BMJ Open

Remote versus face-to-face neuropsychological testing for dementia research: a comparative study in people with Alzheimer's disease, frontotemporal dementia and healthy older individuals

Maï-Carmen Requena-Komuro ,[1,2] Jessica Jiang,[1] Lucianne Dobson,[1] Elia Benhamou,[1,3] Lucy Russell,[1] Rebecca L Bond,[1] Emilie V Brotherhood,[1] Caroline Greaves,[1] Suzie Barker,[1] Jonathan D Rohrer ,[1] Sebastian J Crutch,[1] Jason D Warren ,[1] Chris JD Hardy [1]

To cite: Requena-Komuro M-C, Jiang J, Dobson L, *et al*. Remote versus face-to-face neuropsychological testing for dementia research: a comparative study in people with Alzheimer's disease, frontotemporal dementia and healthy older individuals. *BMJ Open* 2022;**12**:e064576. doi:10.1136/bmjopen-2022-064576

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2022-064576).

M-CR-K and JJ contributed equally.

M-CR-K and JJ are joint first authors.

Received 10 May 2022
Accepted 20 October 2022

For numbered affiliations see end of article.

Correspondence to
Dr Chris JD Hardy;
chris.hardy@ucl.ac.uk

ABSTRACT

Objectives We explored whether adapting neuropsychological tests for online administration during the COVID-19 pandemic was feasible for dementia research.

Design We used a longitudinal design for healthy controls, who completed face-to-face assessments 3–4 years before remote assessments. For patients, we used a cross-sectional design, contrasting a prospective remote cohort with a retrospective face-to-face cohort matched for age/education/severity.

Setting Remote assessments were conducted using video-conferencing/online testing platforms, with participants using a personal computer/tablet at home. Face-to-face assessments were conducted in testing rooms at our research centre.

Participants The remote cohort comprised 25 patients (n=8 Alzheimer's disease (AD); n=3 behavioural variant frontotemporal dementia (bvFTD); n=4 semantic dementia (SD); n=5 progressive non-fluent aphasia (PNFA); n=5 logopenic aphasia (LPA)). The face-to-face patient cohort comprised 64 patients (n=25 AD; n=12 bvFTD; n=9 SD; n=12 PNFA; n=6 LPA). Ten controls who previously participated in face-to-face research also took part remotely.

Outcome measures The outcome measures comprised the strength of evidence under a Bayesian framework for differences in performances between testing environments on general neuropsychological and neurolinguistic measures.

Results There was substantial evidence suggesting no difference across environments in both the healthy control and combined patient cohorts (including measures of working memory, single-word comprehension, arithmetic and naming; Bayes Factors $(BF)_{01}$ >3), in the healthy control group alone (including measures of letter/category fluency, semantic knowledge and bisyllabic word repetition; all BF_{01} >3), and in the combined patient cohort alone (including measures of working memory, episodic memory, short-term verbal memory, visual perception, non-word reading, sentence comprehension and bisyllabic/trisyllabic word repetition; all BF_{01} >3). In the control cohort alone, there was substantial evidence in support of a difference across environments for tests of visual perception (BF_{01}=0.0404) and monosyllabic word repetition (BF_{01}=0.0487).

Conclusions Our findings suggest that remote delivery of neuropsychological tests for dementia research is feasible.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ Diverse patient cohorts representing rare dementias with specific communication difficulties.
⇒ Sampling of diverse and relevant neuropsychological domains.
⇒ Use of Bayesian statistics to quantify the strength of evidence for the putative null hypothesis (no effect between remote and face-to-face testing).
⇒ Relatively small cohort sizes.
⇒ Lack of direct head-to-head comparisons of test environment in the same patients.

INTRODUCTION

The COVID-19 pandemic and associated social distancing and lockdown measures imposed a series of daunting challenges for conducting research with people with dementia. In the UK, three national lockdowns between March 2020 and February 2021 largely prevented face-to-face research. People with dementia are at increased risk of COVID-19,[1] and many participants understandably did not feel safe to travel for research, particularly before widespread vaccination was implemented. Here, we

describe our attempts to translate our traditional neuropsychological and neurolinguistic test batteries (typically administered face-to-face) for remote administration.

Development and implementation of online cognitive assessments for patients with dementia, particularly within communities who experience difficulties in accessing clinical care are not new.[2] Telemedicine has been previously used successfully in Alzheimer's disease (AD)[3 4] and with rarer dementias such as primary progressive aphasia (PPA)[5–7] and behavioural variant frontotemporal dementia (bvFTD).[7] However, due to COVID-19, there has been a more pervasive shift towards the use of online methods to meet clinical, support and research needs.[8 9]

A review by Hunter and colleagues[10] summarises 20 years of research comparing face-to-face and online administration of cognitive tests in healthy older adults (≥40 years old) and participants diagnosed with mild cognitive impairment, AD or other types of dementia (often unspecified). The authors identified 12 studies that used video-conferencing methods. Overall, there was clear evidence to suggest that remote cognitive testing for people living with AD and other forms of dementia is feasible. Additionally, there is evidence to suggest that online performance remains stable over time (with a maximum delay of 3 months between assessments), particularly for the domains of executive function, working memory, verbal episodic memory and language. Minimal evidence was available for visuospatial tasks, and tests of single word and sentence comprehension.

Notwithstanding considerable progress in this area to date, further research into the feasibility of remote neuropsychological testing of patients with neurodegenerative diseases is required. There are three main considerations that need addressing: (1) Adaptation: are similar performance outcomes obtained on neuropsychological tests designed for face-to-face administration when given remotely? If performance across modalities is equivalent, then this could allow for pooling of data collected face-to-face and remotely, potentially allowing more equitable access to research for participants who are not able to physically travel to research centres. (2) Demand: to what extent are research participants with and without dementia willing to engage in remote neuropsychological research? (3) Acceptability: how satisfactory is remote testing for research participants including those with diverse forms of dementia?[11]

Based largely on the face-to-face protocol for general neuropsychological and neurolinguistic testing used at our research centre, we built a protocol for remote testing of patients diagnosed with typical AD, patients representing major variants of PPA (semantic dementia (SD), progressive non-fluent aphasia (PNFA), logopenic aphasia (LPA)) and bvFTD. Patients were tested from their homes via the widely used video-conferencing software, Zoom (Zoom Video Communications). We also recruited a small cohort of healthy older adults who had taken part in our face-to-face research at the Dementia Research Centre 3–4 years before the pandemic. Here, we compared the healthy controls' performance on several neuropsychological and neurolinguistic tests between the two testing environments (face-to-face vs remote). We also compared the performance of patients tested remotely with a historical face-to-face cohort of patients chosen to represent the same syndromes and to match the remote cohort based on age, education and symptom duration. We adopted a Bayesian approach that assesses the amount of evidence in favour of the null hypothesis (ie, that there is no significant difference in performance on a given neuropsychological task between testing environments) relative to the alternative hypothesis (ie, that there is a significant difference in performance on a given neuropsychological task between testing environments).

Following previous research,[10] we did not predict major differences in terms of participants' performances when tested face-to-face and remotely on most neuropsychological and neurolinguistic tests. However, we did consider the potential for poorer performance on tests of speech perception that were administered remotely, given additional difficulties associated with controlling the remote auditory environment.

METHODS

Participant recruitment and group matching

Recruitment for the study took place between February and August 2021. Potential patient participants were identified via the Specialist Cognitive Disorders Clinic at the National Hospital for Neurology and Neurosurgery, direct research referrals from external clinicians or via Rare Dementia Support (www.raredementiasupport.org); healthy controls were recruited via our research participant database. Eighty-seven potentially eligible participants were identified, and all of these were contacted.

An initial telephone screen was conducted for each participant to establish they had access to the necessary equipment (tablet or desktop/laptop computer), a broadband internet connection, a quiet testing space to support the remote research assessment, and no preclusive hearing or visual impairments. We also performed the telephone version of the Mini-Mental State Examination (T-MMSE) with patients to assess their disease severity.[12 13] A minimum score of 12 on the T-MMSE (which corresponds to a converted MMSE score of 16) was used as an inclusion criterion.[12] No participants were excluded after the telephone screen.

Twenty-five patients (eight with typical AD, three bvFTD, four SD, five PNFA, five LPA) were recruited for the remote study. For comparison purposes, a reference historical cohort comprising 64 patients (25 with AD, 12 bvFTD, 9 SD, 12 PNFA, 6 LPA) who had undertaken a face-to-face research assessment at our centre between 2013 and 2020 was selected, matching the cohort assessed remotely as closely as possible for syndromic composition, age, years of education and symptom duration. Henceforth, these are referred to as the 'remote' and

'face-to-face' patient cohorts, respectively. All patients fulfilled consensus diagnostic criteria for the relevant syndromic diagnosis[14–16] and all had clinically mild-to-moderate severity disease. Where available, brain MRI was consistent with the syndromic diagnosis, without evidence of significant cerebrovascular burden. Ten healthy older individuals with no history of neurological or psychiatric illness and who had been seen for face-to-face testing 3–4 years previously also underwent remote assessments. The neuropsychological tests reported in this research article were not used for diagnostic purposes. Demographic and clinical details for all participants are summarised in table 1. All participants gave informed consent for their involvement in the study.

Testing procedure: face-to-face

Data for the reference historical cohort were collected under our face-to-face research assessment protocol, as delivered in experimental sessions at the Dementia Research Centre between 2013 and 2020. Under this protocol, all neuropsychological tests were administered in dedicated quiet testing rooms, with the participant sitting opposite the experimenter. Patients were predominantly tested on their own, unless the informant accompanying them to the study visit requested to be present and the participant agreed to this. In these cases, the informant was explicitly asked not to intervene during testing. No feedback was given on performance and no time limits were imposed (unless timing was intrinsic to the test). A battery of general neuropsychological and neurolinguistic tests (see tables 2 and 3) were administered, following standard methods. The neurolinguistic test was developed specifically to characterise the language profiles of people with PPA and therefore, was not administered to participants with bvFTD or AD.

Modifying the face-to-face battery of neuropsychological tests for remote delivery

We reviewed the battery of general neuropsychological and neurolinguistic tests that had been used historically at our centre for face-to-face administration, in order to identify tests that could be feasibly delivered remotely online while preserving the overall structure of the tests and sampling across cognitive domains as far as possible (see online supplemental table 1). Where a task required visual stimulus presentation, a high-quality copy of the stimuli was made. Images were then imported into Microsoft PowerPoint for subsequent presentation to the participant via screen share.

Tests that were retained for remote testing (ie, tests administered to both the remote and face-to-face patient cohorts) are itemised in tables 2 and 3. Where applicable, we sought permission from the test publishers to adapt tests for remote administration.

Testing procedure: remote

An initial session was conducted via Zoom to accustom participants to the remote testing format, check the screen and sound sharing options on Zoom, and that the quality of their internet connection was acceptable. Technical aspects of the set-up for remote testing are detailed in online supplemental text and online supplemental figure 1.

The remote neuropsychological and neurolinguistic tests each took around an hour to administer. To minimise fatigue,[17] tests were delivered in separate testing sessions typically within a week (and never more than 2 weeks apart).

Feedback on remote testing experience

At the end of each remote testing session, the experimenter debriefed each participant. This provided them with the opportunity to raise any technical issues, give their impressions of the remote testing session and note any distractions that may have occurred for them.

Where time allowed, at the end of the session, participants were also asked by the experimenter to indicate on a 10-point integer scale how comfortable they had felt with the remote testing format, with 10 indicating 'very comfortable'.

Statistical analysis

All statistical analyses were performed in JASP (V.0.16).

The remote and face-to-face patient cohorts were compared on demographic characteristics using independent samples t-tests and Wilcoxon rank-sum tests. Healthy controls' scores in remote and face-to-face testing environments were compared using paired samples t-tests or (where the assumption of normality was not met) Wilcoxon signed-rank tests. Healthy controls' and patients' ratings of comfort after the general neuropsychological and neurolinguistic sessions with the remote testing set-up were compared using Wilcoxon rank-sum tests. To reduce type I error, no corrections for multiple comparisons were applied.

We did not perform between-group comparisons of neuropsychological and neurolinguistic performance as these syndromic profiles of the neuropsychological and neurolinguistic tests have been reviewed and published previously.[18 19]

In comparing testing environments, our null hypothesis was that there would be no effect of testing environment on neuropsychological performance—that is, no differences in performance between remote and face-to-face assessment settings—for any participant group. To critically assess the magnitude of evidence in favour of this null hypothesis versus the alternative hypothesis (ie, that there was in fact an effect of testing environment) particularly in light of the relatively small patient cohorts here, we employed a Bayesian approach.[20] Bayesian independent samples t-tests (and non-parametric equivalents where assumptions of the general linear model were violated) were performed for each general neuropsychological

Table 1 General demographic, clinical and environmental characteristics for all participant groups: comparison of remote and face-to-face cohort characteristics

	CTL	AD		bvFTD		SD		PNFA		LPA	
Characteristic	**Remote**	**F2F**	**Remote**	**F2F**	**Remote**	**F2F**	**Remote**	**F2F**	**Remote**	**F2F**	**Remote**
N	10	25	8	12	3	9	4	12	5	6	5
Gender (M/F)	7/3	15/10	5/3	8/4	3/0	7/2	2/2	6/6	2/3	4/2	4/1
Age (years)	74.0 (4.1)*	69.5 (4.1)	69.8 (5.9)	70.3 (2.9)	72.0 (5.6)	60.3 (5.2)	58.3 (9.3)	67.1 (5.2)	68.2 (6.4)	69.5 (3.6)	71.6 (4.7)
Education (years)	17.6 (0.7)	14.8 (2.5)	16.0 (3.7)	14.9 (2.7)	14.7 (3.1)	16.0 (1.9)	15.5 (3.1)	15.3 (2.3)	14.8 (2.8)	16.0 (1.7)	16.0 (3.3)
Handedness (R/L/A)	9/1/0	23/1/1	8/0/0	11/1/0	3/0/0	9/0/0	3/1/0	10/2/0	5/0/0	5/1/0	5/0/0
Symptom duration (years)	NA	6.8 (2.3)	7.3 (3.7)	4.7 (2.1)	3.7 (3.8)	4.1 (1.9)	3.5 (1.9)	3.4 (1.7)	3.2 (0.8)	4.2 (1.9)	3.8 (1.9)
Desktop/laptop/tablet	4/5/1	NA	3/2/3	NA	1/2/0	NA	1/3/0	NA	1/2/2	NA	1/3/1
Study partner present (Y/N)	NA	NA	5/3	NA	2/1	NA	2/2	NA	2/3	NA	3/2
Interruptions (N)	1	NA	1	NA	1	NA	0	NA	0	NA	0

Mean (SD) values are shown for each group. No significant differences between remote and face-to-face cohorts were found.
*On average, healthy controls were 3.5 years younger when tested face-to-face.
A, ambidextrous AD, patient group with typical Alzheimer's disease; bvFTD, patient group with behavioural variant frontotemporal dementia; CTL, healthy control group; F, female; F2F, face-to-face; L, left; LPA, patient group with logopenic progressive aphasia; M, male; N, number of participants per group; NA, not applicable; PNFA, patient group with progressive non-fluent aphasia; R, right; SD, patient group with semantic dementia.

Table 2 General neuropsychological performance for all participant groups: comparison of remote and face-to-face test administration

	CTL		All patients		AD		bvFTD		SD		PNFA		LPA	
	F2F	Remote	F2F	Remote	F2F	Remote	F2F	Remote	F2F	Remote	F2F	Remote	F2F	Remote
Number tested	10	10	64	25	25	8	12	3	9	4	12	5	6	5
General intellect														
WASI matrix (/32)	26.9 (2.4)	26.3 (3.5)	**16.5 (8.86)**	**17.7 (9.24)**	**14.5 (7.5)**	**13.7 (11.5)**[b]	14.6 (9.5)[a]	7.7 (2.5)	25.8 (4.9)	23.3 (3.4)	18.6 (8.3)[a]	24.0 (2.6)	10.0 (9.2)	17.6 (10.3)
Episodic memory														
RMT faces (short) (/25)	NA	23.3 (2.5)	**18.9 (3.68)**	**19.2 (4.21)**	18.9 (3.7)[h]	15.3 (2.1)	NA	20.3 (4.0)	NA	19.3 (3.7)	NA	22.4 (3.7)	NA	21.6 (3.8)
RMT faces (/50)	42.7 (4.7)	NA	**32.6 (7.40)**	**NA**	30.44 (6.56)[g]	NA	28.56 (6.29)[c]	NA	31.5 (4.47)[a]	NA	40.3 (5.79)[b]	NA	33.17 (8.93)	NA
Working memory														
DS (reverse) (/12)	**8.0 (2.6)**	**8.2 (2.2)**	**4.82 (2.47)**	**4.64 (2.40)**	4.6 (1.9)[c]	4.0 (1.9)	5.0 (3.1)[a]	4.0 (0.0)	6.9 (2.7)	5.5 (2.7)	4.4 (1.7)[d]	5.2 (4.2)	2.4 (1.1)[a]	4.8 (1.8)
Short-term verbal memory														
DS (forward) (/12)	9.8 (2.0)	9.2 (2.0)	**6.68 (2.94)**	**6.24 (2.52)**	**6.5 (2.1)**[a]	**6.6 (2.3)**	7.8 (2.9)	6.3 (0.6)	9.4 (2.2)	7.3 (0.5)	5.1 (2.6)[b]	6.0 (4.4)	3.0 (3.1)[a]	5.0 (2.5)
Language														
BPVS (/150)	**149.0 (1.1)**	**148.9 (1.1)**	**120 (39.3)**	**134 (26.4)**	**135.1 (24.0)**[a]	**133.9 (21.4)**	118.7 (28.9)[a]	148.0 (1.0)	61.9 (44.1)	98.0 (46.3)	136.3 (30.4)	142.8 (8.4)	120.2 (39.6)	145.2 (4.4)
GNT (/30)	27.3 (2.1)	26.8 (1.8)	**11 (9.33)**	**13 (8.62)**	**13.7 (8.5)**[a]	**12.4 (6.3)**	11.5 (11.0)[a]	20.0 (4.4)	0.2 (0.4)	0.5 (1.0)	15.8 (6.8)	20.2 (7.6)	6.2 (7.7)	12.4 (7.0)
NART (/50)	45.2 (3.4)	44.6 (3.0)	26.1 (13.4)	31.4 (13.1)	28.3 (11.9)[a]	37.1 (4.6)	30.7 (13.1)	40.0 (2.0)	16.1 (10.2)	16.3 (13.6)	28.3 (15.0)[f]	30.8 (14.9)	20.4 (18.5)[a]	29.8 (16.9)
Category fluency (60 s, 'animals')	**25.0 (6.6)**	**24.7 (5.4)**	8.46 (5.75)	12.2 (8.10)	**10.0 (5.6)**	**10.4 (6.7)**	8.0 (6.1)	10.3 (6.0)	6.1 (5.3)[a]	9.8 (6.2)	8.6 (5.8)[d]	18.6 (11.5)	5.7 (5.8)	11.6 (8.1)
Arithmetic														
GDA total (/24)	**16.6 (6.4)**	**16.1 (4.0)**	**6.90 (6.25)**	**6.30 (5.60)**	**4.9 (5.0)**[e]	**4.8 (5.4)**	8.2 (7.4)[a]	8.3 (7.2)	13.3 (5.5)	7.3 (4.0)	5 (3.6)[d]	6.4 (7.2)	1.7 (1.5)	7 (5.8)
Visuospatial														
VOSP OD (/20)	*19.5 (0.7)*	*18.4 (1.2)*	**15.6 (3.77)**	**15.2 (3.52)**	*16.1 (2.6)*[a]	*13.0 (3.0)*	14.0 (4.2)	17.0 (1.0)	15.3 (6.3)[a]	14.5 (5.9)	16.8 (3.6)[b]	17.6 (2.1)	15.8 (3.1)	15.8 (2.7)
Executive														
Letter fluency (60 s, 'F')	**19.7 (6.0)**	**20.8 (4.2)**	7.70 (5.27)	11.3 (6.29)	10.3 (5.2)	11.9 (7.4)	6.8 (3.9)[a]	7.3 (2.3)	7.0 (3.6)[b]	10.8 (7.0)	4.8 (5.8)[d]	14.0 (8.2)[a]	*2.6 (3.8)*[a]	*11.3 (3.9)*

Mean (SD) values of performance on neuropsychological tests are shown (maximum scores are indicated in parentheses) for each testing environment (face-to-face vs remote research setting) for each patient group. **Bold** indicates values for which there was substantial evidence in support of the null hypothesis (H0); *italics* indicate values for which there was substantial evidence in support of the alternative hypothesis (H1). Exact values for the Bayes factor comparing H0 against H1 (BF_{01}) are presented in online supplemental table 2. The general neuropsychological tests comprised tests of general intellect (WASI) matrix reasoning,[39] episodic memory (RMT) for faces,[40] working memory (DS forwards[41]), language (GNT,[42] BPVS[43] and NART[44]), arithmetic (GDA),[45] visual perception (VOSP OD task[46]) and executive function (DS reverse,[41] letter ('F') and category fluency ('animal') tasks[47]). A reduced number of participants completed certain tests, as follows: [a]n-1, [b]n-2, [c]n-3, [d]n-4, [e]n-5, [f]n-6, [g]n-9, [h]n-16.

All patients=combined patient cohort.

AD, patient group with typical Alzheimer's disease; BPVS, British Picture Vocabulary Scale; bvFTD, patient group with behavioural variant frontotemporal dementia CTL, healthy control group; DS, digit span; F2F, face-to-face; GDA, Graded Difficulty Arithmetic test; GNT, Graded Naming Test; LPA, patient group with logopenic progressive aphasia; NA, not available; NART, National Adult Reading Test; OD, object decision; PNFA, patient group with progressive non-fluent aphasia; RMT, Recognition Memory Test; SD, patient group with semantic dementia; VOSP, Visual Object and Space Perception; WASI, Wechsler Adult Intelligence Scale.

Table 3 Neurolinguistic performance for all participant groups: comparison of face-to-face and remote test administration

	CTL		All patients		SD		PNFA		LPA	
Test	**F2F**	**Remote**	**F2F**	**Remote**	**F2F**	**Remote**	**F2F**	**Remote**	**F2F**	**Remote**
Number tested	10	10	27	14	9	4	12	5	6	5
Phoneme perception										
PALPA-3 (/36)	34.9 (1.6)	33.8 (2.0)	33.5 (4.28)	33.1 (2.85)	35.3 (1.3)	33.5 (2.6)	31.9 (5.7)	34.0 (1.6)	33.8 (2.4)[a]	32.0 (4.0)
Reading										
Non-word reading (/25)	24.4 (0.8)	24.9 (0.3)	**18.1 (5.51)**	**16.1 (8.58)**	19.6 (4.9)	16.0 (9.8)	19.3 (5.7)[d]	15.8 (8.7)	14.5 (5.4)	16.4 (9.6)
Regular reading (/25)	25.0 (0.0)	25.0 (0.0)	23.3 (2.83)	21.8 (5.07)	22.8 (3.9)[d]	21.3 (6.2)	23.3 (2.2)[g]	22.2 (4.2)	24.5 (0.7)[d]	21.8 (6.1)
Irregular reading (/25)	24.7 (0.7)	25.0 (0.0)	19.6 (5.30)	21 (5.39)	18.8 (5.8)[d]	18.3 (8.9)	19.8 (6.4)[g]	21.6 (4.4)	21.5 (3.5)[d]	22.6 (2.1)
Naming										
BNT (/30)	**29.2 (0.8)**	**29.0 (1.0)**	13.4 (10.9)	17.4 (9.58)	3.3 (3.4)	9.5 (7.1)	24.7 (4.4)[b]	24.4 (7.6)	9.7 (8.0)	16.8 (8.8)
Semantic association										
Camel and cactus (/32)	**31.0 (1.0)**	**30.9 (1.1)**	25.4 (5.58)	28.4 (2.71)	22.3 (5.1)[e]	25.5 (2.6)	28.3 (5.5)[h]	30.0 (1.4)	29.0 (0.0)[e]	29.0 (2.1)
Word comprehension										
Concrete synonyms (/25)	24.5 (0.5)	24.3 (0.5)	**18.7 (5.99)**	**20.2 (4.00)**	15.5 (2.9)[c]	17.0 (4.0)[a]	22.2 (3.2)[b]	21.0 (4.2)	15.6 (9.4)[a]	21.4 (3.5)
Abstract synonyms (/25)	24.8 (0.4)	24.3 (0.8)	18.5 (6.20)	20.6 (3.12)	14.4 (1.5)[d]	17.3 (1.2)[a]	21.4 (3.7)[b]	21.2 (3.4)	16.5 (11.2)[b]	22.3 (2.1)[a]
Sentence comprehension										
PALPA-55 (/24)	23.8 (0.4)	23.1 (1.3)	**20.5 (3.83)**	**20.9 (2.28)**	22.6 (1.6)[a]	21.8 (2.2)	19.6 (4.8)[a]	21.0 (2.6)	18.8 (2.6)[a]	20.0 (2.1)
Speech repetition										
Monosyllabic word repetition (/15)	*14.6 (0.5)*	*12.7 (1.6)*	13.8 (1.92)	10.6 (3.37)	14.8 (0.4)	10.8 (3.2)	12.8 (2.6)[c]	10.0 (4.2)	14.0 (1.2)[b]	11.2 (3.3)
Bisyllabic word repetition (/15)	**14.7 (0.5)**	**14.7 (0.7)**	**13.5 (2.65)**	**13.4 (2.95)**	14.8 (0.7)	13.0 (1.6)	12.8 (3.3)[c]	12.4 (4.7)	12.0 (2.9)[b]	14.8 (0.4)
Trisyllabic word repetition (/15)	15.0 (0.0)	15.0 (0.0)	**12.8 (4.18)**	**12.9 (3.89)**	14.8 (0.4)	12.8 (3.2)	11.1 (5.9)[c]	12.0 (6.2)	12.3 (2.9)[b]	14.0 (1.0)
Graded difficulty sentence repetition (/10)	9.8 (0.4)	9.3 (0.9)	6.37 (2.89)	5.79 (2.36)	8.3 (1.2)	7.3 (1.7)	4.6 (3.0)[e]	5.2 (3.2)	4.7 (2.9)[c]	5.2 (1.6)
Sentence construction										
Spoken sentences (/25)	25.0 (0.0)	24.9 (0.3)	18.6 (6.70)	22.6 (3.34)	19.1 (7.5)	23.5 (1.3)	18.0 (7.4)[d]	21.6 (5.1)	18.0 (5.3)[c]	22.8 (2.7)

Mean (SD) values of performance on neurolinguistic tests are shown (maximum scores are indicated in parentheses) for each testing environment (face-to-face vs remote) for each participant group. **Bold** indicates values for which there was substantial evidence in support of the null hypothesis (H0); *italics* indicate values for which there was substantial evidence in support of the alternative hypothesis (H1). Exact values for the Bayes factor comparing H0 against H1 (BF_{01}) are presented in online supplemental table 3. The neurolinguistic tests comprised tests of phoneme perception (a shortened version of the PALPA Test 3–minimal pair discrimination[48]), reading (graded non-word reading,[49] and graded tests of regular word and irregular word reading adapted from Coltheart *et al*[50]), confrontation naming (a subset of items from the BNT[51]), semantic association (modified camel and cactus test[52]), single-word comprehension (concrete and abstract synonyms[53]), sentence comprehension (a shortened version of the PALPA subtest 55[48]), speech repetition (tests of monosyllabic, bisyllabic and trisyllabic single-word repetition[54]), graded difficulty sentence repetition using a subset of items from the sentence repetition test[55] and sentence construction based on Rohrer *et al*.[56] We opted to include two naming tasks as the GNT is part of the core neuropsychological tests at our centre and therefore administered to patients with all diagnoses. However, as patients with SD often score at floor on this task,[57] we also include a shortened version of the BNT in our neurolinguistic test that is administered to patients with PPA as patients with SD are more likely to be above floor on this task.[26] A reduced number of participants completed certain tests: [a]n-1, [b]n-2, [c]n-3, [d]n-4, [e]n-5, [f]n-6, [g]n-8, [h]n-9. In the spoken sentence construction test, participants are given five different words (walked, radio, throw, green, tree) and are asked to produce a sentence that contains each word in turn. Sentences are then scored based on grammar and semantic sense, each sentence counting for a maximum of 5 points.
All patients=combined patient cohort.
BNT, Boston Naming Test; CTL, healthy control group F2F, face-to-face; GNT, Graded Naming Test; LPA, patient group with logopenic progressive aphasia; N, number of participants per group; PALPA, Psycholinguistic Assessments of Language Processing in Aphasia; PNFA, patient group with progressive non-fluent aphasia; PPA, primary progressive aphasia; SD, patient group with semantic dementia.

and neurolinguistic test in each patient group separately. As numbers in some groups were quite small, we also conducted analyses for a combined patient cohort in both environments. Healthy control performance was compared using Bayesian paired samples t-tests (or appropriate non-parametric equivalent). A Bayes factor, which is the ratio of evidence supporting the null hypothesis over the alternative hypothesis (hereafter BF_{01}), was calculated for each comparison using JASP. A BF_{01} value >3 indicates substantial evidence in favour of the null hypothesis while a value <0.33 supports the alternative hypothesis; BF_{01} values between 0.33 and 3 are classified as 'anecdotal' evidence, comparable with non-significant differences in inferential statistics.[21 22] Bayes factor values are presented in online supplemental tables 2 and 3.

In comparing groups on comfort ratings after the remote sessions, our null hypothesis was that there would be no differences in comfort ratings between healthy controls and patients; our alternative hypothesis was that healthy control participants would report higher comfort ratings than patients.

Finally, we conducted F-tests and Levene's equality of variance tests to evaluate differences in variability between the two testing environments.

Patient and public involvement

In August and September 2020, we contacted 527 people (comprising healthy control participants and people with a diagnosis of a dementia) who had previously taken part in our face-to-face research programmes in the Dementia Research Centre, University College London, or who had expressed an interest in doing so in the future. They were asked, 'Would you consider participating in research remotely (telephone/online)?' Of the 163 people who answered the question, 145 (89%) indicated that they would be happy to take part in remote research. Based on this feedback, we submitted an amendment to our existing research ethics that was approved in October 2020. Following this, we conducted a pilot remote testing session with an older healthy control individual who was also a carer for a family member living with dementia. Their feedback was instrumental in developing and improving our remote testing procedure.

Results from this work will be disseminated to members of the support groups that we run with Rare Dementia Support (www.raredementiasupport.org) through online presentations at webinars and research summaries in newsletters.

RESULTS

General characteristics of participant groups are presented in table 1; performance on the general neuropsychological tests is given in table 2; performance on the neurolinguistic tests is shown in table 3. Figures 1 and 2 show radar plots of performance for each participant group for the general neuropsychological and neurolinguistic tests, respectively. Figures 3 and 4 show performance profiles of healthy control participants on the general neuropsychological and neurolinguistic tests, respectively. online supplemental figure 2 and 3 show performance profiles of the combined patient cohort on the general neuropsychological and neurolinguistic tests, respectively. Bayesian statistics are presented in online supplemental tables 2 and 3; equality of variance analyses is presented in online supplemental table 4; and results for the audibility screening task (see online supplemental methods for more information) are presented in online supplemental table 5.

General participant characteristics

Of the 87 potential participants who were contacted, 35 (40.2%) ultimately took part in the research (25 patients, 10 healthy controls who had taken part in face-to-face research previously). Reasons for declining participation and reports of any technical issues with remote test delivery are detailed in the online supplemental material.

There were no significant differences in age, years of education or symptom duration between the face-to-face and remote testing patient cohorts (table 1).

Below we highlight comparisons where there was substantial evidence in support of either the null (ie, no difference between remote and face-to-face performance) or alternative (ie, difference between remote and face-to-face performance) hypothesis. Comparisons are shown in full in online supplemental tables 2 and 3.

General neuropsychological assessment

Overall, there was little evidence for a significant effect of assessment environment on general neuropsychological test performance in any participant group.

Healthy individuals scored equally well on the digit span reverse, the British Picture Vocabulary Scale (BPVS), the Graded Difficulty Arithmetic test (GDA), and on both letter and category fluency tests (all BF_{01} >3 indicating substantial evidence in favour of the null hypothesis). However, they performed less well on the Visual Object and Spatial Perception object decision task (VOSP) (BF_{01}=0.0404, indicating substantial evidence in favour of the alternative hypothesis) in remote testing than in face-to-face testing, with the remote group (mean=18.4) performing worse than the face-to-face group (mean=19.5) (figures 1 and 3, table 2 and online supplemental table 2).

For the comparisons of the combined remote versus combined face-to-face patient cohorts, there was substantial evidence supporting the null hypothesis for all neuropsychological tests (all BF_{01} >3), except for the National Adult Reading Test and both letter and category fluency tests, where evidence in support of the null hypothesis was anecdotal (table 2 and online supplemental table 2).

For individual patient groups (figure 1, table 2 and online supplemental table 2), there was substantial evidence to suggest that the remote AD cohort performed similarly to the face-to-face AD cohort on Wechsler Adult Intelligence Scale (WASI) matrix reasoning, digit span

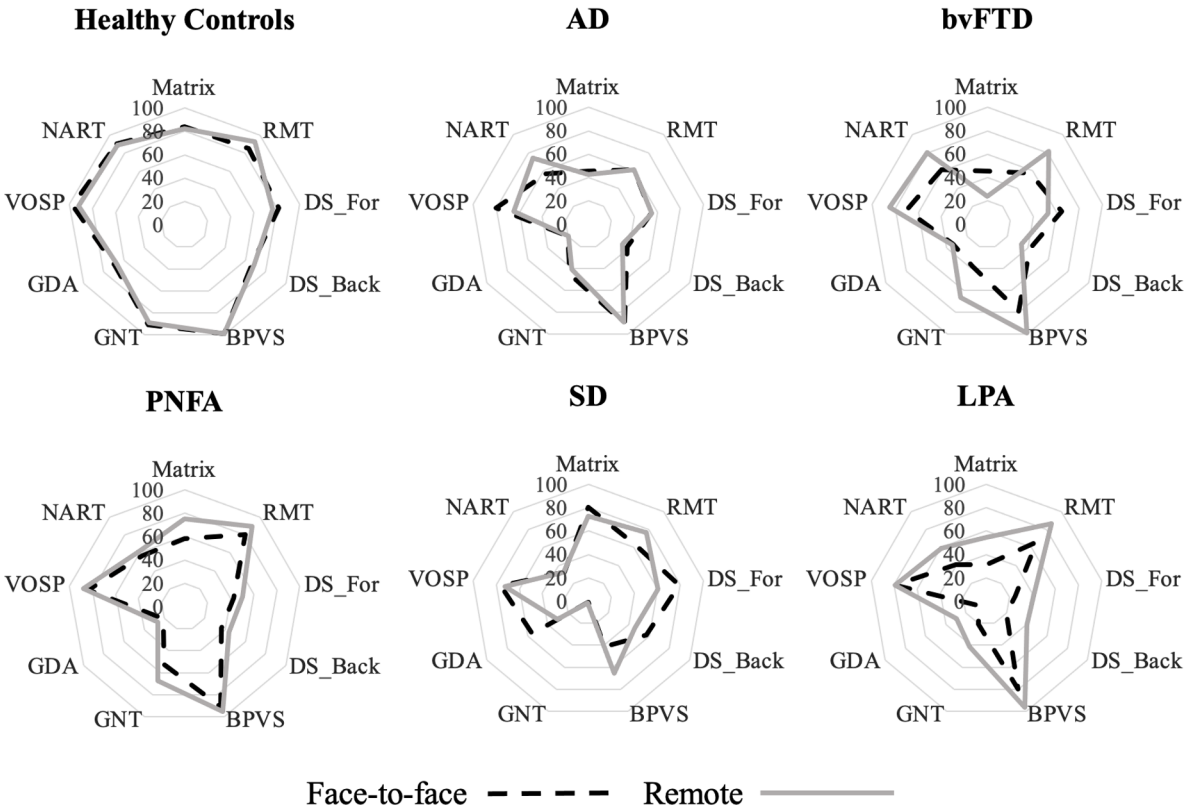


Figure 1 Radar plots of general neuropsychological test performance, by participant group and testing environment. Average percentage correct score (plotted on concentric lines) was calculated for each participant group for each test in the neuropsychological tests, across each testing environment. Scores for the fluency tasks were not included here as responses on these tasks cannot be evaluated as correct/incorrect. AD, patient group with typical Alzheimer's disease; BPVS, British Picture Vocabulary Scale; bvFTD, patient group with behavioural variant frontotemporal dementia; DS_For/Back, digit span forwards/backwards; GDA, Graded Difficulty Arithmetic test; GNT, Graded Naming Test; LPA, patient group with logopenic progressive aphasia; Matrix, WASI matrix reasoning; NART, National Adult Reading Test; PNFA, patient group with progressive non-fluent aphasia; RMT, Recognition Memory Test; SD, patient group with semantic dementia; VOSP, Visual Object Space Perception; WASI, Wechsler Adult Intelligence Scale.

forwards, Graded Naming Test (GNT), BPVS, GDA and category fluency test (all BF_{01} >3). However, the remote AD cohort performed less well on the VOSP (mean=13.0; BF_{01}=0.171, substantial evidence) compared with the face-to-face cohort (mean=16.1). Conversely, patients with LPA who completed the letter fluency test remotely (mean words=11.3) performed better than those who completed the same task face-to-face (mean=2.6; BF_{01}=0.188, substantial evidence).

No other comparisons yielded substantial evidence in support of either hypothesis (online supplemental table 2).

Results of the equality of variance analyses are reported in full in online supplemental table 4. The assumption of homogeneity of variance was not violated for any test in the general neuropsychological tests in either the healthy control or combined patient cohorts.

Neurolinguistic assessment

Overall, there was little evidence for a significant effect of assessment environment on neurolinguistic test performance in any participant group.

Healthy individuals scored equally well on the Boston Naming Test (BNT), the camel and cactus test and the bisyllabic single-word repetition test (all BF_{01} >3, indicating substantial evidence in favour of the null hypothesis). However, they performed less well on the monosyllabic word repetition test (BF_{01}=0.0487, substantial evidence in favour of the alternative hypothesis) in remote testing than in face-to-face testing, with the remote group (mean=12.7) performing worse than the face-to-face group (mean=14.6) (figures 2 and 4, table 3 and online supplemental table 3).

The comparisons between combined patient cohorts for remote versus face-to-face testing showed substantial evidence supporting the null hypothesis for non-word reading, concrete synonyms, the Psycholinguistic Assessments of Language Processing in Aphasia-55, and bisyllabic and trisyllabic single-word repetition tests (all BF_{01} values >3). There was anecdotal evidence supporting the null hypothesis on all other neurolinguistic tests (all BF_{01} values between 1 and 3).

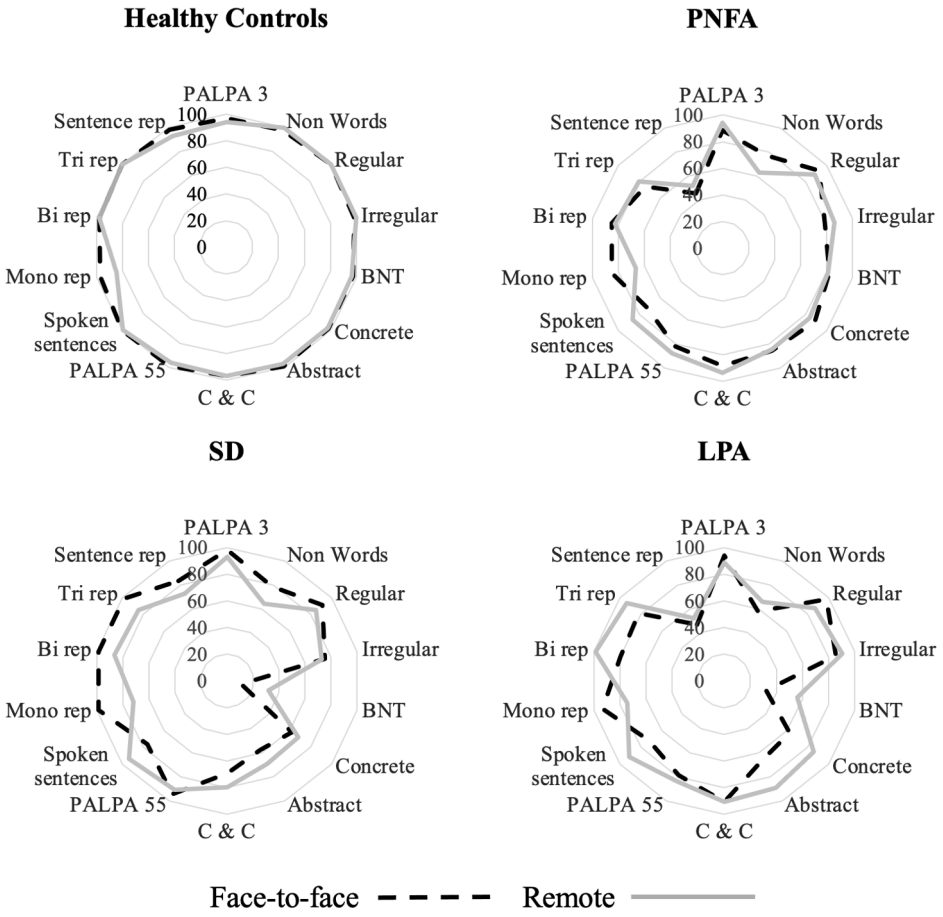


Figure 2 Radar plots of performance on neurolinguistic test performance, by participant group and testing environment. Average percentage score (plotted on concentric lines) was calculated for each participant group for each test in the neurolinguistic tests, across each testing environment. Abstract, abstract synonyms test; bi rep, bisyllabic single-word repetition; BNT, Boston Naming Test; C & C, camel and cactus test; concrete, concrete synonyms test; irregular, irregular word reading test; LPA, patient group with logopenic progressive aphasia; mono rep, monosyllabic single-word repetition test; non word, non-word reading test; PALPA, Psycholinguistic Assessment of Language Processing in Aphasia subtests; PNFA, patient group with progressive non-fluent aphasia; regular, regular word reading test; SD, patient group with semantic dementia; sentence rep, graded difficulty sentence repetition test; spoken sentences, spoken sentences test; tri rep, trisyllabic single-word repetition test.

Individual patient group comparisons across environments did not yield substantial evidence in support of either hypothesis.

Results of the equality of variance analyses are reported in full in online supplemental table 4. For the healthy control group, the assumption of equality of variance was violated for monosyllabic repetition ($F=0.11$, $p<0.05$, with higher variability in the remote condition); for the combined patient cohort, this assumption was violated for monosyllabic repetition ($F=4.89$, $p=0.03$, with higher variability in the remote group), the camel and cactus test ($F=4.25$, $p=0.02$, with higher variability in the face-to-face group) and the spoken sentence task ($F=4.37$, $p<0.05$, with higher variability in the face-to-face group).

Feedback on remote testing experience

We received 20 responses to the question, 'How comfortable did you feel in this new setting (eg, online testing)?' after the general neuropsychology remote session (from 10 healthy control participants and 10 patients); and 22 responses to the same question that was posed after the neurolinguistic remote session (from 10 healthy control participants and 12 patients). There was little evidence for a significant difference across groups on these ratings for either the general neuropsychology session (control mean=9.60, standard deviation (st.d)=0.70; patient mean=9.10, st.d=1.91; $BF_{01}=1.825$, anecdotal evidence supporting the null hypothesis) or neurolinguistic session (control mean=9.10, st.d=1.73; patient mean=8.92, st.d=1.73; $BF_{01}=2.325$, anecdotal evidence supporting the null hypothesis).

DISCUSSION

The present findings suggest that administration of neuropsychological tasks remotely over the internet with healthy older adults and people with a diverse range of dementia phenotypes is feasible according to three key metrics: acceptability, adaptation and demand.[11] In terms

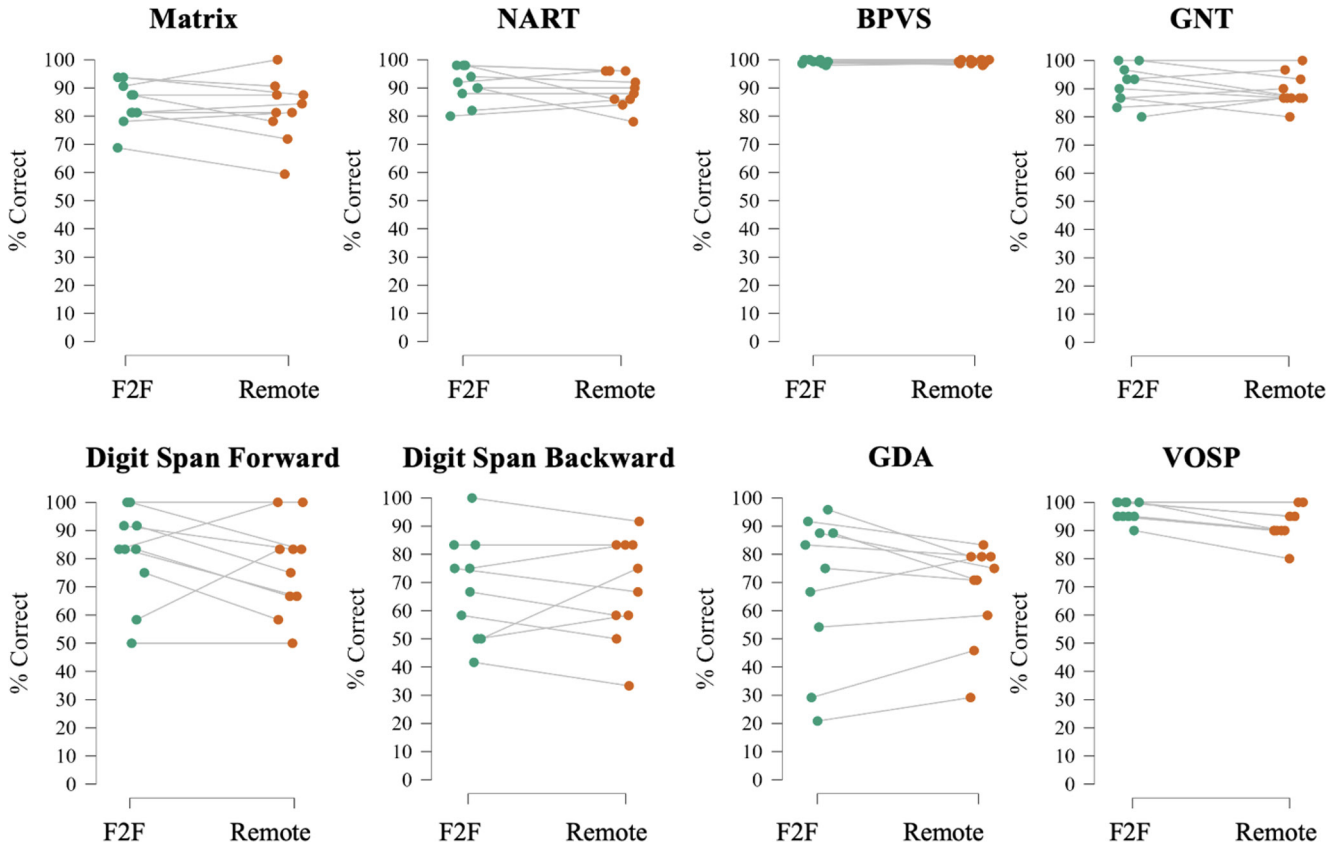


Figure 3 Performance profiles of healthy control participants on tasks in general neuropsychological tests. Line plots showing performance profiles of individual healthy control participants on tasks in the general neuropsychological tests. BPVS, British Picture Vocabulary Scale; F2F, face-to-face; GDA, Graded Difficulty Arithmetic test; GNT, Graded Naming Test; Matrix, WASI matrix reasoning; NART, National Adult Reading Test; VOSP, Visual Object Space Perception object decision task; WASI, Wechsler Adult Intelligence Scale.

of acceptability, results from our feedback questionnaires indicated that patient and healthy control participants were comfortable with the remote testing environment. For adaptation, we have demonstrated that similar performance outcomes were obtained across test settings (with further discussion of these findings below). In terms of demand, only 6.9% of those we contacted about taking part in the research declined due to technological reasons (online supplemental material 1).

Our Bayesian analytical approach demonstrated that there was anecdotal or substantial evidence suggesting comparable performance across testing environments of healthy participants and patients with AD on a range of general neuropsychological and neurolinguistic tests, specifically those targeting working memory (digit span forward), executive functioning (digit span reverse, letter and category fluency tests, WASI matrix reasoning), arithmetic skills (GDA) and general semantic knowledge (BNT, BPVS, GNT). These results corroborate previous reports of preserved neuropsychological performance on executive function, working memory, and language tests across testing environments in both healthy individuals and patients with AD.[10] Our findings also corroborated previous work suggesting that remote assessments are viable for people with PPA.[23]

Healthy controls and participants with AD both performed significantly worse on the remote version of the VOSP object decision task, in which participants are presented with four silhouettes and asked to select the drawing of a real object; the three distractor silhouettes are based on nonsense shapes. The typical amnestic AD phenotype can include prominent visuospatial impairments[24 25] and it is feasible that a reduction in stimulus quality may have stressed cortical apperceptive mechanisms still further, akin to a dynamic 'stress test' of degraded input processing.[26–28] However, it is worth noting that there was no such discrepancy across the AD cohorts for other tasks involving visual administration (eg, WASI matrix). For the healthy controls, the absolute performance difference across environments was relatively small (mean reduction of 1.1 points) for the VOSP; however, we note that even small differences can have important consequences if this change were to yield a lower scaled/percentile score and thus affect interpretations of test performance. It is also possible that this reduction at least in part reflected normal healthy ageing, consistent with previous findings,[29] as the healthy control cohort was tested on the remote test 3–4 years after their face-to-face assessment.

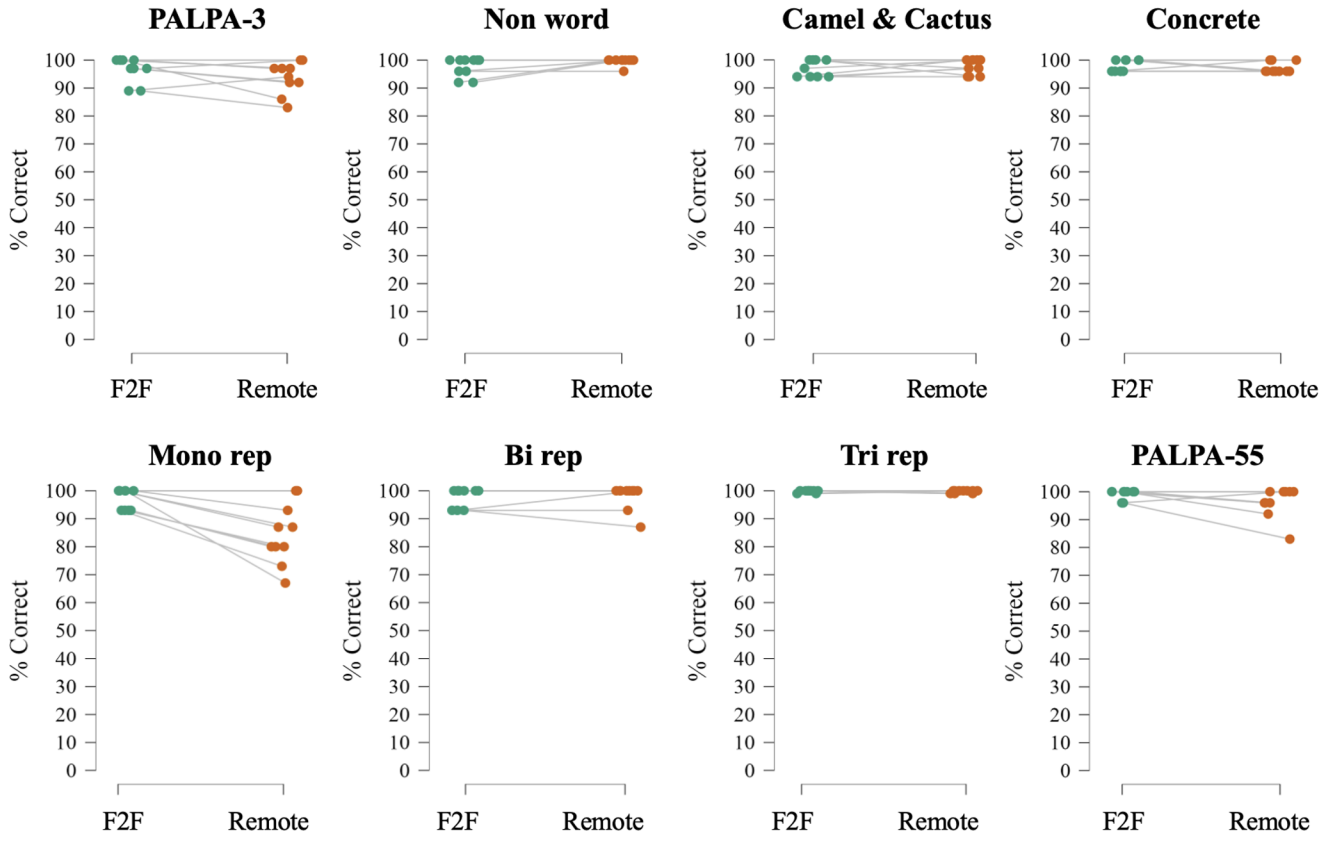


Figure 4 Performance profiles of healthy control participants on tasks in the neurolinguistic tests. Line plots indicating percentage scores for each healthy control on representative tests from the neurolinguistic tests administered face-to-face (F2F) and remotely. Scores on the trisyllabic single-word repetition task were jittered slightly on the x-axis to allow for plotting as participants were uniformly at ceiling in both environments. Bi rep, bisyllabic single-word repetition; Concrete, concrete synonyms test; Mono rep, monosyllabic single-word repetition test; Non word, non-word reading test; PALPA, Psycholinguistic Assessment of Language Processing in Aphasia; Tri rep, trisyllabic single-word repetition.

Healthy control participants also performed significantly worse on the monosyllabic single-word repetition task when delivered remotely, and there was more variability on this task when performed remotely. The videoconferencing software may have degraded the fidelity of the raw speech signal,[30] essentially resulting in a harder task than when administered face-to-face. This is potentially consistent with the controls' preserved performance on the bisyllabic single-word repetition test where top-down information can be used to complement bottom-up auditory information partially degraded by the videoconferencing software.[31] An alternative (or complementary) explanation could again be age-related changes, here affecting hearing function (presbycusis).[32–34]

The finding of significantly better performance on the verbal (letter) fluency task in the LPA cohort tested remotely compared with the cohort tested face-to-face is surprising. The obvious explanation is that the remote cohort was overall less impaired than the patients seen face-to-face; although efforts were made to match the two cohorts for disease severity and other potentially relevant factors. An alternative explanation could be that participants found the remote setting less anxiety provoking than face-to-face testing in an unfamiliar environment. Patients with LPA may be relatively susceptible to anxiety as a factor modulating cognitive performance.[35 36] Additionally, as word retrieval is an intrinsically dynamic process that is likely to be facilitated by the availability of 'prompts',[37] patients may have benefitted from cueing of word retrieval by their more familiar home environments.

The way the neuropsychological and neurolinguistic testing protocol was adapted for remote delivery may have favoured null differences. The testing sessions were shorter and spread out within a week, which may have helped counteract the effect of anxiety related to the unfamiliarity of the remote testing setting, as well as potential 'Zoom' fatigue.[17] The increased flexibility of scheduling compared with face-to-face testing in addition to the absence of potential stressors associated with a face-to-face research visit (eg, travelling, being in a unfamiliar environment) may have led to participants feeling more relaxed when taking part remotely versus face-to-face: future research should explore this. Furthermore, certain tests selected for remote delivery may have been intrinsically less susceptible to changes in testing protocol (eg, BPVS), whereas we deliberately excluded tests that we considered would not be practical or suboptimal for remote delivery (eg, WASI block design, Baxter spelling test, Trails). Anecdotally, participants reported satisfaction with the remote testing protocol.

The current study presents several limitations which should inform future work. First, while most statistical comparisons indicated similar performance between testing environments for healthy participants and those with dementia, they were not all supported by substantial evidence and certain comparisons even led to the opposite conclusion. Second, the present study was not ideally designed to compare the two testing environments, as the patient cohorts were different and the healthy control participants were not tested simultaneously in both environments within the same year. This meant that certain statistical measures that may have been informative (eg, assessing stability of ranking within groups) were not appropriate here. While it is likely that the variability of test results in the patient cohorts will be influenced by testing environment, this needs to be interpreted cautiously, due to unequal sample sizes and individual disease trajectories—and may further depend on the particular test employed. Indeed, our sample sizes across modalities were relatively small. These findings would need to be replicated in larger cohorts with the same patients in each test situation, to rule out the possibility of small differences observed in favour of face-to-face testing—and in particular, to assess the extent of individual variability in any differential effect of test environment. Patients of equivalent disease severity would also need to be tested to compare the differential impact of diagnosis on remote performance over the course of the illness. Third, here we did not control for potential deficits in peripheral hearing as these are difficult to measure remotely without adequate equipment. Fourth, we manually adapted face-to-face tasks for remote administration, but there are now several established fully integrated online neuropsychological tests that have shown success in assessing patients with neurodegenerative disease remotely[38]: future research could explore the extent to which our results are comparable with those obtained by such tests. Fifth, only a proportion of patients had time to respond to the additional question asking for their rating of comfort with the remote testing set-up, and it is possible that there was some selection bias here in that those patients who felt most comfortable with the technology finished the sessions earlier and therefore had time to give this additional feedback. Relatedly, while there was little evidence for a significant effect of 'group' (controls vs patients) on these comfort ratings, the patient cohort did record a lower mean average score than controls after both the remote general neuropsychological and neurolinguistic sessions, something that would warrant further and more in-depth investigation with validated assessment tools. Finally, we note that the study was designed to explore the potential for remote neuropsychological assessments of research participants, and the results and conclusions here may not generalise to clinical settings.

Overall, the present findings demonstrate that, despite challenges in setting up remote testing protocols (specifically due to technological requirements), these may produce similar results to face-to-face testing protocols. These are encouraging findings given the current climate and anticipating that research participants may continue to favour remote (or hybrid) visits over face-to-face assessments for reasons of convenience as well as safety, as we move beyond the COVID-19 pandemic.

Author affiliations
[1]Dementia Research Centre, University College London, London, UK
[2]Kidney Cancer Program, UT Southwestern Medical Center, Dallas, Texas, USA
[3]Cognition and Brain Sciences Unit, Cambridge University, Cambridge, UK

Twitter Jessica Jiang @jessj_siping, Jason D Warren @profjasonwarren and Chris JD Hardy @cjdhardy

Acknowledgements We are grateful to all participants for their involvement.

Contributors M-CR-K, JJ, LD, EB, LR, RLB, EVB, CG, SB, JR, SC, JW and CH contributed to the conception and design of the study. M-CR-K, JJ, LD, EB, LR, RLB, EVB, CG and CH contributed to the acquisition of data. M-CR-K, JJ, JW and CH were involved in the analysis and interpretation of data, and wrote the first draft of the article. M-CR-K, JJ, LD, EB, LR, RLB, EVB, CG, SB, JR, SC, JW and CH were involved in revision of the draft and approved the final version of the manuscript. CH is responsible for the overall content as the guarantor. M-CR-K and JJ contributed equally to this paper.

Funding The Dementia Research Centre is supported by Alzheimer's Research UK, Brain Research UK and the Wolfson Foundation. This work was supported by the Alzheimer's Society, the Royal National Institute for Deaf People, Alzheimer's Research UK, the National Institute for Health Research University College London Hospitals Biomedical Research Centre and the University College London Leonard Wolfson Experimental Neurology Centre (grant PR/ylr/18575). M-CR-K was supported by a Wellcome Trust PhD Studentship (102129/B/13/Z). JJ is supported by a Frontotemporal Dementia Research Studentship in Memory of David Blechner (funded through the National Brain Appeal). EB was supported by a Brain Research UK PhD Studentship. RLB was supported by an MRC PhD Studentship in Mental Health. SC was supported by grants from ESRC-NIHR (ES/L001810/1), EPSRC (EP/M006093/1) and Wellcome Trust (200783). CH was supported by a Royal National Institute for Deaf People–Dunhill Medical Trust Pauline Ashley Fellowship (grant PA23_Hardy) and a Wellcome Institutional Strategic Support Fund Award (204841/Z/16/Z). This research was funded in part by UKRI and the Wellcome Trust (grant 204841/Z/16/Z). For the purpose of Open Access, the author has applied a Creative Commons Attribution (CC BY) public copyright licence to any Author Accepted Manuscript version arising from this submission.

Competing interests None declared.

Patient and public involvement Patients and/or the public were involved in the design, or conduct, or reporting, or dissemination plans of this research. Refer to the Methods section for further details.

Patient consent for publication Not required.

Ethics approval This study involves human participants and ethical approval was granted by the University College London and National Hospital for Neurology and Neurosurgery Joint Research Ethics Committees in accordance with the Declaration of Helsinki (reference numbers 06N032 and 150508). Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. The data that support the findings of this study are available on request from the corresponding author.

ORCID iDs
Maï-Carmen Requena-Komuro http://orcid.org/0000-0002-5624-0527
Jonathan D Rohrer http://orcid.org/0000-0002-6155-8417
Jason D Warren http://orcid.org/0000-0002-5405-0826
Chris JD Hardy http://orcid.org/0000-0002-4900-6492

REFERENCES

1 Wang Q, Davis PB, Gurney ME, *et al*. COVID-19 and dementia: analyses of risk, disparity, and outcomes from electronic health records in the US. *Alzheimers Dement* 2021;17:1297–306.
2 Poon P, Hui E, Dai D, *et al*. Cognitive intervention for community-dwelling older persons with memory problems: telemedicine versus face-to-face treatment. *Int J Geriatr Psychiatry* 2005;20:285–6.
3 Jelcic N, Agostini M, Meneghello F, *et al*. Feasibility and efficacy of cognitive telerehabilitation in early Alzheimer's disease: a pilot study. *Clin Interv Aging* 2014;9:1605–11.
4 Costanzo MC, Arcidiacono C, Rodolico A, *et al*. Diagnostic and interventional implications of telemedicine in Alzheimer's disease and mild cognitive impairment: a literature review. *Int J Geriatr Psychiatry* 2020;35:12–28.
5 Dial HR, Hinshelwood HA, Grasso SM, *et al*. Investigating the utility of teletherapy in individuals with primary progressive aphasia. *Clin Interv Aging* 2019;14:453–71.
6 Rogalski EJ, Saxon M, McKenna H, *et al*. Communication bridge: a pilot feasibility study of Internet-based speech-language therapy for individuals with progressive aphasia. *Alzheimers Dement* 2016;2:213–21.
7 Capozzo R, Zoccolella S, Frisullo ME, *et al*. Telemedicine for delivery of care in frontotemporal lobar degeneration during COVID-19 pandemic: results from southern Italy. *J Alzheimers Dis* 2020;76:481–9.
8 Cuffaro L, Di Lorenzo F, Bonavita S, *et al*. Dementia care and COVID-19 pandemic: a necessary digital revolution. *Neurol Sci* 2020;41:1977–9.
9 Waddington CG, Harding E, Brotherhood EV. The development of virtual Videoconference-Based support for people living with rare dementias and their carers: three-phase support group evaluation protocol. *JMIR Research Protocols*. In Press.
10 Hunter MB, Jenkins N, Dolan C, *et al*. Reliability of telephone and Videoconference methods of cognitive assessment in older adults with and without dementia. *J Alzheimers Dis* 2021;81:1625–47.
11 Bowen DJ, Kreuter M, Spring B, *et al*. How we design feasibility studies. *Am J Prev Med* 2009;36:452–7.
12 Newkirk LA, Kim JM, Thompson JM, *et al*. Validation of a 26-point telephone version of the Mini-Mental state examination. *J Geriatr Psychiatry Neurol* 2004;17:81–7.
13 Kennedy RE, Williams CP, Sawyer P, *et al*. Comparison of in-person and telephone administration of the Mini-Mental state examination in the University of Alabama at Birmingham study of aging. *J Am Geriatr Soc* 2014;62:1928–32.
14 Dubois B, Feldman HH, Jacova C, *et al*. Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria. *Lancet Neurol* 2014;13:614–29.
15 Rascovsky K, Hodges JR, Knopman D, *et al*. Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. *Brain* 2011;134:2456–77.
16 Gorno-Tempini ML, Hillis AE, Weintraub S, *et al*. Classification of primary progressive aphasia and its variants. *Neurology* 2011;76:1006–14.
17 Bailenson JN. Nonverbal overload: a theoretical argument for the causes of Zoom fatigue. *Technol Mind Behav* 2021;2.
18 Sivasathiaseelan H, Marshall CR, Agustus JL, *et al*. Frontotemporal dementia: a clinical review. *Semin Neurol* 2019;39:251–63.
19 Marshall CR, Hardy CJD, Volkmer A, *et al*. Primary progressive aphasia: a clinical approach. *J Neurol* 2018;265:1474–90.
20 Dienes Z. Using Bayes to get the most out of non-significant results. *Front Psychol* 2014;5:781.
21 Quintana DS, Williams DR. Bayesian alternatives for common null-hypothesis significance tests in psychiatry: a non-technical guide using JASP. *BMC Psychiatry* 2018;18:178.
22 Ashworth M, Palikara O, Burchell E, *et al*. Online and face-to-face performance on two cognitive tasks in children with Williams syndrome. *Front Psychol* 2020;11:594465.
23 Rao LA, Roberts AC, Schafer R. The reliability of Telepractice administration of the Western aphasia Battery-Revised in persons with primary progressive aphasia. *Am J Speech Lang Pathol* 2022:1–15.
24 Day BL, Ocal D, Peters A, *et al*. Altered visual and haptic verticality perception in posterior cortical atrophy and Alzheimer's disease. *J Physiol* 2022;600:373–91.
25 Mandal PK, Joshi J, Saharan S. Visuospatial perception: an emerging biomarker for Alzheimer's disease. *J Alzheimers Dis* 2012;31 Suppl 3:S117–35.
26 Hardy CJD, Marshall CR, Bond RL, *et al*. Retained capacity for perceptual learning of degraded speech in primary progressive aphasia and Alzheimer's disease. *Alzheimers Res Ther* 2018;10:70.
27 Zarkali A, Adams RA, Psarras S, *et al*. Increased weighting on prior knowledge in Lewy body-associated visual hallucinations. *Brain Commun* 2019;1:fcz007.
28 Weil RS, Pappa K, Schade RN, *et al*. The Cats-and-Dogs test: a tool to identify visuoperceptual deficits in Parkinson's disease. *Mov Disord* 2017;32:1789–90.
29 Herrera-Guzmán I, Peña-Casanova J, Lara JP, *et al*. Influence of age, sex, and education on the visual object and space perception battery (VOSP) in a healthy normal elderly population. *Clin Neuropsychol* 2004;18:385–94.
30 Weerathunge HR, Segina RK, Tracy L, *et al*. Accuracy of acoustic measures of voice via Telepractice videoconferencing platforms. *J Speech Lang Hear Res* 2021;64:2586–99.
31 Jiang J, Benhamou E, Waters S, *et al*. Processing of degraded speech in brain disorders. *Brain Sci* 2021;11:394.
32 Goodwin MV, Hogervorst E, Maidment DW. The impact of presentation modality on cognitive test performance for adults with hearing loss. *Alzheimers Dement* 2021;17 Suppl 12:e058571.
33 Helfer KS, Mamo SK, Clauss M, *et al*. Listening in 2020: a survey of adults' experiences with Pandemic-Related disruptions. *Am J Audiol* 2021;30:941–55.
34 Naylor G, Burke LA, Holman JA. Covid-19 Lockdown affects hearing disability and handicap in diverse ways: a rapid online survey study. *Ear Hear* 2020;41:1442–9.
35 Rohrer JD, Warren JD. Phenomenology and anatomy of abnormal behaviours in primary progressive aphasia. *J Neurol Sci* 2010;293:35–8.
36 Magnin E, Chopard G, Ferreira S, *et al*. Initial neuropsychological profile of a series of 20 patients with logopenic variant of primary progressive aphasia. *J Alzheimers Dis* 2013;36:799–808.
37 Birn RM, Kenworthy L, Case L, *et al*. Neural systems supporting lexical search guided by letter and semantic category cues: a self-paced overt response fMRI study of verbal fluency. *Neuroimage* 2010;49:1099–107.
38 Alosco ML, Mariani ML, Adler CH, *et al*. Developing methods to detect and diagnose chronic traumatic encephalopathy during life: rationale, design, and methodology for the diagnose CTE research project. *Alzheimers Res Ther* 2021;13:136.
39 Wechsler D. *Wechsler adult intelligence Scale-Revised*. New York: Psychological Corporation, 1981.
40 Warrington EK. *Recognition memory test*. Windsor: NFER-Nelson, 1984.
41 Wechsler D. *Wechsler memory scale: revised*. San Antonio, TX: The Psychological Corporation, 1987.
42 McKenna P, Warrington EK. Testing for nominal dysphasia. *J Neurol Neurosurg Psychiatry* 1980;43:781–8.
43 Dunn LM, Whetton C. *British picture vocabulary scale*. Windsor, England: NFER-Nelson, 1997.
44 Nelson HE, Wilson J. *National adult reading test (NART)*. Windsor, UK: Nfer-Nelson, 1991.
45 Jackson M, Warrington EK. Arithmetic skills in patients with unilateral cerebral lesions. *Cortex* 1986;22:611–20.
46 Warrington EK, James M. *The visual object and space perception battery*. Bury St Edmunds, UK: Thames Valley Test Company, 1991.
47 Delis DC, Kaplan E, Kramer JH. *Delis-Kaplan executive function system (D-KEFS)*. San Antonio, TX: Psychological Corporation, 2001.
48 Kay J, Lesser R, Coltheart M. Psycholinguistic assessments of language processing in aphasia 1992.
49 Snowling M, Stothard J, McLean JF. Graded nonword reading test. *Thames Valley Test* 1996.
50 Coltheart M, Besner D, Jonasson JT, *et al*. Phonological encoding in the lexical decision task. *Q J Exp Psychol* 1979;31:489–507.

51 Kaplan E, Goodglass H, Weintraub S. Boston naming test 1983.
52 Moore K, Convery R, Bocchetta M, *et al*. A modified camel and cactus test detects presymptomatic semantic impairment in genetic frontotemporal dementia within the GENFI cohort. *Appl Neuropsychol Adult* 2022;29:112–9.
53 Warrington EK, Mckenna P, Orpwood L. Single word comprehension: a Concrete and Abstract word synonym test. *Neuropsychol Rehabil* 1998;8:143–54.
54 McCarthy R, Warrington EK. A two-route model of speech production. Evidence from aphasia. *Brain* 1984;107:463–85.
55 Strauss E, Sherman EMS, Spreen O. *A compendium of neuropsychological tests: administration, norms, and commentary*. Oxford University Press, 2006.
56 Rohrer JD, Crutch SJ, Warrington EK, *et al*. Progranulin-Associated primary progressive aphasia: a distinct phenotype? *Neuropsychologia* 2010;48:288–97.
57 Hardy CJD, Agustus JL, Marshall CR, *et al*. Behavioural and neuroanatomical correlates of auditory speech analysis in primary progressive aphasias. *Alzheimers Res Ther* 2017;9:53.

Supplementary Material: Remote versus face-to-face neuropsychological testing for dementia research: a comparative study in people with Alzheimer's disease, frontotemporal dementia and healthy older individuals

by MC Requena-Komuro, J Jiang et al

Technical aspects of the set-up for remote testing

Participants were permitted to use their preferred device (tablet, laptop, or desktop computer; see Table 1). To ensure screen visibility, we did not accept the use of smartphones. Most participants (90%) listened via speakers; six participants used headphones, and device volume was set to a comfortable level by each participant or their caregiver. Remote assessments were scheduled to ensure that testing could be completed in a quiet environment with minimal distractions. Additionally, the experimenters asked that each participant's video remained turned on so the experimenter could keep track of any distractions that may be occurring (see Table 1), as well as to ensure that no additional materials were used during the tests (e.g., paper/calculator). Where a task required visual presentation, this was done by screen sharing the Microsoft Powerpoint presentation containing the scanned stimuli for that task in full screen mode. Each patient's primary caregiver was asked to be available during each research session in case of any problems with using the equipment; in practice, no major technological issues arose. The primary caregiver was also permitted to be within the room during the sessions (see Table 1), but were told explicitly not to interfere with any of the neuropsychological and neurolinguistics tests given to the patient participants.

To check basic audibility in the remote testing environment, before each remote session, participants first listened to a set of 10 sentences from the Bamford-Kowal-Bench (BKB) list[1]. These sentences have previously been validated in hearing-impaired children. The spoken sentences were delivered online using an online experiment builder, Labvanced[2]. In each trial, a spoken sentence was played to the participant via screen and sound share on Zoom and the participant was encouraged to select the last word in the sentence they had just heard from three possible options presented visually via screen share (see Figure S1). A perfect score on the final three items was required for the participant to proceed to the remote testing session proper (this allowed each participant and/or caregiver to manually adjust the volume to a comfortable level during the first seven sentences). Most participants (95%) performed at ceiling across all ten items, and no participant made an error on any of the final three items, meaning that none was rejected based on their BKB performance (see Table S5). The order of sentences was fixed across participants.

Reasons for declining participation

Of those who did not take part, 20 (23.0%) did not respond (four AD, two bvFTD, three SD, two PNFA, three LPA, six controls); six (6.9%) declined due to not being comfortable using videoconferencing technology (two LPA, four controls); two (2.3%) declined due to changing health conditions (one AD, one control); one (1.1%) bvFTD patient declined due to sensory difficulties; one (1.1%) LPA patient declined due to family reasons; two (2.3%) declined due to anxieties about research (one AD, one LPA); three (3.4%) declined due to being too impaired

Requena-Komuro M-C, *et al. BMJ Open* 2022; 12:e064576. doi: 10.1136/bmjopen-2022-064576

(two bvFTD, one LPA); two (2.3%) declined due to being too busy with other activities (one SD, one PNFA); 12 (13.8%) provisionally said they were interested but we were unable to schedule a convenient time for them to take part (eight AD, three PNFA, one LPA); two (2.3%) healthy controls expressed an interest in participating but had not previously undertaken face-to-face research so were not included here; and one (1.1%) patient with atypical PPA expressed an interest in participating but was not included here as no other atypical PPA patients were recruited for the remote research.

Technical issues with remote test delivery

Three minor interruptions were recorded during administration of the remote general neuropsychological battery. On two occasions, the experimenter observed that the internet connection had slowed considerably for a brief period of time; however, this was not explicitly commented on by either participant affected (one healthy control, one bvFTD patient). The third interruption was a dog barking during an AD patient's testing session. In all cases, the trial that the interruption had occurred in was restarted, and the examiner judged that there was no penalty or benefit afforded by the interruption in any case. No interruptions were recorded during administration of the neurolinguistic battery, and no other interruptions were reported by participants.

Table S1. List of general neuropsychological and neurolinguistic tests delivered face-to-face that were not included in the remote battery.

Test	Reason for removal
WASI Vocabulary	Other tests from same domain already included
WASI Similarities	Other tests from same domain already included
Long Recognition Memory Test for Faces	To reduce the length of the remote battery, this was replaced with the Short Recognition Memory Test for Faces
Long Recognition Memory Test for Words	To reduce length of battery
Camden Paired Associates Learning	Other tests from same domain already included
WASI Block Design	Not feasible online as participants would not have blocks to manipulate
Stroop task	Not feasible online as too difficult for the examiner to determine the participant's target and therefore whether or not they had made an error. Also differences in colour display across participants' screens.
Trails A and B	Not feasible online as requires use of pencil and paper
WAIS-R Digit Symbol	Not feasible online as requires use of pencil and paper
Usual/Unusual views	Other tests from same domain already included
Modified Kissing and Dancing	Other tests from same domain already included
Baxter Spelling Test	Not feasible online as face-to-face task requires the participant to use pencil and paper; typing responses was not considered appropriate due to potential interference from spell-check software
Written sentences	Not feasible online as face-to-face task requires the participant to use pencil and paper; typing responses was not considered appropriate due to potential interference from spell-check software
Spatial Span Forwards	To reduce length of battery
Spatial Span Backwards	To reduce length of battery

We reduced the number of tests used in our face-to-face general neuropsychological and neurolinguistic batteries down for remote testing, reflecting a) the need to make the remote testing batteries shorter to minimise fatigue; b) impracticalities of administering certain stimuli remotely; and c) inability to adequately record participants' responses to some tasks. The Table shows the tasks that were not included in the remote batteries.

Table S2. Bayesian statistics comparing general neuropsychological test performance on remote vs face-to-face assessments

Test	CTL	All Patients	AD	bvFTD	SD	PNFA	LPA
General intellect							
WASI Matrix	2.56[t] (Anecdotal)	5.403 (Substantial)	3.134[t] (Substantial)	1.477[t] (Anecdotal)	1.911[t] (Anecdotal)	1.356[t] (Anecdotal)	1.441[t] (Anecdotal)
Episodic memory							
RMT Faces short	NA	3.46 (Substantial)	0.411[t] (Anecdotal)	NA	NA	NA	NA
Working memory							
DS (Reverse)	4.304[t] (Substantial)	5.28 (Substantial)	2.746[t] (Anecdotal)	Variance at 0	1.991[t] (Anecdotal)	2.659 (Anecdotal)	0.417[t] (Anecdotal)
Short-term verbal memory							
DS (Forward)	2.663[t] (Anecdotal)	4.54[t] (Substantial)	3.467[t] (Substantial)	1.948[t] (Anecdotal)	0.799[t] (Anecdotal)	2.528[t] (Anecdotal)	1.598[t] (Anecdotal)
Language							
BPVS	4.304[t] (Substantial)	4.05 (Substantial)	3.363 (Substantial)	1.836 (Anecdotal)	1.382[t] (Anecdotal)	2.754 (Anecdotal)	1.261 (Anecdotal)
GNT	2.923[t] (Anecdotal)	4.16 (Substantial)	3.542 (Substantial)	1.407[t] (Anecdotal)	2.331 (Anecdotal)	1.68[t] (Anecdotal)	1.325[t] (Anecdotal)
NART	2.83[t] (Anecdotal)	1.99 (Anecdotal)	1.552 (Anecdotal)	1.511[t] (Anecdotal)	2.606[t] (Anecdotal)	2.52[t] (Anecdotal)	1.948[t] (Anecdotal)
Category fluency	3.767[t] (Substantial)	1.08 (Anecdotal)	3.532 (Substantial)	2.186[t] (Anecdotal)	1.715[t] (Anecdotal)	1.027 (Anecdotal)	1.29[t] (Anecdotal)
Arithmetic							
GDA Total	3.671 (Substantial)	5.30 (Substantial)	3.33 (Substantial)	2.467[t] (Anecdotal)	0.943[t] (Anecdotal)	2.483[t] (Anecdotal)	1.123[t] (Anecdotal)
Visuospatial							
VOSP	0.0404 (Substantial)	4.76 (Substantial)	0.171[t] (Substantial)	2.215 (Anecdotal)	2.478 (Anecdotal)	2.553 (Anecdotal)	2.592[t] (Anecdotal)
Executive							

Letter fluency	3.159[t] (Substantial)	0.59 (Anecdotal)	2.86[t] (Anecdotal)	2.426[t] (Anecdotal)	1.519[t] (Anecdotal)	0.511[t] (Anecdotal)	0.188[t] (Substantial)

A Bayes factor (BF_{01}) indicates the extent to which the null hypothesis is favoured against the alternative hypothesis (e.g., a BF_{01} value of 4 means that the obtained data are 4 times more likely under the null hypothesis than under the alternative hypothesis). A $BF_{01} > 3$ is therefore considered as substantial evidence in support of the null hypothesis; while a BF_{01} of <1/3 is considered as substantial evidence in support of the alternative hypothesis. Any values in between are categorised as 'anecdotal' evidence, equivalent to a non-significant result in inferential statistics [3]. Results are influenced by the prior (more specifically the shape of the prior influences the strength of the evidence), which can be specified be default using a Cauchy distribution, as here; the Cauchy scale set here is 1.00. The superscript [t] indicates that a parametric Bayesian test was used; else the non-parametric Mann Whitney (with 1000 iterative samples) was employed. Blue shading indicates that the alternative hypothesis (H1, i.e. there was a difference in performance across the two environments) was favoured with substantial evidence; Green shading indicates that the null hypothesis (H0; i.e. there was no difference in performance across environments) was favoured with substantial evidence. AD, patient group with typical Alzheimer's disease; BPVS, British Picture Vocabulary Scale; bvFTD, patient group with behavioural variant frontotemporal dementia; CTL, healthy control group; DS, Digit Span; F2F, face-to-face; GDA, Graded Difficulty Arithmetic test; GNT, Graded Naming Test; LPA, patient group with logopenic progressive aphasia; Matrix, WASI Matrix Reasoning; NART, National Adult Reading Test; PNFA, patient group with progressive nonfluent aphasia; RMT, Recognition Memory Test; SD, patient group with semantic dementia; VOSP, Visual Object Space Perception battery.

Table S3. Bayesian statistics comparing neurolinguistic test performance on remote vs face-to-face assessments

Test	CTL	All Patients	SD	PNFA	LPA
Phoneme perception					
PALPA 3	1.003 (Anecdotal)	2.918 (Anecdotal)	1.594 (Anecdotal)	2.731 (Anecdotal)	2.079 (Anecdotal)
Reading					
Non word reading	1.544 (Anecdotal)	3.239 (Substantial)	1.940^{t} (Anecdotal)	2.033^{t} (Anecdotal)	2.428^{t} (Anecdotal)
Regular reading	Variance at 0	2.989 (Anecdotal)	1.975 (Anecdotal)	2.305 (Anecdotal)	1.923 (Anecdotal)
Irregular reading	Variance at 0	2.996 (Anecdotal)	2.256 (Anecdotal)	2.194^{t} (Anecdotal)	2.005 (Anecdotal)
Naming					
BNT	3.612^{t} (Substantial)	1.995 (Anecdotal)	0.584^{t} (Anecdotal)	2.580 (Anecdotal)	1.490 (Anecdotal)
Semantic association					
Camel and cactus	4.039^{t} (Substantial)	1.067^{t} (Anecdotal)	1.524^{t} (Anecdotal)	2.102 (Anecdotal)	N<2 for F2F
Word comprehension					
Concrete synonyms	2.739^{t} (Anecdotal)	3.622 (Substantial)	2.009^{t} (Anecdotal)	2.400^{t} (Anecdotal)	1.413^{t} (Anecdotal)
Abstract synonyms	1.726^{t} (Anecdotal)	2.718 (Anecdotal)	0.782 (Anecdotal)	2.741 (Anecdotal)	1. 941 (Anecdotal)
Sentence comprehension					
PALPA55	0.944 (Anecdotal)	3.954 (Substantial)	2.246 (Anecdotal)	2.603 (Anecdotal)	2.110 (Anecdotal)
Speech repetition					
Monosyllabic word repetition	0.0487^{t} (Substantial)	0.117 (Anecdotal)	0.477^{t} (Anecdotal)	1.167^{t} (Anecdotal)	1.055^{t} (Anecdotal)
Bisyllabic word repetition	4.304^{t} (Substantial)	3.744 (Substantial)	1.086 (Anecdotal)	2.639 (Anecdotal)	0.650^{t} (Anecdotal)
Trisyllabic word repetition	Variance at 0	3.701 (Substantial)	1.339 (Anecdotal)	2.400 (Anecdotal)	1.677 (Anecdotal)
Graded difficulty sentence repetition	1.779 (Anecdotal)	2.983 (Anecdotal)	1.420^{t} (Anecdotal)	2.535^{t} (Anecdotal)	1.996 (Anecdotal)
Sentence construction					
Spoken	Variance at 0	1.525 (Anecdotal)	2.270 (Anecdotal)	1.960. (Anecdotal)	0.930^{t} (Anecdotal)

A Bayes factor (BF_{01}) is shown for each remote vs face-to-face testing comparison. The interpretation and colour coding are as indicated in the legend to Table S2 above. BNT, Boston Naming Test; CTRL, healthy control group; F2F, face-to-face; LPA, patient group with logopenic progressive aphasia; N, number of participants per group; PALPA, Psycholinguistic Assessment of Language Processing in Aphasia subtests; PNFA, patient group with progressive nonfluent aphasia; SD, patient group with semantic dementia

Table S4. Equality of variance analyses

Test	Healthy controls	Combined patients
General neuropsychology tasks		
WASI Matrix	F = 0.50, p = 0.32	F* = 0.15, p = 0.70
RMT faces short	N/A	F*=0.67, p=0.42
Digit span forwards	F = 1, p = 1	F = 1.36, p = 0.41
Digit span backwards	F = 1, p = 1	F* = 0.20, p = 0.65
BPVS	F = 0.82, p = 0.77	F* = 2.56, p = 0.11
GNT	F = 1.38, p =0.64	F* = 1.15, p = 0. 29
GDA	F* = 1.14, p = 0.30	F* = 0.13, p = 0.72
VOSP	F* = 0.42, p = 0.52	F* = 0.001, p = 0.97
NART	F = 1.15, p = 0.84	F* = 1.72, p = 0.19
Letter fluency	F = 2.20, p = 0.26	F* = 1.28, p = 0.26
Category fluency	F = 2.31, p = 0.23	F* = 0.96, p = 0.33
Neurolinguistic tasks		
PALPA-3	F* = 0.72, p = 0.41	F* = 0.07, p = 0.79
Nonword reading	F* = 3.08, p = 0.10	F* = 1.31, p = 0.26
Regular word reading	Variance=0	F* = 0.70, p = 0.41
Irregular word reading	Variance=0	F* = 0.30, p = 0.59
BNT	F = 0.57, p = 0.42	F* = 1.37, p = 0.25
Mono-syllabic repetition	**F = 0.11, p < 0.05**	**F* = 4.89, p = 0.03**
Bi-syllabic repetition	F = 0.58, p = 0.43	F* = 0.001, p = 0.98
Tri-syllabic repetition	Variance=0	F* = 0.04, p = 0.85
Concrete synonyms	F = 1.19, p = 0.80	F* = 0.91, p = 0.35
Abstract synonyms	F = 0.29, p = 0.09	F* = 2.24, p = 0.15
PALPA-55	F* = 4.41, p=0.05	F* = 1.10, p = 0.30
Sentence repetition	Variance=0	F* = 0.48, p = 0.49
Camel and cactus	F* = 1.30, p = 0.70	**F = 4.25, p = 0.02**
Spoken sentences	Variance=0	**F* = 4.37, p = 0.046**

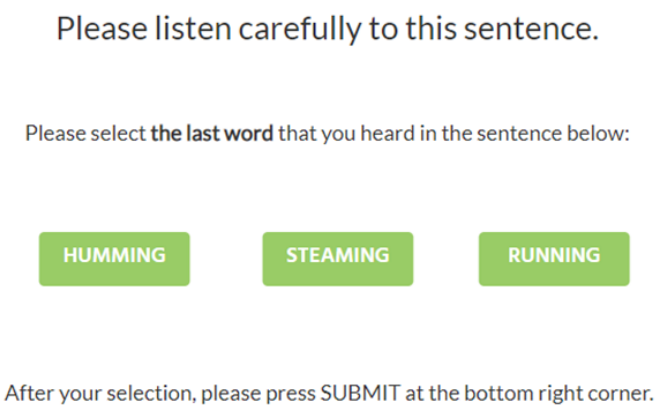
Results of equality of variance analyses. F-test results are reported; *indicates that Levene's test was adopted instead due to violations of the assumption of normality. **Bold** indicates that the test was significant, meaning that the assumption of homogeneity of variance was violated. The short version of the RMT faces task was not administered to healthy control participants as part of their remote testing battery, making an equality of variance test impossible here. BPVS, British Picture Vocabulary Scale; GDA, Graded Difficulty Arithmetic test; GNT, Graded Naming Test; Matrix, WASI Matrix Reasoning; NART, National Adult Reading Test; VOSP, Visual Object Space Perception Object Decision task. BNT, Boston Naming Test; PALPA, Psycholinguistic Assessment of Language Processing in Aphasia.

Table S5. Performance on audibility screening task by remote participant groups

	CTL	AD	bvFTD	SD	PNFA	LPA
Pre neuropsychology battery						
Average number of incorrect items (/10)	0.0 (0.0)	0.1 (0.4)	0.0 (0.0)	0.0 (0.0)	0.2 (0.4)	0.0 (0.0)
Average number of errors on last three items (/3)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
Pre neurolinguistic battery						
Average number of incorrect items (/10)	0.0 (0.0)	NA	NA	0.0 (0.0)	0.4 (0.9)	0.0 (0.0)
Average number of errors on last three items (/3)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)

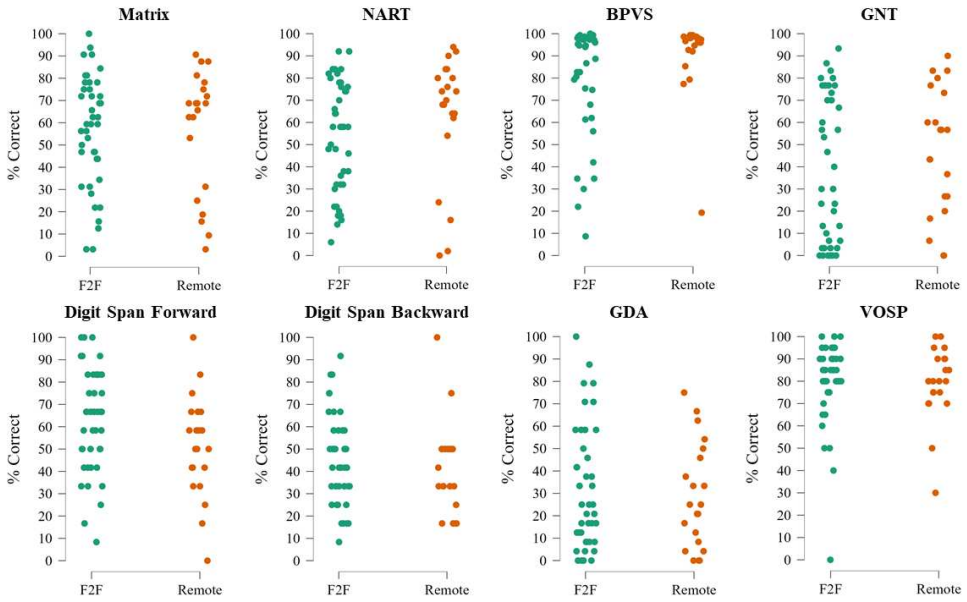
The data indicate that there were no major background listening environmental confounds nor any significant differences between participant groups (all $p>0.05$). AD, patient group with typical Alzheimer's disease; bvFTD, patient group with behavioural variant frontotemporal dementia; CTL, healthy control group; LPA, patient group with logopenic progressive aphasia; NA, not applicable; PNFA, patient group with progressive non-Fluent aphasia; SD, patient group with semantic dementia.

Figure S1. Example of basic audibility screening measure



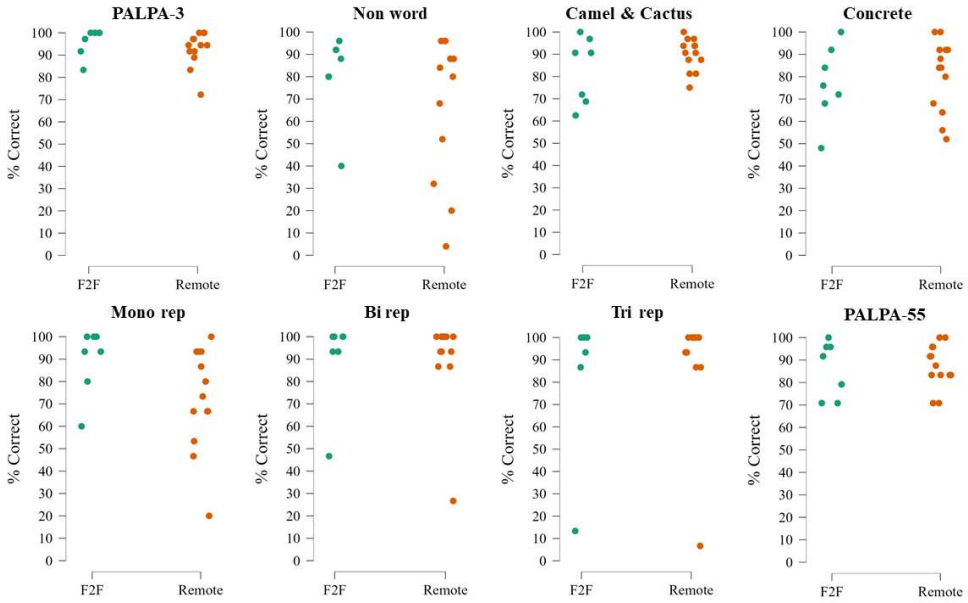
The Figure shows a Labvanced display of the BKB hearing screening measure. In this example, the sentence spoken was "The car engine is running". For each sentence, two foils were displayed alonside the target, both of which made sense in the sentence when replacing the target. One of the foils was also selected to loosely rhyme with the target word (here, "humming").

Figure S2. Performance profiles of patients on tasks in general neuropsychological battery.



Scatter plots showing performance profiles of individual patients on tasks in the general neuropsychological battery. BPVS, British Picture Vocabulary Scale; GDA, Graded Difficulty Arithmetic test; GNT, Graded Naming Test; Matrix, WASI Matrix Reasoning; NART, National Adult Reading Test; VOSP, Visual Object Space Perception Object Decision task.

Figure S3. Performance profiles of patients on tasks in the neurolinguistic battery.



Scatter plots indicating percentage scores for each patient on representative tests from the neurolinguistic battery administered face-to-face and remotely. Bi rep, bisyllabic single word repetition; BNT, Boston Naming Test; Concrete, concrete synonyms test; F2F, face-to-face; Mono rep, monosyllabic single word repetition test; Non word, non-word reading test; PALPA, Psycholinguistic Assessment of Language Processing in Aphasia; Tri rep, trisyllabic single word repetition.

References

1. Bench J, Kowal Å, Bamford J. The Bkb (Bamford-Kowal-Bench) Sentence Lists for Partially-Hearing Children. *British Journal of Audiology* 1979;13(3):108-12. doi: 10.3109/03005367909078884
2. Finger H, Goeke C, Diekamp D, et al. LabVanced: A Unified JavaScript Framework for Online Studies. *International Conference on Computational Social Science* 2016
3. Ashworth M, Palikara O, Burchell E, et al. Online and Face-to-Face Performance on Two Cognitive Tasks in Children With Williams Syndrome. *Front Psychol* 2021;11:594465. doi: 10.3389/fpsyg.2020.594465