



Published in final edited form as:

J Neurol. 2015 September ; 262(9): 2064–2072. doi:10.1007/s00415-015-7821-5.

Sample size calculations for clinical trials targeting tauopathies: A new potential disease target

Jennifer L. Whitwell, PhD¹, Joseph R. Duffy, PhD², Edythe A. Strand, PhD², Mary M. Machulda, PhD³, Nirubol Tosakulwong, BS⁴, Stephen D. Weigand, MS⁴, Matthew L. Senjem, MS^{1,5}, Anthony J. Spychalla, BS¹, Jeffrey L. Gunter, PhD^{1,5}, Ronald C. Petersen, MD, PhD⁶, Clifford R. Jack Jr, MD¹, and Keith A. Josephs, MD, MST, MSc⁶

¹Department of Radiology, Mayo Clinic, Rochester, MN

²Department of Neurology (Division of Speech Pathology), Mayo Clinic, Rochester, MN

³Department of Psychiatry and Psychology (Neuropsychology), Mayo Clinic, Rochester, MN

⁴Department of Health Sciences Research (Biostatistics), Mayo Clinic, Rochester, MN

⁵Department of Information Technology, Mayo Clinic, Rochester, MN

⁶Department of Neurology (Divisions of Behavioral Neurology and Movement Disorders), Mayo Clinic, Rochester, MN

Abstract

Disease-modifying therapies are being developed to target tau pathology, and should, therefore, be tested in primary tauopathies. We propose that progressive apraxia of speech should be considered one such target group. In this study, we investigate potential neuroimaging and clinical outcome measures for progressive apraxia of speech and determine sample size estimates for clinical trials. We prospectively recruited 24 patients with progressive apraxia of speech who underwent two serial MRI with an interval of approximately two years. Detailed speech and language assessments included the Apraxia of Speech Rating Scale (ASRS) and Motor Speech Disorders (MSD) severity scale. Rates of ventricular expansion and rates of whole brain, striatal and midbrain atrophy were calculated. Atrophy rates across 38 cortical regions were also calculated and the regions that best differentiated patients from controls were selected. Sample size estimates required to power placebo-controlled treatment trials were calculated. The smallest sample size estimates were obtained with rates of atrophy of the precentral gyrus and supplementary motor area, with both measures requiring less than 50 subjects per arm to detect a 25% treatment effect with 80% power. These measures outperformed the other regional and global MRI measures and the clinical scales. Regional rates of cortical atrophy therefore provide the best outcome measures in progressive

Corresponding author: Jennifer L. Whitwell, Associate Professor of Radiology, Mayo Clinic, 200 1st St SW, Rochester, MN 55905, Tel: 507-284-5576, Fax: 507-284-9778, whitwell.jennifer@mayo.edu.

Ethical Standards

The study was approved by the Mayo Clinic Institutional Review Board and was therefore performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments. All subjects were consented for enrolment into the study.

Conflict of Interest

The authors declare that they have no conflict of interest.

apraxia of speech. The small sample size estimates demonstrate feasibility for including progressive apraxia of speech in future clinical treatment trials targeting tau.

Keywords

Clinical treatment trials; tau; rates; MRI; apraxia of speech; PPAOS

Introduction

Neurodegenerative diseases, such as Alzheimer's disease, are becoming increasingly prevalent as the population ages. They are becoming a huge societal burden, with Alzheimer's disease, for example, affecting more than 5 million Americans and costing American society \$214 billion annually. Hence, there is an increasingly pressing need to identify disease modifying treatments that can slow or halt progression of these diseases. Drug discovery efforts have focused on targeting the abnormal proteins found in the brain at autopsy, such as beta-amyloid and tau. In order to assess the effects of such treatments, clinical trials need to recruit patients in whom the predicted underlying pathology is homogeneous. In addition, there has been an increasing emphasis on recruiting patients early in their disease course when treatments are most likely to be effective.

A number of clinical treatment trials targeting tauopathies have focused on recruiting patients with progressive supranuclear palsy syndrome (PSPS) [1–3], given that a high proportion of patients with this diagnosis have an underlying neurodegenerative disorder characterized by the presence of a relatively pure tau pathology [4]. Neuroimaging outcome measures have also been developed for PSPS, with rate of midbrain atrophy consistently providing small sample size estimates for acceptably powered clinical treatment trials [5, 6]. It is, however, difficult to identify and recruit patients with PSPS early in their disease course, since few early clinical biomarkers are available. Hence, patients recruited into clinical trials typically have relatively advanced disease with significant functional impairment which may limit the potential treatment benefits.

We have demonstrated that the presence of a progressive apraxia of speech also has a close clinicopathological association with tau pathology [7–9]. Apraxia of speech affects the production of speech and is characterized by a cluster of speech features that most often include slow rate, articulatory distortions, distorted sound substitutions and segmentation of syllables in multisyllabic words or across words [10]. The diagnosis of progressive apraxia of speech is made when apraxia of speech is the dominant symptom of the neurodegenerative disorder [11]. Apraxia of speech can, however, occur in isolation as the only symptom of a neurodegenerative disorder and is then referred to as primary progressive apraxia of speech (PPAOS) [12]. At the time of presentation, most patients with progressive apraxia of speech, especially those with PPAOS, have little functional impairment (e.g. no reduction of speech intelligibility and no parkinsonism or cognitive impairment), suggesting a very early stage of the disease. Importantly, we and others have shown that patients with progressive apraxia of speech typically have underlying tau pathology without beta-amyloidosis [9, 12], with 100% of reported autopsy cases having tau [7–9, 13, 14]. In fact,

progressive apraxia of speech patients often eventually progress into a PSP-like syndrome [15]; hence little surprise that tau is the underlying pathology. No studies have assessed the prevalence of progressive apraxia of speech or PPAOS, although prevalence can be estimated to be approximately 4.4 per 100,000 persons based on the prevalence of progressive apraxia of speech compared to other speech and language disorders in our cohort [16], and previous prevalence studies of speech and language disorders [17, 18]. The typical age at onset of progressive apraxia of speech is also relatively young [11] which will limit the presence of age-related co morbid pathologies that could confound progression. These facts open up the exciting possibility that progressive apraxia of speech and, especially, PPAOS, could be excellent early phenotypic targets for clinical treatment trials targeting tau pathology.

In order for patients with progressive apraxia of speech to be included in such treatment trials, outcome measures will be needed by which to judge the success of the treatment. Neuroimaging outcome measures generally provide smaller sample size estimates than clinical outcome measures in PSPS [6], but the utility of neuroimaging metrics as outcome measures in progressive apraxia of speech is unknown. Patients with progressive apraxia of speech have relatively focal patterns of atrophy on MRI involving the premotor cortex, with mild involvement of the midbrain [11, 12, 19]. However, we have shown that brain atrophy progresses over time, with rates of atrophy in the premotor, motor, prefrontal and parietal cortex, midbrain and striatum greater than those observed in healthy controls [15]. We also found that rates of whole brain atrophy and ventricular expansion, two commonly utilized MRI outcomes measures, were greater in progressive apraxia of speech compared to healthy controls [15]. The aim of the current study was to utilize these neuroimaging measures to determine sample size estimates required for clinical treatment trials in progressive apraxia of speech, and ultimately to determine which neuroimaging measure or clinical measure would provide optimum sample size estimates. We also aimed to specifically assess sample size estimates in patients with PPAOS, since these patients may represent an earlier stage of the disease than patients that have started to develop additional symptoms.

Methods

Subjects

A total of 24 patients diagnosed with progressive apraxia of speech [11, 12] were prospectively recruited from the Department of Neurology, Mayo Clinic, into a NIH-funded grant assessing higher-level speech and language disorders between July 1st 2010 and July 1st 2014, and have since returned for serial testing and undergone a follow-up evaluation that included the identical clinical and neuroimaging protocol that was administered at baseline. All patients had undergone a detailed speech and language examination at baseline [12] and were diagnosed with progressive apraxia of speech if the dominant presenting sign was that of apraxia of speech and any other neurological and aphasia characteristics were absent or considered minimal and less severe. Of the 24 patients, 15 showed no more than equivocal neurological and aphasia characteristics and hence were diagnosed with PPAOS [12]. Diagnosis was made after review of video and audio recordings and speech and language test scores by two speech-language pathologists (J.R.D. and E.A.S.). Detailed

speech and language, neurological and neuropsychological data from 13 of the PPAOS patients at both baseline and follow-up have been previously reported [15]. We also previously reported rates of whole brain atrophy, ventricular expansion and midbrain atrophy for these 13 PPAOS patients [15].

Two clinical measures were selected as potential outcome measures for the study: (1) the Apraxia of Speech Rating Scale (ASRS) [20], which provides a composite measure of the presence, pervasiveness and severity of 16 features of AOS, with possible scores ranging from 0 (no abnormality for rated features) to 64 (all 16 features pervasively evidence and severe), and (2) the Motor Speech Disorders (MSD) scale (adapted from [21]) which rates the severity of functional speech impairment, with possible scores ranging from 1 (essentially nonvocal or completely unintelligible speech) to 10 (no perceptible speech abnormality). All 24 patients were scored using the MSD scale at both baseline and follow-up, and 22 patients were scored at both time-points using the ASRS. The ASRS could not be completed at follow-up on two patients due to lack of speech output. Rate of change in each clinical scale was calculated as number of points changed per year.

The progressive apraxia of speech cohort was matched by age, gender and scan interval to a cohort of 20 healthy control subjects that had two volumetric MRI performed using the same protocol as the progressive apraxia of speech patients. All healthy controls had been recruited into the Mayo Clinic Study of Aging (MCSA) and the controls were identified from the MCSA database. The demographic features of the progressive apraxia of speech patients, the PPAOS subset, and controls are shown in Table 1.

MRI acquisition

All subjects had both volumetric MRI performed at 3T using a standardized protocol on the same MRI scanner. A 3D magnetization prepared rapid acquisition gradient echo (MPRAGE) was performed at each time-point using the following parameters: TR/TE/T1, 2300/3/900 ms; flip angle 8°, 26-cm FOV; 256 × 256 in-plane matrix with a phase FOV of 0.94, voxel sizes of 1×1×1.2mm. All MPRAGE images underwent pre-processing correction for gradient nonlinearity and intensity non-uniformity.

MRI analysis

A number of whole-brain, subcortical and cortical neuroimaging metrics were selected as potential outcome measures for the study.

The whole brain metrics included rate of whole brain atrophy and rate of ventricular expansion. These metrics were selected because they have proven to be useful disease biomarkers in other neurodegenerative diseases and are already routinely utilized as outcome measures in clinical trials. All serial MRI were registered to baseline using 9 degrees-of-freedom registration. A differential-bias correction was then performed in order to balance image intensities across each scan-pair [22]. Annualized rates of whole brain atrophy and ventricular expansion were then calculated for each scan-pair using the boundary shift integral (BSI) [23, 24].

Rate of midbrain atrophy was also selected as an outcome measure because we have previously shown that midbrain volume is reduced in progressive apraxia of speech [19], and that rates of midbrain atrophy are increased compared to controls [15]. Midbrain volumes were calculated for each scan using atlas-based parcellation. A manually traced midbrain mask (Figure 1) was propagated onto each native-space MPRAGE. Annualized rates of atrophy were calculated as a change in volume expressed as a percentage of the baseline volume, divided by scan interval (i.e. $[\text{repeat volume} - \text{baseline volume} / \text{baseline volume} * 100] / \text{scan interval}$, years).

Regional cortical and subcortical metrics were also calculated utilizing an in-house developed version of tensor-based morphometry using symmetric normalization (TBM-SyN) [25]. The symmetric normalization (SyN) algorithm was applied to each 9 degrees-of-freedom registered and differential bias corrected scan-pair. The SyN deformation was calculated between each pair of scans, in both directions explicitly, saving an image of the log of the Jacobian determinants for each scan-pair. The deformation was applied to the moving image to create a “softmean” of the “fixed” and the “moved” image, resulting in a softmean image for both time-points. We then applied unified segmentation to each softmean image, and propagated automated anatomical labeling (AAL) [26] masks from template space to the softmean space, to obtain regional volume estimates as well as a regional atrophy measurement for every scan-pair. Based on our previous findings [15], we selected the striatum as a region-of-interest (ROI). We also identified a set of 38 cortical ROIs from the AAL atlas covering frontal, temporal, parietal and occipital cortices (Figure 2) and calculated rates of atrophy for each ROI. Data reduction was performed (described in the statistical analysis) to determine which of the 38 cortical ROIs provided the best discrimination between progressive apraxia of speech and controls, and between the PPAOS subset and controls, and hence would be considered as potential outcome measures for the study. This method of unbiased data reduction was chosen because a large number of ROIs showed significant differences between the patient groups and controls. Annualized rates of atrophy were calculated by normalizing the log of the Jacobian determinant by the interscan time interval, and computing the average within each ROI. The left and right hemispheres were combined for all TBM-SyN outcome measures given that progressive apraxia of speech is a relatively symmetric disease [11, 12, 19].

Statistical analysis

We summarized differences in the 38 cortical ROIs between progressive apraxia of speech and normal controls, and between PPAOS and the normal controls, using the area under the receiver operating characteristic curve (AUROC), a nonparametric effect size measure that is independent of the underlying scales of the data [27]. P-values and 95% confidence intervals for 38 ROIs were based on the Mann-Whitney U statistic divided by the product of the sample sizes [28]. In general, the higher the AUROC estimates, the better the discrimination between groups; in particular, the AUROC can be interpreted as the estimated proportion of times that a patient would have a higher rate of atrophy in that ROI compared to a normal subject. We selected nine cortical ROIs that showed AUROC estimates greater than 0.90 (precentral gyrus, supplementary motor area (SMA), paracentral gyrus, superior parietal lobe, postcentral gyrus, rolandic operculum, Broca’s area (pars

opercularis + pars triangularis), superior frontal lobe and middle frontal lobe) (Figure 1 and 2) as outcome measures to carry forward into our sample size calculations. Therefore, sample size estimates were calculated using rates of atrophy for the following 13 ROIs: whole brain, ventricle, striatum, midbrain, precentral gyrus, SMA, paracentral gyrus, superior parietal lobe, postcentral gyrus, rolandic operculum, Broca's area, superior frontal lobe and middle frontal lobe.

Estimated sample sizes needed for a hypothetical 24 month clinical trial were calculated to detect an effect size of a 25% reduction in the mean rate of atrophy in the 13 ROIs and the mean rate of clinical score change in the two clinical scales. All reported estimates are based on 80% or 90% power and a 5% two-sided Type I error rate. For the atrophy-based sample size estimates, we accounted for the fact that some degree of atrophy is expected even in normal subjects and that a desired effect size needs to take this into account. Specifically, we accounted for "background" atrophy in normal subjects by subtracting the mean rate of atrophy in normal control subjects from the mean rate of atrophy in the patients and basing the effect size on a 25% improvement in this difference. Specifically, the effect size was calculated as

$$\text{Effect size to detect} = 0.25 \times (\text{mean rate in patients} - \text{mean rate in controls}) / \text{standard deviation in patients}$$

For the clinical score estimates, we do not factor in the change in controls since it is assumed to be zero, given that none of the healthy control subjects had apraxia of speech.

To account for sampling variability in our estimates of the mean rates in patients and controls, and the standard deviation in patients, we used a parametric bootstrap procedure to generate $B = 1000$ replicates of each of these parameters via posterior simulation of an intercept-only regression model [29]. We present the sample size estimates for 80% and 90% powers ranked from the smallest sample size to the largest.

A 95% confidence interval that does not include the null value of interest, usually zero, can be interpreted as providing a statistical test that is significant at $p < 0.05$. We use this duality between confidence intervals and hypothesis tests to calculate p-values for tests of differences in sample size requirements. For each pairwise comparison, we estimated bootstrap confidence intervals of the difference in sample size and identified the widest confidence interval (i.e., one with the highest level) that still excluded the null value of zero. For example, if a 98% confidence interval did not include zero but a 99% confidence interval included zero, we define the p-value as $p = 0.02$.

Results

Sample size estimates per treatment arm for a placebo-controlled treatment trial are shown in Table 2 and Figure 3. The smallest sample size estimates for both the whole progressive apraxia of speech cohort, and the PPAOS subset, occurred with rates of atrophy of the precentral gyrus followed by the SMA, with both measures requiring less than 50 subjects per arm to detect a 25% treatment effect with 80% power. The precentral gyrus significantly

outperformed most of the other regional rates of atrophy, as well as rate of ventricular expansion and rate of change on the ASRS, in the whole cohort. However, rate of atrophy of the rolandic operculum, rate of expansion of the ventricles, and rate of change on the ASRS did perform well in the PPAOS subset, requiring less than 100 subjects per treatment arm at 80% power. Sample size estimates were generally lower for the PPAOS subset compared to the whole progressive apraxia of speech cohort for all metrics, except for rate of whole brain atrophy and rate of change on the MSD.

Discussion

This study provides evidence that utilizing progressive apraxia of speech as a target population would be feasible for future clinical treatment trials; given the small sample size estimates obtained with regional neuroimaging outcome measures.

The neuroimaging biomarkers that provided optimum sample size were rates of atrophy measured from the precentral and premotor regions of the brain, specifically the precentral gyrus and SMA. These findings concur with our previous work showing that progressive apraxia of speech is a relatively focal disease that first targets the SMA [11, 12] before spreading throughout the premotor cortex and into motor cortex [15]. We have demonstrated that the SMA is particularly associated with the severity of apraxia of speech [30]. The focal nature of the disease, therefore, allows the appropriate targeting of neuroimaging outcome measures and explains why global measures, such as rate of whole brain atrophy and rate of ventricular expansion, may have performed less well. The precentral gyrus and SMA provided optimal sample size in both the whole cohort and the smaller subset of patients with PPAOS, although sample sizes were slightly smaller in the PPAOS cohort. This suggests that targeting patients with an isolated apraxia of speech may be beneficial for future trials. These patients have more focal and less variable patterns of atrophy compared to subjects who also show mild aphasia [11].

Importantly, the sample size estimates obtained, especially from the precentral gyrus and SMA, were similar to those that we and others have previously reported in clinical and neuroimaging studies assessing PSPS [5, 6, 31], as well as studies assessing mild cognitive impairment and Alzheimer's disease [32]. Our study shows that by using the precentral gyrus as an outcome measure, only 31 PPAOS patients would be required per arm of a two year treatment trial in order to detect a 25% change in the rate of atrophy with 80% power. This compares to 63 subjects that would be required per treatment arm to detect a 30% change in rate of midbrain atrophy with 80% power in a one year trial of patients with PSPS [6]. In contrast to studies investigating PSPS, rates of midbrain atrophy performed poorly in progressive apraxia of speech and PPAOS. This concurs with our previous findings that, although midbrain atrophy is a feature of the disease [15], midbrain loss is less striking than observed in PSPS [19] and hence is a more appropriate outcome measure for trials targeting PSPS.

The sample size estimates calculated from the two clinical measures were larger than those observed for the precentral gyrus and SMA, albeit not significantly so. A potential limitation of the ASRS is that it cannot be scored in patients with very little or no speech output, as it

relies on scoring the severity of specific features of apraxia of speech [20]. This would not be a problem in trials seeking subjects very early in the disease course. While the MSD scale can be scored regardless of speech output, once patients become mute they will reach the limit of the scale and no further progression will be detected. Neuroimaging does not suffer from these limitations, since high quality scans can still be obtained from progressive apraxia of speech patients even when mute. In fact, it may be easier to obtain high quality MRI, that are not degraded by motion, in patients with progressive apraxia of speech compared to those with PSPS which have more severe parkinsonism and motor features [19].

The strengths of our study include the TBM-SyN technique for measuring regional rates of brain atrophy. This technique performs high-dimensional normalization across scan-pairs, fulfills the recommended requirements for symmetric normalization [33], whereby absolute changes from time A to time B equal the changes from time B to time A, and has proven sensitive to changes over time in patients with progressive apraxia of speech [15]. Others have similarly found that TBM provides a powerful method to track changes in the brain [32], suggesting that it should be considered as a method of choice for future clinical treatment trials. Our study also calculated sample size estimates accounting for change in normal aging to avoid producing unrealistically low sample size estimates and potentially under powering future studies [33]. The relatively small number of subjects in our cohort, particularly in the PPAOS subset, resulted in a large degree of overlap in confidence intervals across outcome measures. Caution is therefore needed when interpreting the specific sample size estimates. With-that said, it should be noted that our cohort size was larger than the two previous studies that have provided sample size estimates using neuroimaging for clinical trials in PSPS, which were based on only 16 [6] and 17 [5] subjects. Although we compared a number of regions, the modest advantage of the precentral and SMA regions seem unlikely to be due to chance alone. These two regions produced similar sample sizes estimates for both the whole cohort and the PPAOS subset, a similarity that adds validity to the findings. Further, that precentral and SMA perform well is not unexpected given these regions are close anatomical correlates of apraxia of speech. For these reasons we are confident that our findings are robust and that our sample sizes are not likely to be overly optimistic. The average scan interval in our cohort was also two years. It is likely that samples would need to be larger in a one year trial given the reduced power to detect change over the shorter interval [34]. As a guideline, one study on Alzheimer's disease observed an 18% increase in required sample size when going from a two to one year trial [35]. Our results should generalize to patients with disease duration of approximately four years.

Our findings suggest that patients with progressive apraxia of speech could be included in future clinical trials assessing disease modifying treatments aimed at targeting tau, and they could even be considered as another target patient population for such trials. We have demonstrated that rates of regional atrophy provide optimum outcome measures for this population and reported sample size estimates to help guide the development of future clinical trials.

Acknowledgments

This study was funded by R01-DC12519 (PI Whitwell), R01-DC010367 (PI Josephs) and U01-AG06786 (PI Petersen).

References

1. Boxer AL, Lang AE, Grossman M, Knopman DS, Miller BL, Schneider LS, et al. Davunetide in patients with progressive supranuclear palsy: a randomised, double-blind, placebo-controlled phase 2/3 trial. *The Lancet Neurology*. 2014; 13:676–685. [PubMed: 24873720]
2. Litvan I, Agid Y, Calne D, Campbell G, Dubois B, Duvoisin RC, et al. Clinical research criteria for the diagnosis of progressive supranuclear palsy (Steele-Richardson-Olszewski syndrome): report of the NINDS-SPSP international workshop. *Neurology*. 1996; 47:1–9. [PubMed: 8710059]
3. Tolosa E, Litvan I, Hoglinger GU, Burn D, Lees A, Andres MV, et al. A phase 2 trial of the GSK-3 inhibitor tideglusib in progressive supranuclear palsy. *Movement disorders : official journal of the Movement Disorder Society*. 2014; 29:470–478. [PubMed: 24532007]
4. Josephs KA, Dickson DW. Diagnostic accuracy of progressive supranuclear palsy in the Society for Progressive Supranuclear Palsy brain bank. *Movement disorders : official journal of the Movement Disorder Society*. 2003; 18:1018–1026. [PubMed: 14502669]
5. Paviour DC, Price SL, Lees AJ, Fox NC. MRI derived brain atrophy in PSP and MSA-P. Determining sample size to detect treatment effects. *Journal of neurology*. 2007; 254:478–481. [PubMed: 17401522]
6. Whitwell JL, Xu J, Mandrekar JN, Gunter JL, Jack CR Jr, Josephs KA. Rates of brain atrophy and clinical decline over 6 and 12-month intervals in PSP: determining sample size for treatment trials. *Parkinsonism & related disorders*. 2012; 18:252–256. [PubMed: 22079523]
7. Golden E, Duffy JR, Strand EA, Parisi JE, Dickson DW, Josephs KA. Primary progressive apraxia of speech associated with progressive supranuclear palsy pathology. *Am J Neurodegener Dis*. 2014; 3:131. conference abstract.
8. Josephs KA, Boeve BF, Duffy JR, Smith GE, Knopman DS, Parisi JE, et al. Atypical progressive supranuclear palsy underlying progressive apraxia of speech and nonfluent aphasia. *Neurocase*. 2005; 11:283–296. [PubMed: 16093229]
9. Josephs KA, Duffy JR, Strand EA, Whitwell JL, Layton KF, Parisi JE, et al. Clinicopathological and imaging correlates of progressive aphasia and apraxia of speech. *Brain : a journal of neurology*. 2006; 129:1385–1398. [PubMed: 16613895]
10. Duffy, JR. *Motor speech disorders: substrates, differential diagnosis, and management*. Mosby; St Louis, MI: 2005.
11. Josephs KA, Duffy JR, Strand EA, Machulda MM, Senjem ML, Lowe VJ, et al. Syndromes dominated by apraxia of speech show distinct characteristics from agrammatic PPA. *Neurology*. 2013; 81:337–345. [PubMed: 23803320]
12. Josephs KA, Duffy JR, Strand EA, Machulda MM, Senjem ML, Master AV, et al. Characterizing a neurodegenerative syndrome: primary progressive apraxia of speech. *Brain : a journal of neurology*. 2012; 135:1522–1536. [PubMed: 22382356]
13. Deramecourt V, Lebert F, Debachy B, Mackowiak-Cordoliani MA, Bombois S, Kerdraon O, et al. Prediction of pathology in primary progressive language and speech disorders. *Neurology*. 2010; 74:42–49. [PubMed: 19940270]
14. Dickson DW, Ahmed Z, Algom AA, Tsuboi Y, Josephs KA. Neuropathology of variants of progressive supranuclear palsy. *Current opinion in neurology*. 2010; 23:394–400. [PubMed: 20610990]
15. Josephs KA, Duffy JR, Strand EA, Machulda MM, Senjem ML, Gunter JL, et al. The evolution of primary progressive apraxia of speech. *Brain : a journal of neurology*. 2014; 137:2783–2795. [PubMed: 25113789]
16. Wicklund MR, Duffy JR, Strand EA, Whitwell JL, Machulda MM, Josephs KA. Aphasia with left occipitotemporal hypometabolism: a novel presentation of posterior cortical atrophy? *Journal of*

- clinical neuroscience : official journal of the Neurosurgical Society of Australasia. 2013; 20:1237–1240. [PubMed: 23850398]
17. Borroni B, Alberici A, Grassi M, Turla M, Zanetti O, Bianchetti A, et al. Is frontotemporal lobar degeneration a rare disorder? Evidence from a preliminary study in Brescia county, Italy. *Journal of Alzheimer's disease : JAD*. 2010; 19:111–116. [PubMed: 20061630]
 18. Johnson JK, Diehl J, Mendez MF, Neuhaus J, Shapira JS, Forman M, et al. Frontotemporal lobar degeneration: demographic characteristics of 353 patients. *Archives of neurology*. 2005; 62:925–930. [PubMed: 15956163]
 19. Whitwell JL, Duffy JR, Strand EA, Machulda MM, Senjem ML, Gunter JL, et al. Neuroimaging comparison of primary progressive apraxia of speech and progressive supranuclear palsy. *Eur J Neurol*. 2013; 20:629–637. [PubMed: 23078273]
 20. Strand EA, Duffy JR, Clark HM, Josephs K. The apraxia of speech rating scale: A tool for diagnosis and description of apraxia of speech. *Journal of communication disorders*. 2014
 21. Yorkson K, Strand EA, Miller R, Hillel A, Smith K. Speech deterioration in amyotrophic lateral sclerosis: Implications for the timing of intervention. *J Medical Speech-Language Pathology*. 1993; 1:35–46.
 22. Lewis EB, Fox NC. Correction of differential intensity inhomogeneity in longitudinal MR images. *NeuroImage*. 2004; 23:75–83. [PubMed: 15325354]
 23. Freeborough PA, Fox NC. The boundary shift integral: an accurate and robust measure of cerebral volume changes from registered repeat MRI. *IEEE Trans Med Imaging*. 1997; 16:623–629. [PubMed: 9368118]
 24. Gunter JL, Shiung MM, Manduca A, Jack CR Jr. Methodological considerations for measuring rates of brain atrophy. *J Magn Reson Imaging*. 2003; 18:16–24. [PubMed: 12815635]
 25. Jack CR Jr, Wiste HJ, Knopman DS, Vemuri P, Mielke MM, Weigand SD, et al. Rates of beta-amyloid accumulation are independent of hippocampal neurodegeneration. *Neurology*. 2014; 82:1605–1612. [PubMed: 24706010]
 26. Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*. 2002; 15:273–289. [PubMed: 11771995]
 27. Acion L, Peterson JJ, Temple S, Arndt S. Probabilistic index: an intuitive non-parametric approach to measuring the size of treatment effects. *Statist Med*. 2006; 25:591–602.
 28. Newcombe RG. Confidence intervals for an effect size measure based on the Mann-Whitney statistic. Part 1: general issues and tail-area-based methods. *Statist Med*. 2006; 25:543–557.
 29. Gelman, A.; Hill, J. *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press; New York: 2007.
 30. Whitwell JL, Duffy JR, Strand EA, Xia R, Mandrekar J, Machulda MM, et al. Distinct regional anatomic and functional correlates of neurodegenerative apraxia of speech and aphasia: an MRI and FDG-PET study. *Brain and language*. 2013; 125:245–252. [PubMed: 23542727]
 31. Litvan I, Kong M. Rate of decline in progressive supranuclear palsy. *Movement disorders : official journal of the Movement Disorder Society*. 2014; 29:463–468. [PubMed: 24615741]
 32. Hua X, Lee S, Yanovsky I, Leow AD, Chou YY, Ho AJ, et al. Optimizing power to track brain degeneration in Alzheimer's disease and mild cognitive impairment with tensor-based morphometry: an ADNI study of 515 subjects. *NeuroImage*. 2009; 48:668–681. [PubMed: 19615450]
 33. Fox NC, Ridgway GR, Schott JM. Algorithms, atrophy and Alzheimer's disease: cautionary tales for clinical trials. *NeuroImage*. 2011; 57:15–18. [PubMed: 21296168]
 34. Ard MC, Edland SD. Power calculations for clinical trials in Alzheimer's disease. *Journal of Alzheimer's disease : JAD*. 2011; 26(Suppl 3):369–377. [PubMed: 21971476]
 35. Hua X, Lee S, Hibar DP, Yanovsky I, Leow AD, Toga AW, et al. Mapping Alzheimer's disease progression in 1309 MRI scans: power estimates for different inter-scan intervals. *NeuroImage*. 2010; 51:63–75. [PubMed: 20139010]

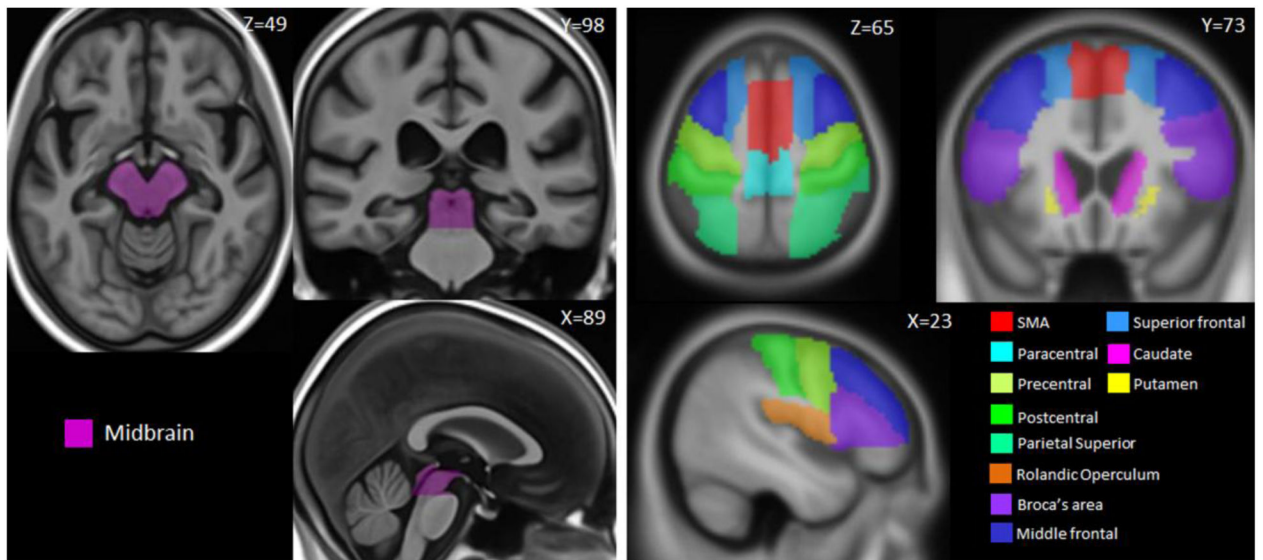


Figure 1. MRI atlas-based regions-of-interest. The left panel shows the manually traced midbrain mask, and the right panel shows the selected regions of-interest from the automated anatomical labeling atlas.

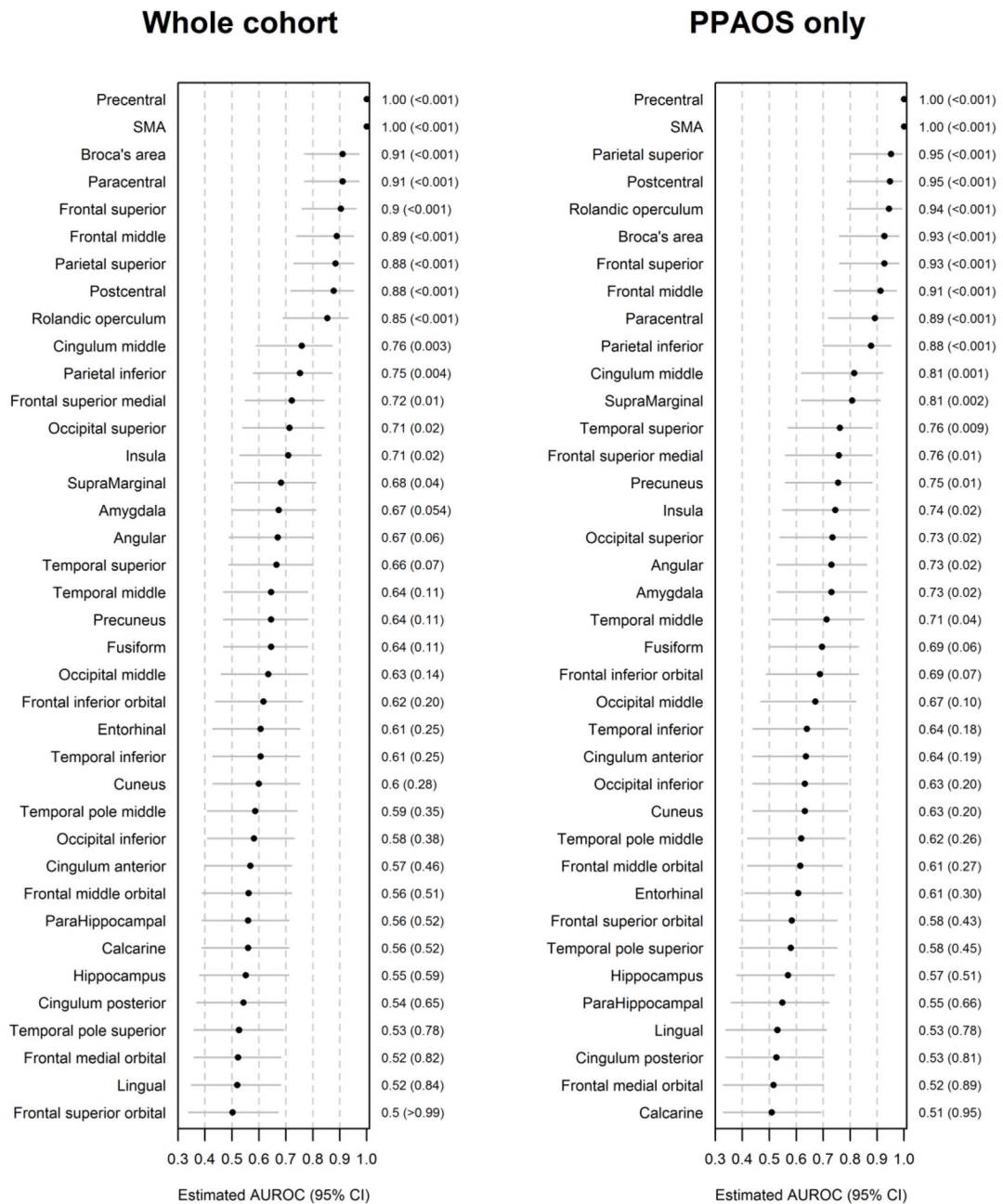


Figure 2. Forest plots illustrating discrimination of the whole cohort of progressive apraxia of speech patients, and for the PPAOS subset, from controls for the 38 cortical ROIs. Regions-of-interest are ranked from best (rank of 1) to worst according to AUROC, and p values and 95% confidence intervals are shown.

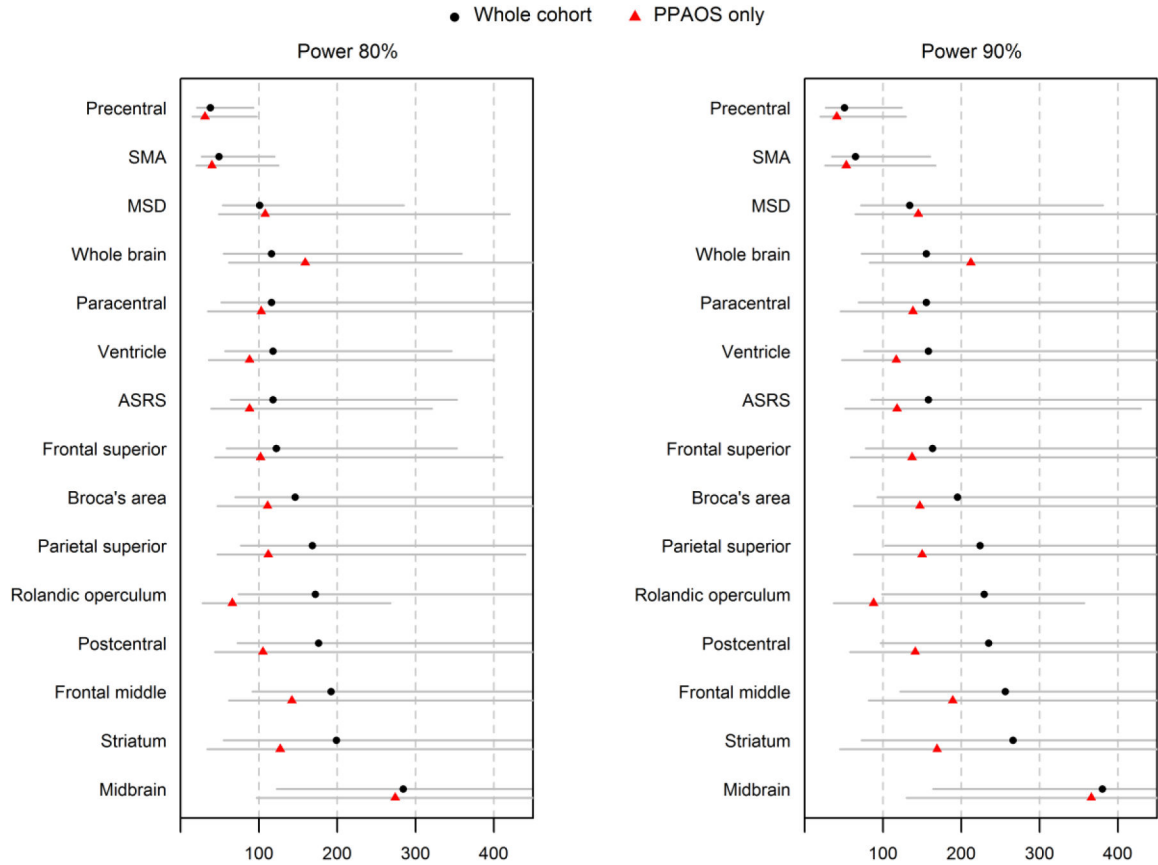


Figure 3.

Forest plots illustrating sample size estimates and 95% bootstrap confidence intervals for all neuroimaging and clinical outcome measures. Sample size estimates are given for both the whole progressive apraxia of speech cohort and the PPAOS subset, at 80% and 90% power. The 95% CIs are shown to account for uncertainty in the means and SDs used in the sample size formulas. SMA = supplementary motor area; ASRS = Apraxia of Speech Rating Scale; MSD = Motor Speech Disorders scale.

Table 1

Subject demographics

	Total progressive apraxia of speech cohort (n=24)	PPAOS subset (n=15)	Controls (n=20)
Age at baseline MRI, years	72 (65–76)	74 (67–76)	72 (67–76)
Gender, % female	15 (62%)	8 (53%)	11 (55%)
Scan interval, years	2 (1.5–2.4)	2 (1.7–2.6)	2 (1.7–2.6)
Disease duration, years	4 (2.6–4.9)	4 (2.9–5.4)	NA
ASRS at baseline (/64, 0 normal)	16 (12–23)	16 (12–20)	NA
ASRS at follow-up (/64, 0 normal) ^a	26 (20–31)	25 (20–31)	NA
MSD scale at baseline (/10, 10 normal)	7 (6–8)	7 (7–8)	NA
MSD scale at follow-up (/10, 10 normal)	6 (4–6)	6 (4–6)	NA
WAB-AQ at baseline (/100, 100 normal)	95 (85–98)	97 (95–100)	NA
WAB-AQ at follow-up (/100, 100 normal) ^b	87 (82–96)	96 (87–97)	NA

Data shown as median (inter-quartile range); ASRS = Apraxia of Speech Rating Scale; MSD = Motor Speech Disorder scale; WAB – AQ = Western Aphasia Battery Aphasia Quotient; no significant differences were observed between progressive apraxia of speech and controls or between PPAOS and controls.

^aTwo patients (one diagnosed with PPAOS) were unable to complete the ASRS at follow-up.

^bFour patients (two diagnosed with PPAOS) were unable to complete the WAB-AQ at follow-up

Table 2

Mean change and sample size estimates for the whole cohort of progressive apraxia of speech patients and for the subset of PPAOS patients. Sample sizes represent number of subjects needed per arm to detect a 25% treatment effect in a two year clinical trial accounting for the rate of change in controls. We report 95% CIs for the sample size to account for uncertainty in the means and SDs used in the sample size formulas.

ROI ^a	Mean change (standard deviation)		Number of subjects needed per arm (95% CI)					
	Whole cohort		Control		Whole cohort		PPAOS subset only	
	Whole cohort	PPAOS subset only	Control	power 80%	power 90%	power 80%	power 90%	
Precentral	-3.9 (1.6)	-3.4 (1.2)	0.19 (1.5)	38 (21, 93)	51 (27, 124)	31 (15, 97)	41 (20, 129)	
SMA	-4.1 (1.8)	-3.7 (1.4)	-0.11 (1.3)	49 (27, 120)	65 (35, 160)	40 (20, 125)	53 (26, 167)	
MSD	0.97 (0.61)	0.98 (0.64)	0 (0)	101 (54, 285)	134 (72, 381)	108 (49, 420)	145 (65, 561)	
Whole brain	-1.5 (0.72)	-1.4 (0.73)	-0.44 (0.5)	116 (55, 359)	155 (73, 480)	159 (62, 762) ^c	212 (83, 1020) ^c	
Paracentral	-2 (1.3)	-1.9 (1.2)	-0.04 (1.5)	116 (52, 467)	155 (69, 624)	103 (35, 551)	138 (46, 737)	
Ventricle	9.3 (4)	8.4 (2.9)	3.4 (2.5)	118 (57, 346) ^c	158 (76, 462) ^c	88 (36, 400)	117 (48, 534)	
ASRS ^b	-4.4 (3)	-4.4 (2.6)	0 (0)	118 (64, 353) ^c	158 (85, 472) ^c	88 (39, 321)	118 (52, 429)	
Frontal superior	-2.3 (1.5)	-2.2 (1.3)	-0.13 (0.85)	122 (59, 353)	163 (78, 472)	102 (44, 411)	137 (59, 550)	
Broca's area	-3.8 (2.9)	-3.7 (2.5)	0.045 (1.6)	146 (70, 456) ^c	195 (93, 610) ^c	111 (47, 538)	147 (63, 719)	
Parietal superior	-2 (1.6)	-2.2 (1.5)	0.008 (0.94)	168 (77, 626) ^c	224 (103, 837) ^c	112 (47, 440)	150 (63, 588)	
Rolandic operculum	-1.6 (1.4)	-1.7 (0.91)	0.07 (0.97)	172 (74, 686) ^c	229 (99, 918) ^c	66 (28, 268)	88 (37, 357)	
Postcentral	-1.5 (1.2)	-1.5 (0.99)	0.04 (0.8)	176 (73, 652) ^c	235 (97, 872) ^c	105 (44, 502)	141 (58, 672)	
Frontal middle	-1.8 (1.6)	-1.8 (1.4)	0.045 (0.63)	192 (92, 692) ^{c,d}	256 (122, 926) ^{c,d}	142 (62, 573) ^c	189 (82, 767) ^c	
Striatum	-2.4 (2.2)	-2.6 (1.9)	0.07 (3.9)	199 (55, 4983) ^c	266 (73, 6671) ^c	127 (34, 2434)	169 (45, 3258)	
Midbrain	-1.6 (1.9)	-1.6 (1.9)	0.19 (0.74)	284 (123, 1298) ^{c,d}	380 (164, 1737) ^{c,d}	274 (97, 3458) ^{c,d}	366 (130, 4628) ^{c,d}	

SMA = supplementary motor area; ASRS = Apraxia of Speech Rating Scale; MSD = Motor Speech Disorder scale

^aValues shown for MRI-based measures are expressed as annualized percentage change while values shown for ASRS and MSD are annual change in points.

^bTwo patients (one diagnosed with PPAOS) were unable to complete the ASRS at follow-up.

^cSample size estimates significantly larger than precentral sample size

^dSample size estimates significantly larger than SMA sample size