G.*, et al.* Alpha-synuclein RT-QuIC in the CSF of patients with alpha-synucleinopathies. *Ann Clin Transl Neurol* **3**, 812-818 (2016).
6. Siderowf, A.*, et al.* Assessment of heterogeneity among participants in the Parkinson's Progression Markers Initiative cohort using α-synuclein seed amplification: a cross-sectional study. *Lancet Neurol* **22**, 407-417 (2023).
7. Simuni, T.*, et al.* A biological definition of neuronal α-synuclein disease: towards an integrated staging system for research. *Lancet Neurol* **23**, 178-190 (2024).
8. Höglinger, G.U.*, et al.* A biological classification of Parkinson's disease: the SynNeurGe research diagnostic criteria. *Lancet Neurol* **23**, 191-204 (2024).
9. Doty, R.L. Olfaction in Parkinson's disease and related disorders. *Neurobiol Dis* **46**, 527-552 (2012).
10. Brown, E.G.*, et al.* Staged Screening Identifies People with Biomarkers Related to Neuronal Alpha-Synuclein Disease. *Ann Neurol* **97**, 730-740 (2025).
11. Venuto, C.S., Herbst, K., Chahine, L.M. & Kieburtz, K. Predicting Cerebrospinal Fluid Alpha-Synuclein Seed Amplification Assay Status from Demographics and Clinical Data. *medRxiv* (2025).
12. Brumm, M.C.*, et al.* Updated percentiles for the University of Pennsylvania smell identification test in adults 50 years of age and older. *Neurology* **100**, e1691-e1701 (2023).
13. Coughlin, D.G.*, et al.* α-Synuclein Seed Amplification Assay Amplification Parameters and the Risk of Progression in Prodromal Parkinson Disease. *Neurology* **104**, e210279 (2025).
14. Marek, K.*, et al.* Widespread synuclein pathology in hyposmics precedes dopamine transporter deficit in PARS. *Mov. Disord.* **38 (supp 1)**, S105-S105 (2023).
15. Deb, K., Pratap, A., Agarwal, S. & Meyarivan, T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* **6**, 182-197 (2002).
16. Fortin, F.-A., De Rainville, F.-M., Gardner, M.-A.G., Parizeau, M. & Gagné, C. DEAP: Evolutionary algorithms made easy. *JMLR* **13**, 2171-2175 (2012).
17. Mitchell, E.*, et al.* Hyposmia in Parkinson’s disease; exploring selective odour loss. *npj Parkinson's Disease* **11**, 67 (2025).

18. Liu, G., Zong, G., Doty, R.L. & Sun, Q. Prevalence and risk factors of taste and smell impairment in a nationwide representative sample of the US population: a cross-sectional study. *BMJ open* **6**, e013246 (2016).
19. Virameteekul, S*., et al.* PREDICT-PD: A Two-Stage Approach to Early Identification of Parkinson's Disease. *Mov Disord Clin Pract* **13**, 2421-2427 (2026).
20. Flores-Torres, M.H*., et al.* Identifying Individuals in the Prodromal Phase of Parkinson's Disease: A Prospective Cohort Study. *Ann. Neurol.* **97**, 720-729 (2025).
21. Bestwick, J.P*., et al.* Optimising classification of Parkinson's disease based on motor, olfactory, neuropsychiatric and sleep features. *npj Parkinson's Dis.* **7**, 87 (2021).
22. Sensonics International, Smell Identification Test™ (UPSIT®). https://sensonics.com/product/smell-identification-test/. (2024).
23. Joseph, T*., et al.* Screening performance of abbreviated versions of the UPSIT smell test. *J Neurol* **266**, 1897-1906 (2019).
24. Auger, S.D*., et al.* Testing Shortened Versions of Smell Tests to Screen for Hyposmia in Parkinson's Disease. *Mov Disord Clin Pract* **7**, 394-398 (2020).
25. Oosterhof, T.H*., et al.* Slow-SPEED: protocol for three randomised trials of remotely delivered exercise to prevent Parkinson's disease. *npj Parkinson's Dis.* (2026).
26. Kenneth Marek 2024. The Parkinson's Progression Marker Initiative (PPMI) Clinical - Establishing a Deeply Phenotyped PD Cohort AM 4. protocols.io.
27. AccessBrainHealth Study, https://accessall.umed.io/. (2026).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarymaterial.pdf
- Figures.pdf