Predicting Cerebrospinal Fluid Alpha-Synuclein Seed Amplification Assay Status from Demographics and Clinical Data

Charles S. Venuto, PharmD[1,2] (ORCID: 0000-0002-2394-9513); Konnor Herbst, MA[1] (ORCID: 0000-0002-3629-9337); Lana M. Chahine, MD[3] (ORCID: 0000-0003-2521-7196); Karl Kieburtz, MD, MPH[1,2] (ORCID: 0000-0002-1645-6921)

[1]Center for Health + Technology, University of Rochester, Rochester, NY, USA

[2]Department of Neurology, University of Rochester, Rochester, NY, USA

[3]Department of Neurology, University of Pittsburgh, Pittsburgh, PA, USA

Correspondence to: Charles S. Venuto, Center for Health + Technology, University of Rochester, 265 Crittenden Blvd., Rochester, NY 14642; Telephone: 585-276-5622; E-mail: Charles.Venuto@chet.rochester.edu

ABSTRACT

Objective: To develop and externally validate models to predict probabilities of alpha-synuclein (a-syn) positive or negative status in vivo in a mixture of people with and without Parkinson's disease (PD) using easily accessible clinical predictors.

Methods: Uni- and multi-variable logistic regression models were developed in a cohort of participants from the Parkinson Progression Marker Initiative (PPMI) study to predict cerebrospinal fluid (CSF) a-syn status as measured by seeding amplification assay (SAA). Models were externally validated in a cohort of participants from the Systemic Synuclein Sampling Study (S4) that had also measured CSF a-syn status using SAA.

Results: The PPMI model training/testing cohort consisted of 1260 participants, of which 76% had manifest PD with a mean (± standard deviation) disease duration of 1.2 (±1.6) years. Overall, 68.7% of the overall PPMI cohort (and 88.0% with PD of those with manifest PD) had positive CSF a-syn SAA status results. Variables from the full multivariable model to predict CSF a-syn SAA status included age- and sex-specific University of Pennsylvania Smell Identification Test (UPSIT) percentile values, sex, self-reported presence of constipation problems, leucine-rich repeat kinase 2 (*LRRK2*) genetic status and pathogenic variant, and *GBA* status. Internal performance of the model on PPMI data to predict CSF a-syn SAA status had an area under the receiver operating characteristic curve (AUROC) of 0.920, and sensitivity/specificity of 0.881/0.845. When this model was applied to the external S4 cohort, which included 71 participants (70.4% with manifest PD for a mean 5.1 (±4.8) years), it performed well, achieving an AUROC of 0.976, and sensitivity/specificity of 0.958/0.870. Models using only UPSIT percentile performed similarly well upon internal and external testing.

Conclusion: Data-driven models using non-invasive clinical features can accurately predict CSF a-syn SAA positive and negative status in cohorts enriched for people living with PD. Scores from the UPSIT were highly significant in predicting a-syn SAA status.

INTRODUCTION

The introduction of the α-synuclein (a-syn) cerebrospinal fluid (CSF) seeding amplification assay (SAA) offers a promising approach for the diagnosis of Parkinson's disease (PD) and other synucleinopathies. This method is notable for its high sensitivity and specificity in distinguishing between people with and without PD.[1,2] Additionally, there is an emerging emphasis on a-syn SAA positive results to redefine and stage PD biologically.[3-5] These approaches promote a shift away from solely relying on symptom-based diagnostics, and instead focus towards a framework that integrates biological markers of a-syn aggregation, thus enhancing the potential for earlier detection and intervention.

However, the implementation of CSF a-syn SAA is not without its challenges. In particular, the assay requires a spinal tap for the collection of CSF samples, an invasive procedure that may limit its routine clinical use. Therefore, efforts are ongoing to extend SAA applicability to other more accessible fluids and tissues. For instance, detection of a-syn in skin biopsy samples has shown good performance similar to those using CSF samples for differentiating PD from controls.[6-10] However, fewer studies have evaluated skin biopsies, and technical aspects such as number of biopsies, sampling locations, and sample processing still need to be optimized. Additional biospecimens, such as saliva, blood, submandibular gland, and colonic mucosa are still under development and are not as readily available.[11-14] Therefore, at present, CSF remains the leading biomatrix for the detection of a-syn using SAA.

Intriguingly, in the largest study to date that measured CSF a-syn via SAA in a predominantly PD population, the proportion of individuals yielding a positive CSF a-syn SAA result was not uniform. For instance, only 67.5% of individuals with PD carrying genetic mutations in the leucine-rich repeat kinase 2 (*LRRK2*) gene were positive for a-syn CSF SAA compared to 93.3% of sporadic PD individuals. For PD participants without hyposmia, the sensitivity of the a-syn SAA assay was only 63.0%.[15] This variation when stratified by genetic and clinical characteristics, points to a nuanced relationship between select clinical features and the disease pathology. We aimed to leverage this variation by developing data-driven models that can predict CSF a-syn SAA status probabilities based on simple, non-invasively collected assessments.

METHODS

Study Data

Model training data came from the Parkinson's Progression Marker Initiative (PPMI) study cohort. PPMI is a longitudinal observational study that is recruiting participants from outpatient neurology clinics at academic centers in North America and Europe (registered at ClinicalTrials.gov NCT01141023).[16] The objective of the PPMI study is to identify clinical and biological markers of PD risk, onset, and progression. Enrollment began in 2010 and is ongoing. Participants in PPMI were included in one of the following cohorts: untreated sporadic PD; PD with presence of pathogenic genetic variants (*LRRK2*, *GBA*, *SNCA*, *Parkin*, *PINK1*); at-risk non-manifest PD based on presence of rapid eye movement (REM) sleep behavior disorder (RBD), hyposmia with dopamine transporter (DAT) deficit measured using single photon emission computed tomography (SPECT), or presence of a genetic risk variant; 'healthy controls' who have no neurologic disorder and no first-degree relative with PD; and, people with parkinsonism but with DAT SPECT scans without evidence of dopamine deficiency (SWEDD). The diagnosis for each cohort was made by site investigators who are movement disorder specialists. Data used in preparation for this article were obtained on February 13th, 2024 from the PPMI database (http://www.ppmi-info.org/access-data-specimens/download-data), RRID:SCR_006431. For up-to-data information on the study, visit http://www.ppmi-info.org.

External validation data came from the Systemic Synuclein Sampling Study (S4) cohort. S4 was a cross-sectional study evaluating a-syn measurement in tissue and biofluid samples among people with and without PD. Enrollment occurred between October 2015 and August 2017 at six sites in North America. The PD group included those with a decreased striatal dopamine transporter binding on SPECT (based on age-matched normative data), while the non-PD ('healthy control') group had normal DAT SPECT. Additional details about the S4 study design have been described elsewhere.[17,18]

a-Syn Seed Amplification Assay (SAA)

a-Syn detection in the PPMI study was measured in CSF samples using the Amprion a-syn SAA, details of which have been previously described.[15,19] Initial a-syn SAA testing in PPMI

occurred in 2022 using an assay with a runtime length of 150 hours to produce aggregation curves for kinetic analysis of a-syn (PPMI biospecimen analysis project 155). In 2023, a-syn testing in PPMI was conducted using a more optimized assay that had a reduced runtime length of 24 hours and a reduced required volume of recombinant a-syn (PPMI biospecimen analysis project 237). For both assay versions, sample reactions were run in triplicate, and a final qualitative determination of being a-syn SAA positive, negative, or inconclusive was determined using maximum fluorescence (F_{max}) data from all three replicates as described previously.[15,19] Briefly, a sample was considered a-syn positive when all three replicates were positive according to a probabilistic algorithm based on F_{max}. A sample was considered a-syn negative when zero or one replicates were positive, and was inconclusive when two replicates were positive. There were 1384 out of 3874 PPMI participants with CSF a-syn SAA data at the time of download. Data were excluded from the analysis for participants who had inconclusive results, or with re-tested samples showing inconsistent a-syn SAA results.

In the S4 study, cyclic amplification of protein misfolding adapted for a-syn was performed using CSF samples as previously described.[20] Samples were run in triplicate. For each participant, the average F_{max} value from three samples was calculated to classify a-syn SAA status as positive or negative. Receiver-operating curve analysis identified an optimal F_{max} cutoff of $\geq$ 1375 for SAA status, based on the maximum Youden Index for differentiating PD from healthy controls.[21] There were 75 out of 82 S4 participants with CSF a-syn SAA data.

Clinical Demographics and Assessments Included in Analyses

To predict a-syn SAA positive versus negative status, we selected clinical and genetic variables that demonstrated subgroup differences in SAA status according to prior analysis.[15] These variables included presence or absence of *LRRK2* pathogenic variants (regardless of PD manifestation) and the *LRRK2* variant type located within the kinase domain (G2019S or I2020T) versus the Ras of complex (ROC) proteins domain (R1441G/C or N1437H); any *GBA* pathogenic variant; University of Pennsylvania Smell Identification Test (UPSIT) scores; and, age and sex. We also evaluated constipation symptoms as a variable, which is a symptom that has been suggested to at least partially stem from a-syn accumulation in myenteric neurons.[22,23]

Constipation was measured by the Scale for Outcomes in PD – Autonomic Dysfunction constipation symptom (SCOPA-AUT5) categorized as never/sometimes or regular/often.

UPSIT data were utilized in multiple formats with varying criteria for missingness among the forty different scent items to assess the impact on modeling performance. First, we evaluated UPSIT data as age- and sex-specific percentiles, as described previously.[24] This approach included participants missing at most one of the forty scent items. Missing responses were imputed as one, assuming correct identification of the scent; analyses were repeated with an imputation as zero, assuming an incorrect response. Results using an imputation of zero were nearly identical to those using an imputation of one, and therefore are not reported. We also used UPSIT data in a more general approach by calculating the percentage of correctly answered scent items without age and sex-specific scores. This UPSIT percent correct method included participants who completed more than twenty of the forty individual scent test items, and was calculated by dividing the total number of correct responses by the total number of items answered. Results for the UPSIT percent correct approach are provided in the Appendix.

Statistical Analysis

Logistic regression models were developed using PPMI data that included participants from all cohorts (as described above) to predict the probability of SAA a-syn positive status. Candidate models were evaluated as follows: a null model without covariates; a baseline model that uses only UPSIT data; partially adjusted models that include UPSIT data plus age and/or sex; another partially adjusted model including UPSIT data, sex (or age), and constipation history (SCOPA-Aut item 5); and, a fully adjusted model that includes UPSIT data, sex (or age), constipation history, *LRRK2* pathogenic variant type, and *GBA* status. The relative goodness of fit between the nested models was assessed using Likelihood Ratio Chi-Squared tests. A significance level of $p < 0.05$ was used to determine whether the inclusion of additional predictors significantly improved the model fit. Regression coefficients from each logistic model and their odds ratios are reported to indicate the relative risk of SAA positive status associated with each predictor.

The predictive performance of the models was first examined through internal validation using area under the Receiver Operating Characteristic curve (AUROC). The optimal probability cut-off for defining SAA status for each model was determined using the Youden Index, maximizing the sum of sensitivity and specificity, from the internally trained models on PPMI data. Confusion matrices corresponding to the cut-offs were generated, and metrics such as sensitivity, specificity, accuracy, positive predictive value, and negative predictive value were calculated.

For internal validation, an optimism bootstrap validation approach on the PPMI dataset was implemented. We resampled, refit, and reevaluated on 10000 bootstrap re-samples. For external validation, the logistic regression model trained on the full PPMI dataset was subsequently applied to the independent S4 data cohort. Performance metrics in the external S4 sample were assessed to evaluate how well the models built on the PPMI cohort generalized to an external cohort that also consisted of participants with PD and healthy controls.

The models were also evaluated for their ability to predict CSF a-syn SAA status when considering only participants with manifest PD to determine their capabilities in a PD-specific setting. For these analyses, model performance metrics were re-calculated as previously described but when considering only those individuals who had a PD diagnosis at enrollment in either PPMI or S4.

Descriptive statistics were calculated for PPMI and S4 cohorts. Categorical variables were compared using chi-squared test of independence, and numeric variables were compared using Welch's two-sample t-tests. Analyses were conducted in R 4.3.1 (packages tidyverse, pROC, and rms).

RESULTS

Model Development

Among the 1384 participants in the PPMI study who had CSF a-syn SAA status assayed, 1260 (91.0%) and 1313 (94.9%) participants met the data inclusion criteria when considering age- and sex-specific UPSIT percentile, and UPSIT percent correct, respectively (**Appendix**

Figure 1). Forty-three participants were excluded due to inconsistent a-syn SAA status results, and four were excluded because of inconclusive results. An additional 23 individuals were excluded for missing > 50% of their UPSIT items, and one person did not have SCOPA-AUT5 data. Fifty-three participants who were missing more than one UPSIT item were excluded from the age- and sex-specific UPSIT percentile analysis, while 143 participants required imputation of one UPSIT item for this analysis. The baseline characteristics of participants included in the age- and sex-specific UPSIT percentile modeling analysis are summarized in **Table 1** (see **Appendix Table 1** for baseline characteristics from UPSIT percent correct analysis).

For the candidate logistic regression models that used age- and sex-specific UPSIT percentiles, incorporation of sex, constipation symptoms, and *LRRK2* and *GBA* genetic information significantly improved model fit, as evidenced by the likelihood ratio tests (**Appendix Table 2**). For the logistic regression models that used UPSIT percent correct, the inclusion of age but not sex significantly improved model fit, while inclusion of constipation symptoms and genetics data remained beneficial (**Appendix Table 3**).

The parameter estimates from partially and fully adjusted models using age- and sex-specific UPSIT percentiles are reported in **Table 2**. Overall, the fully adjusted model indicates that higher odds of a-syn SAA(+) status are associated with a lower UPSIT age- and sex-specific percentile, male sex, the presence of constipation problems, and carrying a *GBA* mutation. In contrast, having a *LRRK2* variant, particularly R1441G/C or N1437H, is associated with lower odds of a-syn SAA(+) status. For the fully adjusted model using UPSIT percent correct, age, constipation history, and *LRRK2* and *GBA* genetic information, the general patterns were similar (**Appendix Table 4**); however, the coefficient estimate (β) for the UPSIT percent correct was larger in magnitude (β = -0.12), and increasing age was associated with lower odds of a-syn SAA(+) status (β = -0.047).

The odds ratios (ORs) of the fully adjusted logistic regression model using age- and sex-specific UPSIT percentiles are presented in **Figure 1**. The a-syn SAA(+) OR for UPSIT age- and sex-specific percentile is 0.932, indicating that between two individuals whose only difference is a 1-percentile difference in UPSIT, the individual in the higher percentile has 0.932 times the odds of being a-syn SAA(+). Rescaling this OR for 10-percentile unit changes shows that for someone whose age- and sex-specific UPSIT is higher by 10 percentile, they will have 0.497

times the odds of being a-syn SAA(+). Carrying the R1441G/C *LRRK2* variant was another significant impactful variable, with carriers having 0.134 times the odds of being a-syn SAA(+) compared to non-carriers. In contrast, *GBA* mutation carriers had 2.8 times the odds of being a-syn SAA(+) compared to non-*GBA* mutation carriers. Male sex compared to female sex, and more frequent constipation problems versus no or less constipation problems were associated with 1.6 and 2.6 higher odds, respectively, for being a-syn SAA(+). A similar OR plot is provided in the Appendix for the UPSIT percent correct fully adjusted model (**Appendix Figure 2**).

Alternatively, log-odds can be viewed as probabilities for additional interpretations. In **Figure 2**, we provide probability plots for a-syn SAA(+) status versus UPSIT age- and sex-specific percentiles by constipation status and genetic status among males (**Figure 2A**) and females (**Figure 2B**). To facilitate further exploration of the models, an interactive calculator is available online at https://chet-analytics.shinyapps.io/saa_r_shiny/, allowing users to calculate the probabilities of a-syn SAA(+) by entering values of the predictor variables.

Model-Based Predictions: Internal Performance

Using UPSIT age- and sex-specific percentiles to predict a-syn SAA status within the PPMI cohort, both the partially- and fully-adjusted models demonstrated high discrimination abilities in internal testing, as shown by their ROC curves (**Figure 3A**). Similar performance was observed when using UPSIT percent correct (**Appendix Figure 3**).

Applying cut-offs optimized for sensitivity and specificity, the probabilities derived from the models were assessed for accuracy, sensitivity, specificity, positive predictive value, and negative predictive value in determining a-syn SAA(+) versus SAA(-) status (**Table 3**). Results from the internal and optimized bootstrap PPMI data demonstrate that all of the models performed well in predicting a-syn SAA status, with the fully adjusted model achieving the highest accuracy. Results were similar when using UPSIT percent correct instead of UPSIT age- and sex-specific percentiles (**Appendix Table 5**).

Model-Based Predictions: External Performance

For external validation, the models constructed from the PPMI data were used to predict a-syn SAA status in the S4 cohort. The S4 dataset included 75 participants with a-syn SAA data; however, 71 participants were ultimately included because 4 were also co-enrolled in the PPMI study and were therefore excluded to prevent data leakage between the model training/testing and external validation datasets. Differences between the PPMI and S4 cohorts included a longer disease duration for PD participants in the S4 dataset, a higher percentage of healthy controls in S4, and no identified *LRRK2* or *GBA* pathogenic carriers in S4 (**Table 1**).

The partially- and fully-adjusted models trained on PPMI data maintained strong discrimination abilities in predicting a-syn SAA status in the S4 cohort when using UPSIT age- and sex-specific percentiles (**Figure 3B**) or UPSIT percent correct (**Appendix Figure 3**). Applying the same Youden Index cut-offs from the PPMI models to predict a-syn SAA(+) versus SAA(-) status in S4, the partially- and fully-adjusted models demonstrated strong performances in distinguishing between the groups (**Table 3; Appendix Table 5**).

Model-Based Predictions: PD-Only

Using the partially- and fully-adjusted models, performance metrics were recalculated for the PPMI and S4 datasets, focusing solely on participants with manifest PD at enrollment, in contrast to a group comprised of both with, without, and at risk for PD. In the PPMI dataset, this comprised 966 participants, of whom 87.5% were a-syn SAA(+). In the S4 dataset, the subset included 50 participants, with 96.0% being a-syn SAA(+). **Table 4** shows the performance metrics of the three different models in distinguishing a-syn SAA(+) from SAA(-) status using age- and sex-specific UPSIT percentile-based models. The consistently high sensitivities, PPVs, and accuracies across datasets for all models indicate they remained effective in detecting positive cases. However, there were decreases in specificity and NPV indicating a diminished ability to rule out negative cases among manifest PD participants. Results were similar when using UPSIT percent correct (**Appendix Table 6**).

DISCUSSION

This study explores the potential of data-driven models to predict CSF a-syn SAA status in a group of individuals either with or at-risk for PD using readily available clinical data. We found that models using UPSIT data, sex, constipation history, and simple genetics information can predict CSF a-syn SAA status with good accuracy. The predictive capability of the models, demonstrated through AUROC values exceeding 0.900 in both internal and external validation cohorts, highlights the robustness of the models in PD-enriched populations. Interestingly, models that only included UPSIT data showed only marginally reduced performance compared to the partially- and fully-adjusted models. These results underscore the feasibility of leveraging non-invasive measures to help infer the likelihood of the presence of seeding-capable a-syn as detected by CSF a-syn SAA.

Our models suggests that olfactory dysfunction is a useful and potent indicator of CSF a-syn SAA status, reinforcing its significance as an early marker in neurodegenerative processes, particularly in PD. Previous research demonstrate that olfactory dysfunction predicts conversion to PD and dementia with Lewy bodies among patients with isolated REM sleep behavior disorder (iRBD).[26-28] For example, Coughlin and colleagues reported that age- and sex-specific UPSIT percentile had the highest predictive capacity among a range of clinical and demographic features, with an AUROC of 0.87 in predicting CSF a-syn positivity among participants with dementia with Lewy bodies.[29] Furthermore, significant associations have been observed between a-syn pathology and worse olfactory function in PD, iRBD, and normal aging populations.[30-33]

The olfactory tract is considered an initiation site for the seeding and spread of pathological a-syn in PD. Recent disease progression modeling of 814 brain donors indicates that most with Lewy body pathology show earliest pathology in the olfactory bulb, while a smaller subset shows later olfactory involvement after initial brainstem pathology and spread.[34] Notably, those inferred to have a spatiotemporal trajectory involving later olfactory bulb pathology but earlier brainstem pathology and more widespread regional brain involvement had worse pre-mortem smell dysfunction compared to those inferred to have early olfactory bulb pathology, suggesting this symptom is linked to overall brain pathology rather than localized changes. Clinicopathological studies in non-PD populations support the notion that olfactory dysfunction could be more associated with widespread Lewy body pathology rather than localized olfactory

changes.[32,33] However, there are experimental studies in mice that challenge this view, demonstrating that olfactory dysfunction can occur due to localized a-syn aggregation in the olfactory bulb by disrupting neural activity and synaptic transmission.[35] Additionally, RNA sequencing has revealed significant gene alterations in the olfactory bulb of later-stage PD patients, implicating pathways related to neuroinflammation, neurotransmitter dysfunction, and disruptions in olfactory transduction signaling.[36] While the exact mechanisms of olfactory dysfunction in PD will require further investigation, we clearly show that olfactory dysfunction is a strong indicator of CSF a-syn SAA positivity in PD.

Our modeling also found that *LRRK2* pathogenic carriers had reduced odds of being a-syn CSF SAA positive compared to PD non-carriers, while *GBA* pathogenic carriers had increased odds of positivity. In the PPMI training dataset, 63% of *LRRK2* PD pathogenic carriers had positive a-syn CSF SAA results, while 93% of the sporadic PD population were positive. This aligns with prior research reporting 40% to 78% of *LRRK2* carriers having positive a-syn seeding activity compared to approximately 90% of sporadic PD participants.[15,37,38]

Our models suggest that the type of *LRRK2* pathogenic variant is also important to consider when predicting a-syn SAA status. Carriers of R1441G/C/N1437H had the greatest reduction in odds of being a-syn CSF SAA positive compared to non-carriers, while G2019S/I2020T carriers also had reduced odds but to a lesser extent. Several neuropathological studies of R1441G/C/H PD carriers have shown that Lewy body pathology is absent or rarely present at autopsy.[39-41] Histopathology from individuals carrying the G2019S mutation seem to show a-syn pathology is more common but is still highly variable, which aligns with our modeling.[42,43] For instance, among 37 *LRRK2*-related PD cases, all reporting neuronal loss within the substantia nigra, Lewy body pathology was present in 11/17 (65%) G2019S mutation carriers, and 2/6 (33%) R1441C/G carriers.[44] It is unclear as to the specific biochemical mechanism(s) that could be driving these differences in a-syn seeding activities between different *LRRK2* pathogenic variants, and it will be important to investigate these differences further for potential therapies that target a-syn aggregation and *LRRK2* carriers.

A history of constipation symptoms self-reported as “regular” or “often” was also a significant predictor of CSF a-syn SAA positive status in the fully adjusted model. Individuals with more frequent constipation had more than 2.5 times higher odds of being a-syn SAA(+)

than those with no or infrequent constipation issues. Constipation, a common non-motor symptom of PD, is also a predictor of developing the disease.[45] Neuropathological studies have shown a-syn accumulation in the enteric nervous system in both PD and Lewy body disease, and some experimental studies suggest that constipation may be related to the formation of a-syn in the GI tract.[23,46] However, few clinical studies have directly measured the association of a-syn pathology with constipation. While immunohistochemistry studies have shown minimal evidence linking localized Lewy pathology in colonic and mucosal biopsies with constipation in PD and non-PD subjects, a-syn seeding assays have demonstrated an association.[22] Kuzkina and colleagues found higher a-syn seeding activity from skin biopsy samples in PD patients with constipation compared to those without.[8] Thus, incorporating a simple self-reported history of constipation symptoms questionnaire seems to be a plausible and helpful predictor in identifying some individuals at-risk of having positive a-syn SAA results.

Male sex was associated with higher odds of a-syn SAA(+) results when using age- and sex-specific UPSIT percentiles, but it was not a significant predictor in models that used UPSIT percent correct scores. Conversely, increasing age was linked to lower odds of being a-syn SAA(+) when using UPSIT percent correct scores, but this association was not observed with the age- and sex-adjusted percentiles. These findings suggest that the influence of sex and age on a-syn status may be driven more by the method in which UPSIT data are incorporated into the models rather than by true biological differences. Nevertheless, it would be interesting to investigate further how these variables interact and whether different approaches to incorporating UPSIT data could reveal more nuanced insights into the relationship between sex, age, and a-syn pathology.

These models could be useful in several ways. First, they could facilitate identification of individuals who would most benefit from CSF a-syn SAA testing in the research setting, optimizing resource allocation, and minimizing unnecessary invasive procedures. For example, if an individual has a very low probability of being a-syn SAA(+) according to the models, they may not need further invasive testing. Additionally, these models can be applied to historical datasets that lack CSF samples but have UPSIT data, allowing researchers to understand the influence of a-syn SAA status on various exploratory analyses. Furthermore, the models can be utilized in research settings where CSF sample collection is not feasible. This flexibility enables

researchers to infer SAA status likelihood across a broader spectrum of PD studies, potentially enriching our understanding of disease patterns and treatment responses in more diverse patient populations. By enhancing the ability to estimate a-syn pathology without direct sample analysis, these models promise to expand the scope and applicability of this area of PD research.

Our study has some limitations. The external validation cohort was relatively small compared to the training dataset, and did not include participants with *LRRK2* or *GBA* PD-associated genetic mutations. Therefore, we were unable to truly test how well our fully-adjusted model could perform in the setting of a more genetically diverse PD population whose a-syn pathology is more variable. It would be of great interest to test and recalibrate these models using additional datasets that have collected CSF a-syn SAA data with UPSIT information in populations with PD and other neurodegenerative diseases. We also did not include rapid eye movement behavior disorder (RBD) as a potential predictor in our models, which has been shown to be a risk-factor for the presence a-syn pathology.[29] While the PPMI study collected RBD data through the RBD Questionnaire, the S4 validation dataset did not. Therefore, we elected not to pursue the use of RBD symptoms in our models. There were also limitations of the models when using them specifically in a PD manifest population, particularly around the specificity and NPV of the models, indicating a higher rate of false positives. This suggests that while the models are effective at identifying true positives, they may still require further refinement to reduce false positive rates and improve their overall diagnostic utility when only considering manifest PD.

REFERENCES

1. Bellomo G, De Luca CMG, Paoletti FP, Gaetani L, Moda F, Parnetti L. α-Synuclein Seed Amplification Assays for Diagnosing Synucleinopathies: The Way Forward. Neurology. 2022 Aug 2;99(5):195-205. doi: 10.1212/WNL.0000000000200878. Epub 2022 Jun 3. PMID: 35914941.
2. Zheng Y, Li S, Yang C, Yu Z, Jiang Y, Feng T. Comparison of biospecimens for α-synuclein seed amplification assays in Parkinson's disease: A systematic review and network meta-analysis. Eur J Neurol. 2023 Dec;30(12):3949-3967. doi: 10.1111/ene.16041. Epub 2023 Aug 28. PMID: 37573472.
3. Chahine LM, Merchant K, Siderowf A, Sherer T, Tanner C, Marek K, Simuni T. Proposal for a Biologic Staging System of Parkinson's Disease. J Parkinsons Dis. 2023;13(3):297-309. doi: 10.3233/JPD-225111. PMID: 37066922; PMCID: PMC10200239.
4. Simuni T, Chahine LM, Poston K, Brumm M, Buracchio T, Campbell M, Chowdhury S, Coffey C, Concha-Marambio L, Dam T, DiBiaso P, Foroud T, Frasier M, Gochanour C, Jennings D, Kieburtz K, Kopil CM, Merchant K, Mollenhauer B, Montine T, Nudelman K, Pagano G, Seibyl J, Sherer T, Singleton A, Stephenson D, Stern M, Soto C, Tanner CM, Tolosa E, Weintraub D, Xiao Y, Siderowf A, Dunn B, Marek K. A biological definition of neuronal α-synuclein disease: towards an integrated staging system for research. Lancet Neurol. 2024 Feb;23(2):178-190. doi: 10.1016/S1474-4422(23)00405-2. PMID: 38267190.
5. Höglinger GU, Adler CH, Berg D, Klein C, Outeiro TF, Poewe W, Postuma R, Stoessl AJ, Lang AE. A biological classification of Parkinson's disease: the SynNeurGe research diagnostic criteria. Lancet Neurol. 2024 Feb;23(2):191-204. doi: 10.1016/S1474-4422(23)00404-0. Erratum in: Lancet Neurol. 2024 Mar;23(3):e7. PMID: 38267191.
6. Wang Z, Becker K, Donadio V, Siedlak S, Yuan J, Rezaee M, Incensi A, Kuzkina A, Orrú CD, Tatsuoka C, Liguori R, Gunzler SA, Caughey B, Jimenez-Capdeville ME, Zhu X, Doppler K, Cui L, Chen SG, Ma J, Zou WQ. Skin α-Synuclein Aggregation Seeding Activity as a Novel Biomarker for Parkinson Disease. JAMA Neurol. 2020 Sep 28;78(1):1–11. doi: 10.1001/jamaneurol.2020.3311. Epub ahead of print. Erratum in: JAMA Neurol. 2021 Jan 1;78(1):120. doi: 10.1001/jamaneurol.2020.4087. PMID: 32986090; PMCID: PMC7522783.
7. Mammana A, Baiardi S, Quadalti C, Rossi M, Donadio V, Capellari S, Liguori R, Parchi P. RT-QuIC Detection of Pathological α-Synuclein in Skin Punches of Patients with Lewy Body Disease. Mov Disord. 2021 Sep;36(9):2173-2177. doi: 10.1002/mds.28651. Epub 2021 May 18. PMID: 34002890; PMCID: PMC8518528.
8. Kuzkina A, Bargar C, Schmitt D, Rößle J, Wang W, Schubert AL, Tatsuoka C, Gunzler SA, Zou WQ, Volkmann J, Sommer C, Doppler K, Chen SG. Diagnostic value of skin RT-QuIC in Parkinson's disease: a two-laboratory study. NPJ Parkinsons Dis. 2021 Nov 15;7(1):99. doi: 10.1038/s41531-021-00242-2. PMID: 34782640; PMCID: PMC8593128.

9. Donadio V, Wang Z, Incensi A, Rizzo G, Fileccia E, Vacchiano V, Capellari S, Magnani M, Scaglione C, Stanzani Maserati M, Avoni P, Liguori R, Zou W. In Vivo Diagnosis of Synucleinopathies: A Comparative Study of Skin Biopsy and RT-QuIC. Neurology. 2021 May 18;96(20):e2513-e2524. doi: 10.1212/WNL.0000000000011935. Epub 2021 Apr 9. Erratum in: Neurology. 2022 Jul 5;99(1):40-42. doi: 10.1212/WNL.0000000000200689. PMID: 33837116; PMCID: PMC8205473.
10. Kuzkina A, Rößle J, Seger A, Panzer C, Kohl A, Maltese V, Musacchio T, Blaschke SJ, Tamgüney G, Kaulitz S, Rak K, Scherzad A, Zimmermann PH, Klussmann JP, Hackenberg S, Volkmann J, Sommer C, Sommerauer M, Doppler K. Combining skin and olfactory α-synuclein seed amplification assays (SAA)-towards biomarker-driven phenotyping in synucleinopathies. NPJ Parkinsons Dis. 2023 May 29;9(1):79. doi: 10.1038/s41531-023-00519-8. PMID: 37248217; PMCID: PMC10226020.
11. Yoo D, Bang JI, Ahn C, Nyaga VN, Kim YE, Kang MJ, Ahn TB. Diagnostic value of α-synuclein seeding amplification assays in α-synucleinopathies: A systematic review and meta-analysis. Parkinsonism Relat Disord. 2022 Nov;104:99-109. doi: 10.1016/j.parkreldis.2022.10.007. Epub 2022 Oct 19. PMID: 36289019.
12. Okuzumi A, Hatano T, Matsumoto G, Nojiri S, Ueno SI, Imamichi-Tatano Y, Kimura H, Kakuta S, Kondo A, Fukuhara T, Li Y, Funayama M, Saiki S, Taniguchi D, Tsunemi T, McIntyre D, Gérardy JJ, Mittelbronn M, Kruger R, Uchiyama Y, Nukina N, Hattori N. Propagative α-synuclein seeds as serum biomarkers for synucleinopathies. Nat Med. 2023 Jun;29(6):1448-1455. doi: 10.1038/s41591-023-02358-9. Epub 2023 May 29. PMID: 37248302; PMCID: PMC10287557.
13. Zheng Y, Yu Z, Cai H, Kou W, Yang C, Li S, Zhang N, Feng T. Detection of α-Synuclein in Oral Mucosa by Seed Amplification Assay in Synucleinopathies and Isolated REM Sleep Behavior Disorder. Mov Disord. 2024 May 7. doi: 10.1002/mds.29828. Epub ahead of print. PMID: 38715177.
14. Gibbons CH, Levine T, Adler C, Bellaire B, Wang N, Stohl J, Agarwal P, Aldridge GM, Barboi A, Evidente VGH, Galasko D, Geschwind MD, Gonzalez-Duarte A, Gil R, Gudesblatt M, Isaacson SH, Kaufmann H, Khemani P, Kumar R, Lamotte G, Liu AJ, McFarland NR, Miglis M, Reynolds A, Sahagian GA, Saint-Hillaire MH, Schwartzbard JB, Singer W, Soileau MJ, Vernino S, Yerstein O, Freeman R. Skin Biopsy Detection of Phosphorylated α-Synuclein in Patients With Synucleinopathies. JAMA. 2024 Apr 16;331(15):1298-1306. doi: 10.1001/jama.2024.0792. PMID: 38506839; PMCID: PMC10955354.
15. Siderowf A, Concha-Marambio L, Lafontant DE, Farris CM, Ma Y, Urenia PA, Nguyen H, Alcalay RN, Chahine LM, Foroud T, Galasko D, Kieburtz K, Merchant K, Mollenhauer B, Poston KL, Seibyl J, Simuni T, Tanner CM, Weintraub D, Videnovic A, Choi SH, Kurth R, Caspell-Garcia C, Coffey CS, Frasier M, Oliveira LMA, Hutten SJ, Sherer T, Marek K, Soto C; Parkinson's Progression Markers Initiative. Assessment of heterogeneity among participants in the Parkinson's Progression Markers Initiative cohort using α-synuclein seed amplification: a cross-sectional study. Lancet Neurol. 2023 May;22(5):407-417. doi: 10.1016/S1474-4422(23)00109-6. PMID: 37059509; PMCID: PMC10627170.

16. Parkinson Progression Marker Initiative. The Parkinson Progression Marker Initiative (PPMI). Prog Neurobiol. 2011 Dec;95(4):629-35. doi: 10.1016/j.pneurobio.2011.09.005. Epub 2011 Sep 14. PMID: 21930184; PMCID: PMC9014725.
17. Visanji NP, Mollenhauer B, Beach TG, Adler CH, Coffey CS, Kopil CM, Dave KD, Foroud T, Chahine L, Jennings D; Systemic Synuclein Sampling Study (S4). The Systemic Synuclein Sampling Study: toward a biomarker for Parkinson's disease. Biomark Med. 2017 Apr;11(4):359-368. doi: 10.2217/bmm-2016-0366. Epub 2017 Mar 29. PMID: 28353371.
18. Chahine LM, Beach TG, Seedorff N, Caspell-Garcia C, Coffey CS, Brumm M, Adler CH, Serrano GE, Linder C, Mosovsky S, Foroud T, Riss H, Ecklund D, Seibyl J, Jennings D, Arnedo V, Riley L, Dave KD, Mollenhauer B; Systemic sSynuclein Sampling study. Feasibility and Safety of Multicenter Tissue and Biofluid Sampling for α-Synuclein in Parkinson's Disease: The Systemic Synuclein Sampling Study (S4). J Parkinsons Dis. 2018;8(4):517-527. doi: 10.3233/JPD-181434. PMID: 30248065; PMCID: PMC6226302.
19. Concha-Marambio L, Pritzkow S, Shahnawaz M, Farris CM, Soto C. Seed amplification assay for the detection of pathologic alpha-synuclein aggregates in cerebrospinal fluid. Nat Protoc. 2023 Apr;18(4):1179-1196. doi: 10.1038/s41596-022-00787-3. Epub 2023 Jan 18. PMID: 36653527; PMCID: PMC10561622.
20. Kang UJ, Boehme AK, Fairfoul G, Shahnawaz M, Ma TC, Hutten SJ, Green A, Soto C. Comparative study of cerebrospinal fluid α-synuclein seeding aggregation assays for diagnosis of Parkinson's disease. Mov Disord. 2019 Apr;34(4):536-544. doi: 10.1002/mds.27646. Epub 2019 Mar 6. PMID: 30840785; PMCID: PMC6519150.
21. Chahine LM, Beach TG, Adler CH, Hepker M, Kanthasamy A, Appel S, Pritzkow S, Pinho M, Mosovsky S, Serrano GE, Coffey C, Brumm MC, Oliveira LMA, Eberling J, Mollenhauer B; Systemic Synuclein Sampling Study. Central and peripheral α-synuclein in Parkinson disease detected by seed amplification assay. Ann Clin Transl Neurol. 2023 May;10(5):696-705. doi: 10.1002/acn3.51753. Epub 2023 Mar 27. PMID: 36972727; PMCID: PMC10187727.
22. Lebouvier T, Chaumette T, Damier P, Coron E, Touchefeu Y, Vrignaud S, Naveilhan P, Galmiche JP, Bruley des Varannes S, Derkinderen P, Neunlist M. Pathological lesions in colonic biopsies during Parkinson's disease. Gut. 2008 Dec;57(12):1741-3. doi: 10.1136/gut.2008.162503. PMID: 19022934.
23. Wang L, Magen I, Yuan PQ, Subramaniam SR, Richter F, Chesselet MF, Taché Y. Mice overexpressing wild-type human alpha-synuclein display alterations in colonic myenteric ganglia and defecation. Neurogastroenterol Motil. 2012 Sep;24(9):e425-36. doi: 10.1111/j.1365-2982.2012.01974.x. Epub 2012 Jul 11. PMID: 22779732; PMCID: PMC3712640.
24. Brumm MC, Pierz KA, Lafontant DE, Caspell-Garcia C, Coffey CS, Siderowf A, Marek K. Updated Percentiles for the University of Pennsylvania Smell Identification Test in Adults 50 Years of Age and Older. Neurology. 2023 Apr 18;100(16):e1691-e1701. doi: 10.1212/WNL.0000000000207077. Epub 2023 Feb 27. PMID: 36849448; PMCID: PMC10115503.

25. S Xu, M Chen, T Feng, L Zhan, L Zhou, G Yu. Use ggbreak to effectively utilize plotting space to deal with large datasets and outliers. Frontiers in Genetics. 2021, 12:774846. doi: 10.3389/fgene.2021.774846
26. Mahlknecht P, Iranzo A, Högl B, Frauscher B, Müller C, Santamaría J, Tolosa E, Serradell M, Mitterling T, Gschliesser V, Goebel G, Brugger F, Scherfler C, Poewe W, Seppi K; Sleep Innsbruck Barcelona Group. Olfactory dysfunction predicts early transition to a Lewy body disease in idiopathic RBD. Neurology. 2015 Feb 17;84(7):654-8. doi: 10.1212/WNL.0000000000001265. Epub 2015 Jan 21. PMID: 25609758.
27. Iranzo A, Marrero-González P, Serradell M, Gaig C, Santamaria J, Vilaseca I. Significance of hyposmia in isolated REM sleep behavior disorder. J Neurol. 2021 Mar;268(3):963-966. doi: 10.1007/s00415-020-10229-3. Epub 2020 Sep 23. PMID: 32968939.
28. Miyamoto T, Miyamoto M. Odor identification predicts the transition of patients with isolated RBD: A retrospective study. Ann Clin Transl Neurol. 2022 Aug;9(8):1177-1185. doi: 10.1002/acn3.51615. Epub 2022 Jun 29. PMID: 35767550; PMCID: PMC9380141.
29. Coughlin DG, MacLeod KR, Middleton JS, Bozoki AC, Galvin JE, Irwin DJ, Lippa CF, Litvan I, Lopez OL, Berman S, Tsuang DW, Zabetian CP, Honig LS, Marder KS, Fleisher JE, Sabbagh M, Wint D, Taylor AS, Bekris L, Leverenz JB, Galasko D. Association of CSF α-Synuclein Seeding Amplification Assay Results With Clinical Features of Possible and Probable Dementia With Lewy Bodies. Neurology. 2024 Aug 13;103(3):e209656. doi: 10.1212/WNL.0000000000209656. Epub 2024 Jul 16. PMID: 39013126; PMCID: PMC11238940.
30. Stefani A, Iranzo A, Holzknecht E, Perra D, Bongianni M, Gaig C, Heim B, Serradell M, Sacchetto L, Garrido A, Capaldi S, Sánchez-Gómez A, Cecchini MP, Mariotto S, Ferrari S, Fiorini M, Schmutzhard J, Cocchiara P, Vilaseca I, Brozzetti L, Monaco S, Jose Marti M, Seppi K, Tolosa E, Santamaria J, Högl B, Poewe W, Zanusso G; SINBAR (Sleep Innsbruck Barcelona) group. Alpha-synuclein seeds in olfactory mucosa of patients with isolated REM sleep behaviour disorder. Brain. 2021 May 7;144(4):1118-1126. doi: 10.1093/brain/awab005. PMID: 33855335.
31. Kuzkina A, Panzer C, Seger A, Schmitt D, Rößle J, Schreglmann SR, Knacke H, Salabasidou E, Kohl A, Sittig E, Barbe M, Berg D, Volkmann J, Sommer C, Oertel WH, Schaeffer E, Sommerauer M, Janzen A, Doppler K. Dermal Real-Time Quaking-Induced Conversion Is a Sensitive Marker to Confirm Isolated Rapid Eye Movement Sleep Behavior Disorder as an Early α-Synucleinopathy. Mov Disord. 2023 Jun;38(6):1077-1082. doi: 10.1002/mds.29340. Epub 2023 Feb 7. PMID: 36750755.
32. Tremblay C, Serrano GE, Intorcia AJ, Sue LI, Wilson JR, Adler CH, Shill HA, Driver-Dunckley E, Mehta SH, Beach TG. Effect of olfactory bulb pathology on olfactory function in normal aging. Brain Pathol. 2022 Sep;32(5):e13075. doi: 10.1111/bpa.13075. Epub 2022 Apr 29. PMID: 35485279; PMCID: PMC9424999.
33. Nag S, Yu L, VanderHorst VG, Schneider JA, Bennett DA, Buchman AS, Wilson RS. Neocortical Lewy bodies are associated with impaired odor identification in community-dwelling elders without clinical PD. J Neurol. 2019 Dec;266(12):3108-3118. doi:

10.1007/s00415-019-09540-5. Epub 2019 Sep 18. PMID: 31535271; PMCID: PMC6851442.

34. Mastenbroek, S.E., Vogel, J.W., Collij, L.E. *et al*. Disease progression modelling reveals heterogeneity in trajectories of Lewy-type α-synuclein pathology. *Nat Commun* **15**, 5133 (2024). https://doi.org/10.1038/s41467-024-49402-x
35. Chen F, Liu W, Liu P, Wang Z, Zhou Y, Liu X, Li A. α-Synuclein aggregation in the olfactory bulb induces olfactory deficits by perturbing granule cells and granular-mitral synaptic transmission. NPJ Parkinsons Dis. 2021 Dec 13;7(1):114. doi: 10.1038/s41531-021-00259-7. PMID: 34903719; PMCID: PMC8668919.
36. Tremblay C, Aslam S, Walker JE, Lorenzini I, Intorcia AJ, Arce RA, Choudhury P, Adler CH, Shill HA, Driver-Dunckley E, Mehta S, Piras IS, Belden CM, Atri A, Beach TG, Serrano GE. RNA sequencing of olfactory bulb in Parkinson's disease reveals gene alterations associated with olfactory dysfunction. Neurobiol Dis. 2024 Jun 15;196:106514. doi: 10.1016/j.nbd.2024.106514. Epub 2024 Apr 24. PMID: 38663633; PMCID: PMC11132317.
37. Garrido A, Fairfoul G, Tolosa ES, Martí MJ, Green A; Barcelona LRRK2 Study Group. *α*-synuclein RT-QuIC in cerebrospinal fluid of LRRK2-linked Parkinson's disease. Ann Clin Transl Neurol. 2019 May 9;6(6):1024-1032. doi: 10.1002/acn3.772. PMID: 31211166; PMCID: PMC6562027.
38. Brockmann K, Quadalti C, Lerche S, Rossi M, Wurster I, Baiardi S, Roeben B, Mammana A, Zimmermann M, Hauser AK, Deuschle C, Schulte C, Waniek K, Lachmann I, Sjödin S, Brinkmalm A, Blennow K, Zetterberg H, Gasser T, Parchi P. Association between CSF alpha-synuclein seeding activity and genetic status in Parkinson's disease and dementia with Lewy bodies. Acta Neuropathol Commun. 2021 Oct 30;9(1):175. doi: 10.1186/s40478-021-01276-6. PMID: 34717775; PMCID: PMC8556894.
39. Vinagre-Aragón A, Campo-Caballero D, Mondragón-Rezola E, Pardina-Vilella L, Hernandez Eguiazu H, Gorostidi A, Croitoru I, Bergareche A, Ruiz-Martinez J. A More Homogeneous Phenotype in Parkinson's Disease Related to R1441G Mutation in the LRRK2 Gene. Front Neurol. 2021 Mar 2;12:635396. doi: 10.3389/fneur.2021.635396. PMID: 33763016; PMCID: PMC7982912.
40. Takanashi M, Funayama M, Matsuura E, Yoshino H, Li Y, Tsuyama S, Takashima H, Nishioka K, Hattori N. Isolated nigral degeneration without pathological protein aggregation in autopsied brains with LRRK2 p.R1441H homozygous and heterozygous mutations. Acta Neuropathol Commun. 2018 Oct 17;6(1):105. doi: 10.1186/s40478-018-0617-y. PMID: 30333048; PMCID: PMC6192197.
41. Zimprich A, Biskup S, Leitner P, Lichtner P, Farrer M, Lincoln S, Kachergus J, Hulihan M, Uitti RJ, Calne DB, Stoessl AJ, Pfeiffer RF, Patenge N, Carbajal IC, Vieregge P, Asmus F, Müller-Myhsok B, Dickson DW, Meitinger T, Strom TM, Wszolek ZK, Gasser T. Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology. Neuron. 2004 Nov 18;44(4):601-7. doi: 10.1016/j.neuron.2004.11.005. PMID: 15541309.

42. Schneider SA, Alcalay RN. Neuropathology of genetic synucleinopathies with parkinsonism: Review of the literature. Mov Disord. 2017 Nov;32(11):1504-1523. doi: 10.1002/mds.27193. PMID: 29124790; PMCID: PMC5726430.
43. Henderson MX, Sengupta M, Trojanowski JQ, Lee VMY. Alzheimer's disease tau is a prominent pathology in LRRK2 Parkinson's disease. Acta Neuropathol Commun. 2019 Nov 16;7(1):183. doi: 10.1186/s40478-019-0836-x. PMID: 31733655; PMCID: PMC6858668.
44. Kalia LV, Lang AE, Hazrati LN, Fujioka S, Wszolek ZK, Dickson DW, Ross OA, Van Deerlin VM, Trojanowski JQ, Hurtig HI, Alcalay RN, Marder KS, Clark LN, Gaig C, Tolosa E, Ruiz-Martínez J, Marti-Masso JF, Ferrer I, López de Munain A, Goldman SM, Schüle B, Langston JW, Aasly JO, Giordana MT, Bonifati V, Puschmann A, Canesi M, Pezzoli G, Maues De Paula A, Hasegawa K, Duyckaerts C, Brice A, Stoessl AJ, Marras C. Clinical correlations with Lewy body pathology in LRRK2-related Parkinson disease. JAMA Neurol. 2015 Jan;72(1):100-5. doi: 10.1001/jamaneurol.2014.2704. PMID: 25401511; PMCID: PMC4399368.
45. Yao L, Liang W, Chen J, Wang Q, Huang X. Constipation in Parkinson's Disease: A Systematic Review and Meta-Analysis. Eur Neurol. 2023;86(1):34-44. doi: 10.1159/000527513. Epub 2022 Dec 5. PMID: 36470230.
46. Cersosimo MG, Benarroch EE. Pathological correlates of gastrointestinal dysfunction in Parkinson's disease. Neurobiol Dis. 2012 Jun;46(3):559-64. doi: 10.1016/j.nbd.2011.10.014. Epub 2011 Oct 25. PMID: 22048068.

FUNDING

PPMI – a public-private partnership – is funded by the Michael J. Fox Foundation for Parkinson's Research and funding partners, including 4D Pharma, Abbvie, AcureX, Allergan, Amathus Therapeutics, Aligning Science Across Parkinson's, AskBio, Avid Radiopharmaceuticals, BIAL, BioArctic, Biogen, Biohaven, BioLegend, BlueRock Therapeutics, Bristol-Myers Squibb, Calico Labs, Capsida Biotherapeutics, Celgene, Cerevel Therapeutics, Coave Therapeutics, DaCapo Brainscience, Denali, Edmond J. Safra Foundation, Eli Lilly, Gain Therapeutics, GE HealthCare, Genentech, GSK, Golub Capital, Handl Therapeutics, Insitro, Jazz Pharmaceuticals, Johnson & Johnson Innovative Medicine, Lundbeck, Merck, Meso Scale Discovery, Mission Therapeutics, Neurocrine Biosciences, Neuron23, Neuropore, Pfizer, Piramal, Prevail Therapeutics, Roche, Sanofi, Servier, Sun Pharma Advanced Research Company, Takeda, Teva, UCB, Vanqua Bio, Verily, Voyager Therapeutics, the Weston Family Foundation and Yumanity Therapeutics.

S4 was funded by the Michael J. Fox Foundation for Parkinson's Research.

Research reported in this publication was supported by the National Institute of Neurological Disorders and Stroke (Award Number P50NS108676). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

DATA AVAILABILITY

The data supporting the findings from this study are available at the Parkinson's Progression Marker Initiative (PPMI) website (https://www.ppmi-info.org/). This analysis used data openly available from the PPMI (Tier 1 data). Protocol information for The Parkinson's Progression Markers Initiative (PPMI) Clinical - Establishing a Deeply Phenotyped PD Cohort AM 3.2. can be found on protocols.io or by following this link: https://dx.doi.org/10.17504/protocols.io.n92ldmw6ol5b/v2. Additional data from the Systemic Synuclein Sampling Study (S4) were available through Zenodo (https://zenodo.org/records/7821869) after registration for approved data access. Statistical

analysis codes used to perform the analyses in this article are shared on Zenodo (https://doi.org/10.5281/zenodo.14867028).

TABLES

Table 1. Demographics and clinical characteristics of PPMI (training) and S4 (validation) datasets.

Variable	PPMI (n=1260)	S4 (n=71)	p-value
CSF a-syn SAA Status, n *(%)*			
SAA(+)	866 *(68.7%)*	48 *(67.6%)*	0.84
SAA(-)	394 *(31.3%)*	23 *(32.4%)*	
Cohort, n *(%)*			
Parkinson's Disease (PD)	959 *(76.1%)*	50 *(70.4%)*	0.28[f]
PD Sporadic	722 *(57.3%)*	50	
PD LRRK2+	145 *(11.5%)*[a]	0	
PD GBA+	74 *(5.9%)*	0	
PD SNCA+	10 *(0.8%)*	0	
PD PRKN+	8 *(0.6%)*	0[e]	
Non-PD LRRK2+	2 *(0.2%)*	0	
Non-PD GBA+	2 *(0.2%)*	0	
Healthy Control	234 *(18.6%)*[b]	21 *(29.6%)*	**0.022**[g]
SWEDD	60 *(4.8%)*	0	
Non-PD, Non-Healthy Control	3 *(0.2%)*	0	
Sex, n *(%)*			
Male	792 *(62.9%)*	43 *(60.6%)*	0.70
Female	468 *(37.1%)*	28 *(39.4%)*	
Age, years			
Mean ± SD	62.4 ± 10.0	63.0 ± 8.0	0.51
Disease duration among PD, years			
Mean ± SD	1.2 ± 1.6	5.1 ± 4.8	**<0.0001**
LRRK2 Carrier Status among LRRK2+, n *(%)*			
Non-Carrier	1113 *(88.3%)*	71 *(100%)*	**0.0023**
G2019S or I2020T	132 *(10.5%)*[c]	0	
R1441G/C or N1437H	15 *(1.2%)*[d]	0	
UPSIT Age-, Sex-Adjusted Percentile			
Mean ± SD	24.6 ± 29.9	22.6 ± 27.7	0.55
MoCA Score			
Mean ± SD	27.1 ± 2.4	27.4 ± 2.3	0.22
Problems with Constipation (SCOPA item 5), n *(%)*			
Never to Sometimes	1142 *(90.6%)*	65 *(91.5%)*	0.80
Regular to Often	118 *(9.4%)*	6 *(8.5%)*	

[a]Includes seven (n=7) participants who are LRRK2 + GBA PD
[b]Includes two (n=2) participants who were found to carry GBA mutation. Includes thirteen (n=13) participants who later deemed to be non-HC due to SAA+ result at baseline or later.
[c]Includes one (n=1) participant with I2020T mutation and one (n=1) participant indicated to have LRRK2 mutation by consensus committee but specific variant not available.
[d]Includes one (n=1) participant with N1437H mutation
[e]Data on PRKN not available in S4
[f]Comparison of Parkinson's disease proportions
[g]Comparison of healthy control proportions

Table 2. Logistic regression coefficient estimates from models using age- and sex-specific UPSIT percentile

Variables	Model					
	UPSIT%ile		UPSIT%ile + Sex + SCOPA5		UPSIT%ile + Sex + SCOPA5 + LRRK2/GBA	
	β (SE)	p-value	β (SE)	p-value	β (SE)	p-value
Intercept	2.62 (0.13)	**< 0.0001**	2.25 (0.16)	**< 0.0001**	2.33 (0.17)	**< 0.0001**
UPSIT%ile	-0.069 (0.0040)	**< 0.0001**	-0.069 (0.0040)	**< 0.0001**	-0.070 (0.0040)	**< 0.0001**
Male Sex	---	---	0.50 (0.18)	**0.0053**	0.45 (0.18)	**0.013**
Constipation problems	---	---	0.85 (0.36)	**0.019**	0.96 (0.37)	**0.010**
LRRK2 G2019S or I2020T carrier	---	---	---	---	-0.43 (0.27)	0.11
LRRK2 R1441G/C or N1437H carrier	---	---	---	---	-2.01 (0.59)	**0.00067**
GBA Carrier	---	---	---	---	1.01 (0.47)	**0.030**

β = regression coefficient; SE = standard error

Table 3. Internal and external model performances to discriminate a-Syn SAA+ vs. SAA- status using age- and sex-specific UPSIT percentile

*Performance Metric and Dataset	Model (Youden Index Cut-Off)		
	UPSIT%ile (Cut-Off: 0.810)	UPSIT%ile + Sex + SCOPA5 (Cut-Off: 0.760)	UPSIT%ile + Sex + SCOPA5 + LRRK2/GBA (Cut-Off: 0.760)
Specificity			
PPMI Full (Internal)	0.860	0.822	0.845
PPMI Bootstrap**	0.860 [0.828, 0.890]	0.824 [0.788, 0.858]	0.843 [0.807, 0.875]
S4 (External)	0.870	0.870	0.870
Sensitivity			
PPMI Full (Internal)	0.848	0.888	0.881
PPMI Bootstrap**	0.841 [0.794, 0.872]	0.879 [0.847, 0.905]	0.877 [0.845, 0.902]
S4 (External)	0.938	0.958	0.958
Positive Predictive Value			
PPMI Full (Internal)	0.930	0.917	0.926
PPMI Bootstrap**	0.930 [0.915, 0.944]	0.917 [0.902, 0.931]	0.925 [0.910, 0.938]
S4 (External)	0.938	0.939	0.939
Negative Predictive Value			
PPMI Full (Internal)	0.720	0.770	0.764
PPMI Bootstrap**	0.712 [0.669, 0.751]	0.756 [0.716, 0.794]	0.757 [0.717, 0.794]
S4 (External)	0.870	0.909	0.909
Accuracy			
PPMI Full (Internal)	0.852	0.867	0.870
PPMI Bootstrap**	0.847 [0.818, 0.871]	0.862 [0.838, 0.883]	0.866 [0.843, 0.887]
S4 (External)	0.915	0.930	0.930

*Performance metrics were calculated based on the specific Youden Index Cut-Off for each model;
**Mean of bootstrap and 95% bootstrap percentile intervals

Table 4. Model performance metrics among manifest PD participants to discriminate a-Syn SAA+ vs. SAA- status using age- and sex-specific UPSIT percentile

*Performance Metric and Dataset	Model (Youden Index Cut-Off)		
	UPSIT%ile (Cut-Off: 0.810)	UPSIT%ile + Sex + SCOPA5 (Cut-Off: 0.760)	UPSIT%ile + Sex + SCOPA5 + LRRK2/GBA (Cut-Off: 0.760)
Specificity			
PPMI (Internal)	0.785	0.711	0.785
S4 (External)	0.500	0.500	0.500
Sensitivity			
PPMI (Internal)	0.853	0.893	0.885
S4 (External)	0.935	0.957	0.957
Positive Predictive Value			
PPMI (Internal)	0.965	0.956	0.966
S4 (External)	0.956	0.957	0.957
Negative Predictive Value			
PPMI (Internal)	0.434	0.489	0.495
S4 (External)	0.400	0.500	0.500
Accuracy			
PPMI (Internal)	0.845	0.871	0.873
S4 (External)	0.900	0.920	0.920

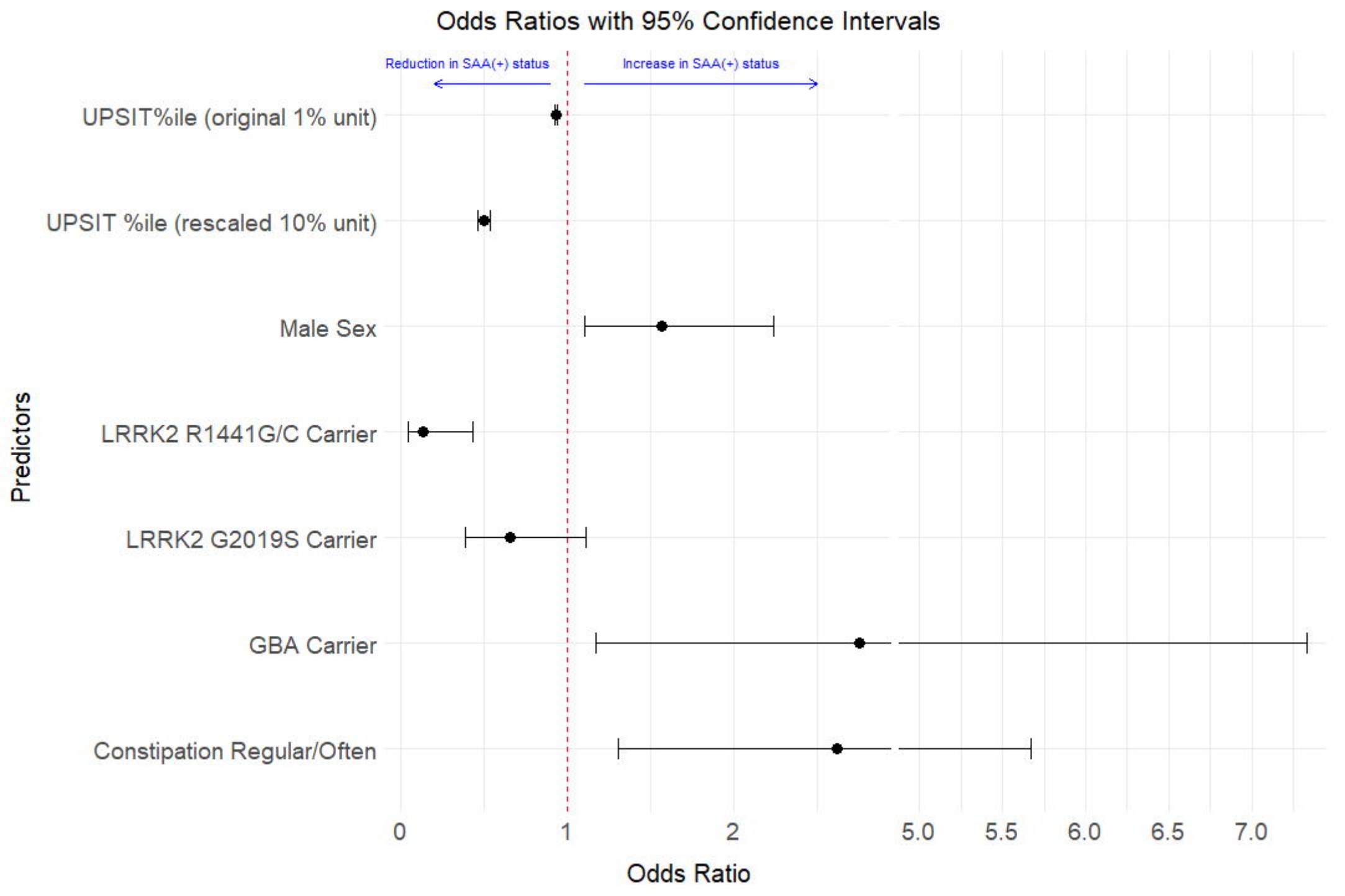
*Performance metrics were calculated based on the specific Youden Index Cut-Off for each model

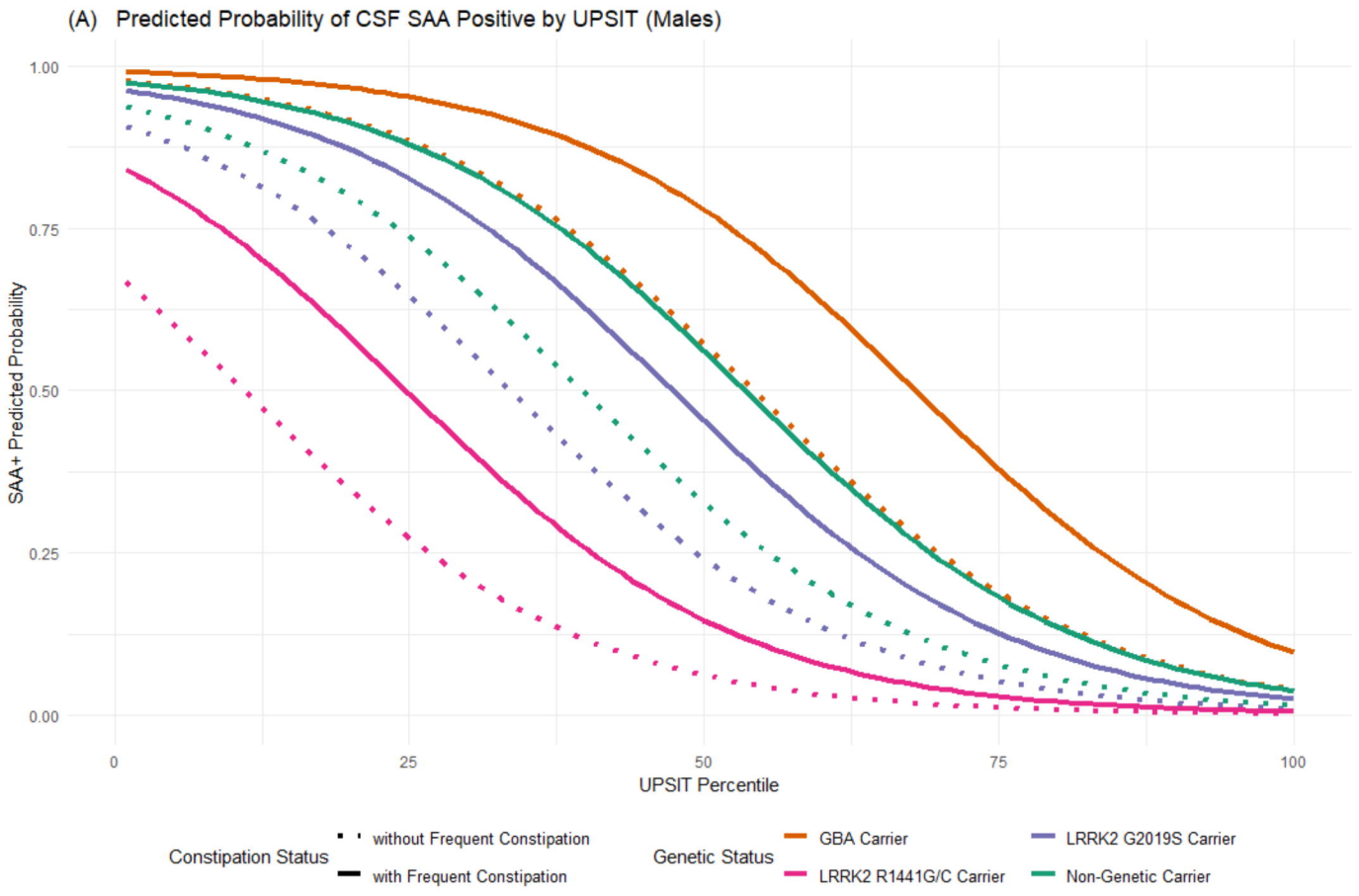
FIGURE LEGENDS

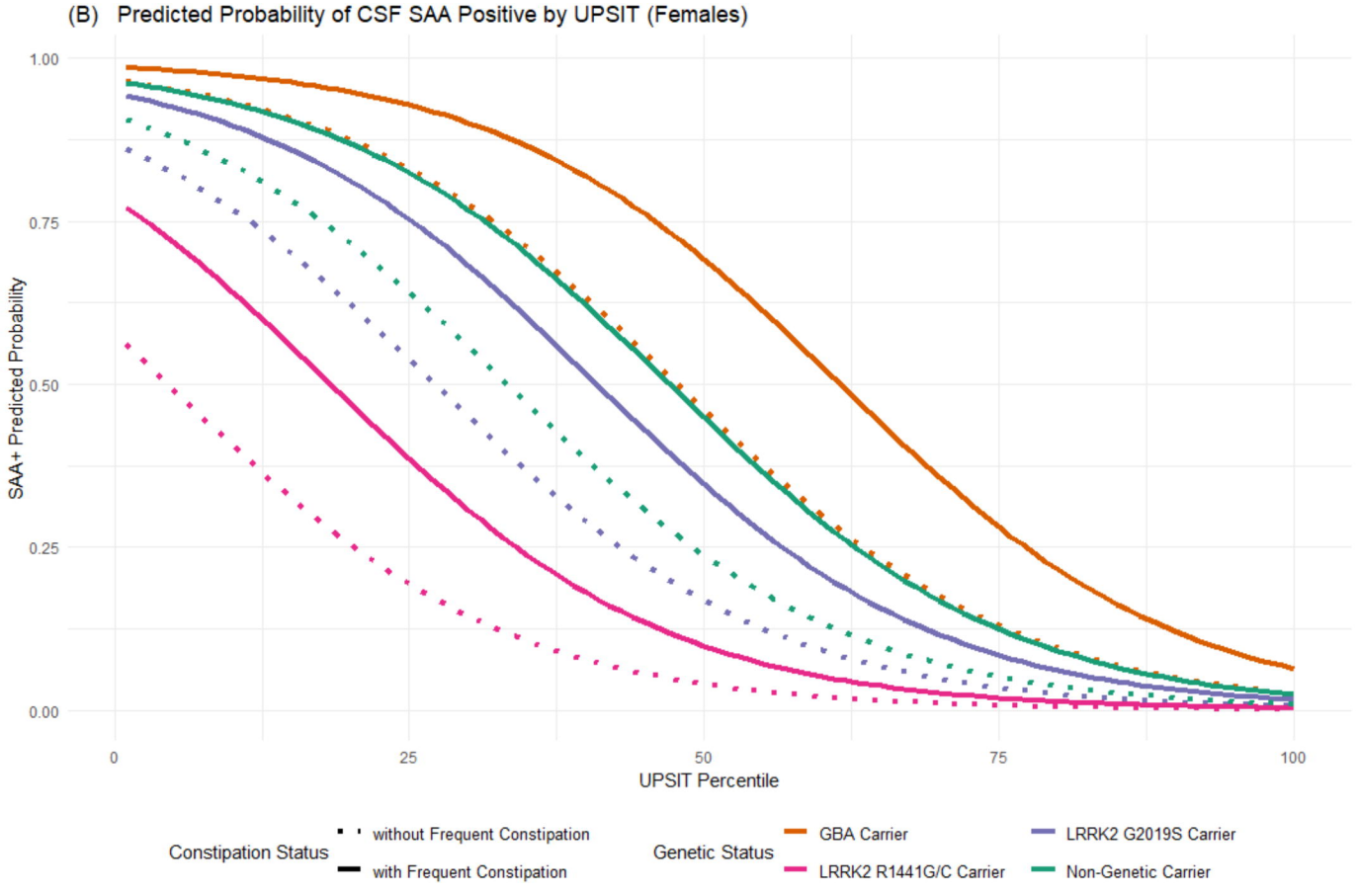
Figure 1. *Odds Ratios (with 95% Confidence Intervals) from Fully Adjusted Logistic Regression Model. Values less than one are associated with a reduced risk of positive alpha-synuclein results from the cerebrospinal fluid seeding amplification assay (a-syn SAA+), and values greater than one are associated with an increased risk of a-syn SAA+, relative to the average population sample. UPSIT percent correct are plotted using 1% and 10% unit changes. Plotted with 'ggbreak' R package.*[25]

Figure 2. *Probability of a-syn SAA(+) versus UPSIT age- and sex-specific percentile according to the fully adjusted model for (A) males and (B) females. Each line is shown for individuals who have a specific LRRK2 pathogenic variant, or carry GBA pathogenic variant, or do not carry those specific LRRK2 or GBA variants, while varying constipation status.*

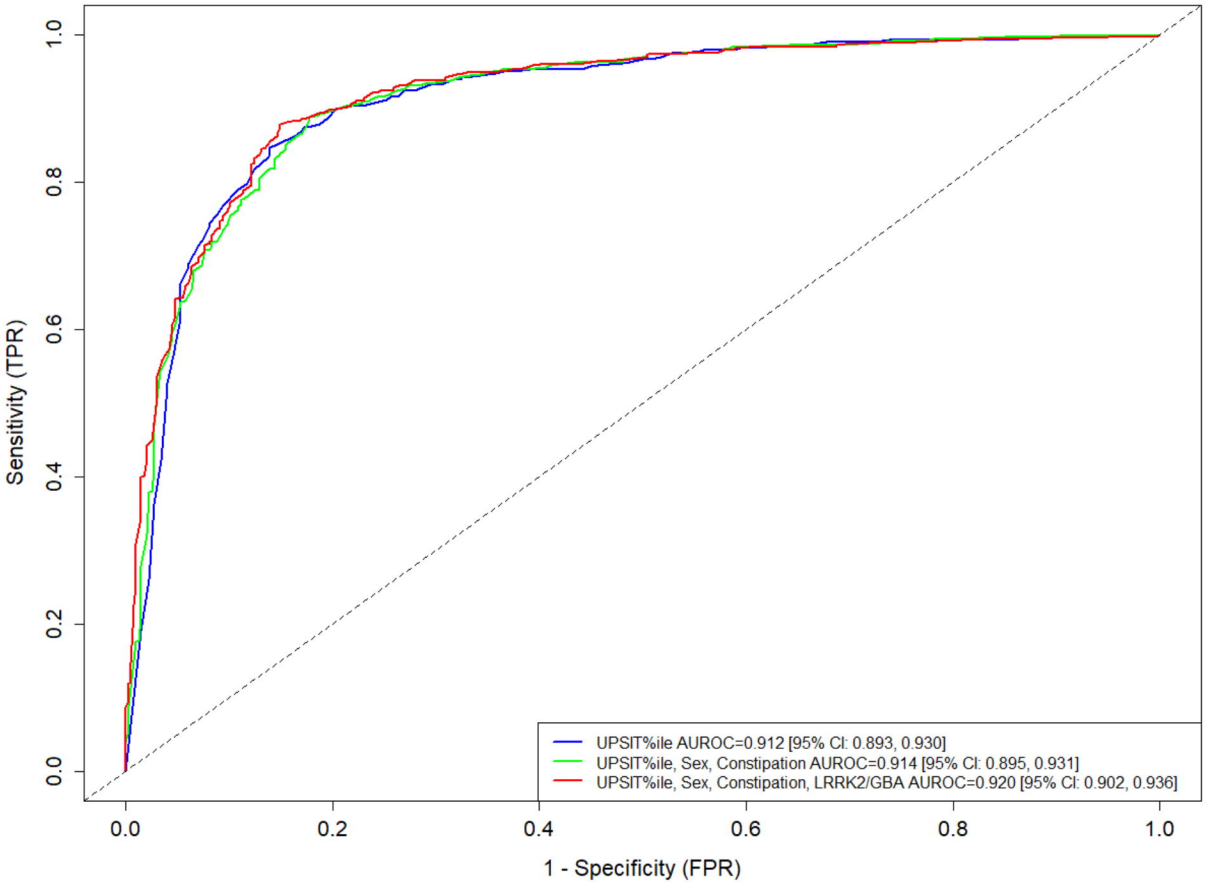
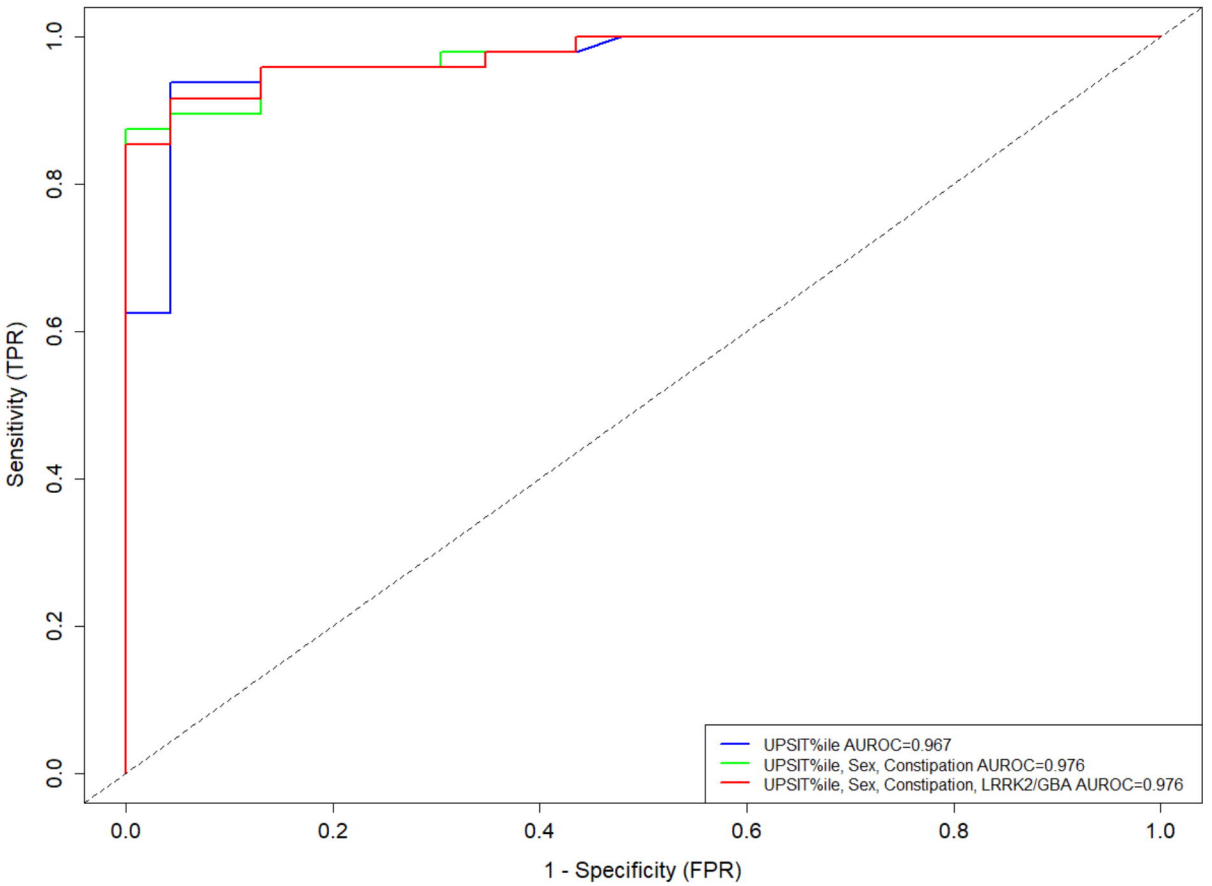
Figure 3. *Training and validation sample AUROC curves from partially and fully adjusted models using UPSIT age- and sex-specific percentile to predict a-syn SAA status. The figures show the (A) PPMI training sample AUROC curves and values with 95% Confidence Intervals [95% CI] from the bootstrap analysis, and (B) the S4 validation sample AUROC curves.*


Odds Ratios with 95% Confidence Intervals
Reduction in SAA(+) status
Increase in SAA(+) status
UPSIT%ile (original 1% unit)
UPSIT %ile (rescaled 10% unit)
Male Sex
LRRK2 R1441G/C Carrier
LRRK2 G2019S Carrier
GBA Carrier
Constipation Regular/Often
Predictors
0
1
2
5.0
5.5
6.0
6.5
7.0
Odds Ratio


(A) Predicted Probability of CSF SAA Positive by UPSIT (Males)
SAA+ Predicted Probability
1.00
0.75
0.50
0.25
0.00
0
25
50
75
100
UPSIT Percentile
Constipation Status
without Frequent Constipation
with Frequent Constipation
Genetic Status
GBA Carrier
LRRK2 G2019S Carrier
LRRK2 R1441G/C Carrier
Non-Genetic Carrier


(B) Predicted Probability of CSF SAA Positive by UPSIT (Females)
SAA+ Predicted Probability
1.00
0.75
0.50
0.25
0.00
0
25
50
75
100
UPSIT Percentile
Constipation Status
without Frequent Constipation
with Frequent Constipation
Genetic Status
GBA Carrier
LRRK2 G2019S Carrier
LRRK2 R1441G/C Carrier
Non-Genetic Carrier

(A) AUROC Curves for Logistic Models using UPSIT Percentile (PPMI)



(B) AUROC Curves for Logistic Models using UPSIT Percentile (S4 External Validation)