Creation of an Open Science Dataset from PREVENT-AD, a Longitudinal Cohort Study of Pre-symptomatic Alzheimer's Disease.

Jennifer Tremblay-Mercier[1], Cécile Madjar[1,2], Samir Das[2], Stephanie O.M. Dyke[2,5,8], Pierre Étienne[1,5], Marie-Elyse Lafaille-Magnan[1,4,5], Pierre Bellec[3,6], D. Louis Collins[5,8], M. Natasha Rajah[1,5], Veronique D. Bohbot[1,5], Jeannie-Marie Leoutsakos[7], Yasser Iturria-Medina[2,5,8], Justin Kat[1,2], Richard D. Hoge[2,5,6,8], Serge Gauthier[1,5,9], M. Mallar Chakravarty[1,5], Jean-Baptiste Poline[5,8] Pedro Rosa-Neto[1,5,8,9], Sylvia Villeneuve[1,2,8], Alan C. Evans[1,2,5,8], Judes Poirier[1,5], John C. S. Breitner[1,5] & the PREVENT-AD Research Group**

1. *StoP-AD Centre, Douglas Mental Health Institute Research Centre, Montréal, QC, Canada*
2. *McGill Centre for Integrative Neuroscience, Montreal Neurological Institute, McGill University, Montréal, QC, Canada*
3. *CRIUGM - Université de Montréal, Montréal, QC, Canada*
4. *Centre for Child Development and Mental Health, Jewish General Hospital. Montréal, QC, Canada*
5. *McGill University, Montréal, QC, Canada*
6. *Université de Montréal, Montréal, QC, Canada*
7. *John Hopkins University School of Medicine, Baltimore, MD, USA,*
8. *McConnell Brain Imaging Center, Montreal Neurological Institute, McGill University, Montréal, QC, Canada*
9. *McGill University Research Centre for Studies in Aging, McGill University, Montréal, QC, Canada*

Corresponding author: Judes Poirier (judes.poirier@mcgill.ca)

***Data used in preparation of this article were obtained from the Pre-symptomatic Evaluation of Novel or Experimental Treatments for Alzheimer's Disease (PREVENT-AD) program (https://douglas.research.mcgill.ca/stop-ad-centre), data release 5.0 (November 30, 2017). A complete listing of the PREVENT-AD Research Group can be found at:* https://preventad.loris.ca/acknowledgements/acknowledgements.php?date=[2017-12-01].

ABSTRACT:

We describe the creation of an open science dataset from a cohort of cognitively unimpaired aging individuals with a parental or multiple-sibling history of Alzheimer's disease (AD). Our purpose was to enable PResymptomatic EValuation of Novel or Experimental Treatments for AD ("PREVENT-AD"). To characterize this population, possibly progressing in the pre-symptomatic phase of AD, we studied genetic variants and obtained longitudinal measures of cognition, brain structure and function, blood and cerebral fluid biochemistry and neurosensory capacities. Two nested prevention trials were also conducted. Data were hosted in LORIS, a platform that facilitates data organization, curation and sharing. We initially assessed 425 individuals, 385 meeting criteria for sustained investigation and 330 remaining active for longitudinal follow-ups. Between 2011 and 2017, we obtained quality-controlled data from 1704 MRI scans, 532 CSF samples, and 1882 cognitive evaluations. To date, 310 active participants (94%) have agreed that their data be openly shared. In addition to being a living resource for continued data acquisition, therefore, PREVENT-AD offers shared data to facilitate understanding of AD pathogenesis.

BACKGROUND AND SUMMARY

Dementia is the final stage of Alzheimer's disease (AD), representing the culmination of a process that begins decades before onset of symptoms.[1-3] Characterizing and tracking the pre-symptomatic stage of AD requires methods sensitive to the disease's early manifestations. These may include not only subtle cognitive decline, but also biochemical changes and structural or functional brain alterations. Studying these pre-symptomatic changes is crucial to a full understanding of AD, and their precise measurement is critical for trials of interventions that seek to prevent symptom onset.

To meet this challenge, in 2010, investigators at McGill University and the Douglas Mental Health University Institute Research Centre created a Centre for

Studies **o**n **P**revention of **A**lzheimer's **D**isease (StoP-AD). The Centre's prime objective was to pursue innovative studies of pre-symptomatic AD, with efforts to provide a relatively enriched population sample for prevention trials requiring individuals 'at-risk' of developing the disease.[8] To this end, the StoP-AD Centre developed an observational cohort for **PR**e-symptomatic **EV**aluation of **E**xperimental or **N**ovel **T**reatments for **AD** ("PREVENT-AD"). This cohort consists of cognitively unimpaired persons with a parental or multiple-sibling history of AD-like dementia, a population having a 2-3 fold relative increase in risk of AD dementia.[9,10] The work began with naturalistic longitudinal follow-up of cognitive, neurosensory, biochemical, and magnetic resonance imaging (MRI) measurements accompanied by clinical and medical assessments. Two clinical trials of putative pharmaco-preventive agents were nested in this observational cohort study. To enrich the characterization of the participants, new evaluations were added over the years. StoP-AD collaborative structure facilitates the development of methods and models with the capacity of tracking AD progression with greater inference than the study of individual biomarkers, by combining a variety of indicators in composite metrics for example. The StoP-AD Centre's collaborative structure is shown in Figure 1.

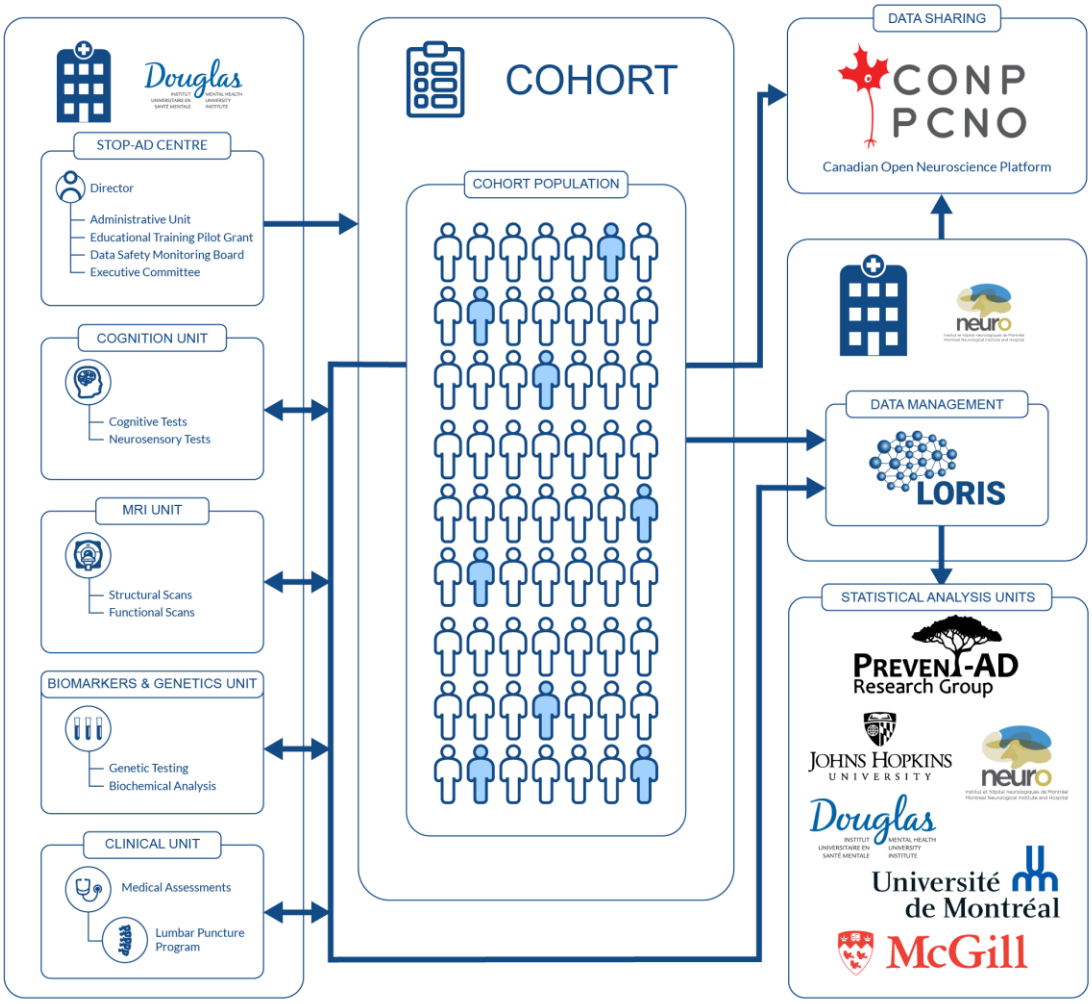

Figure 1: Structural organization of the StoP-AD Centre. Left panel: Research units. Middle panel: Study population (blue shading representing participants enrolled in nested trials). Right panels: Collaborating institutions and research groups involved in data collection, management, and sharing.

The StoP-AD Centre uses LORIS as its data management platform for storage and curation of data.[11,12] LORIS was designed to facilitate sharing of data with research collaborators. The Centre shares data and with more than 12 collaborative research groups. Recently, a portion of the PREVENT-AD data was made available more widely under the principles of open science[13-15] (https://openpreventad.loris.ca/). The PREVENT-AD data sharing resource is a major initiative of the Canadian Open Neuroscience Platform (CONP; https://conp.ca), and the Tanenbaum Open Science Institute.[16]

METHODS

A. Overview

Here we briefly describe the PREVENT-AD cohort and associated clinical trials, including the data acquisition strategy, methods and the infrastructure used for the curation and dissemination of data to the wider research community.

1. Observational Cohort

Recruitment to the observational PREVENT-AD cohort began in November 2011 but was suspended, owing to funding constraints, in May 2017. To increase the probability that cognitively intact participants would harbor early changes of pre-symptomatic AD, entry criteria rested on two broad principles of advanced age and a parental or multiple-sibling history of AD. Participants were 60 years of age or older, excepting persons between 55 - 59 years old who were eligible if their own age was within 15 years of symptom onset in their youngest-affected first-degree relative. Participants' family history of "AD-like dementia" was ascertained either by compelling report of an AD diagnosis from an experienced clinician or, if such was not available, by use of a structured questionnaire developed for the Cache County Study and intended to establish memory or concentration issues in first-degree relatives sufficiently severe to cause disability or loss of function, with an insidious onset or gradual progression (as opposed to obvious consequences of a stroke or other sudden insult). Enrollment further required confirmation of intact cognition, stable general health and availability of a study partner to provide information on daily functioning (Table 1). For more details about recruitment and eligibility determination, see section B.1.

Table 1: Inclusion and Exclusion Criteria

Inclusion criteria
➔ Parental or multiple-sibling (defined by 2 or more) history of Alzheimer-like dementia ➔ Age 60 years or older (persons aged 55-59 years and < 15 years younger than their affected index relative were also eligible.) ➔ Minimum of 6 years of formal education ➔ Study partner available to provide information on cognitive status ➔ Sufficient fluency in spoken and written French and/or English ➔ Ability and intention to participate in regular visits ➔ Provision of informed consent ➔ Agreement for periodic donation of blood and urine samples ➔ Agreement to participate in periodic multimodal assessments via MRI and LP for CSF collection (LP optional at first, then mandatory for participation) ➔ Agreement to limit use of medicines as required by investigational protocols, if applicable
Exclusion criteria
➔ Cognitive disorders - Known or identified during eligibility assessments (MoCA and CDR) ➔ Use of acetyl-cholinesterase inhibitors including tacrine, donepezil, rivastigmine, galantamine ➔ Use of memantine or other approved prescription cognitive enhancer ➔ Use of vitamin E at greater than 600 i.u. / day or aspirin at >325 mg / day ➔ Use of opiates (oxycodone, hydrocodone, tramadol, meperidine, hydromorphone) ➔ Use of NSAIDs or regular use of systemic or inhalation corticosteroids ➔ Clinically significant hypertension (accepted if controlled medically), anemia, significant liver or kidney disease ➔ Concurrent use of warfarin, ticlopidine, clopidrogel, or similar anti-coagulant ➔ Current plasma Creatinine >1.5 mg/dl (132 mmol/l) ➔ Current alcohol, barbiturate or benzodiazepine abuse/dependence

CSF: cerebrospinal fluid; MRI: magnetic resonance imaging; NSAID: non-steroidal anti-inflammatory drug; LP: lumbar puncture

After telephone and on-site screening, eligible participants were enrolled and followed annually with structured evaluations. Cognitive performances (immediate memory, delayed memory, language, attention and visuospatial capacities) were assessed by the **R**epeatable **B**attery for **A**ssessment of **N**europsychological **S**tatus (RBANS)[26] and neurosensory abilities were evaluated by measuring olfactory identification abilities using the standardized **U**niversity of **P**ennsylvania **S**mell Identification **T**est (UPSIT).[35] At each visit, the clinical team obtained blood and urine samples and performed neurological and physical examinations, including electrocardiogram. Further 'in-house' medical history and review of systems questionnaires were also administered. Participants also underwent an MRI scanning session of 1 to 1.5 hours including numerous

structural and functional acquisitions. On a separate day, participants who consented to the procedure donated CSF samples via lumbar puncture (LP). Initially, we performed lumbar punctures only on participants enrolled in clinical trials. In 2016, however, considering the overall success of the LP program (acceptance, tolerability, and retention through serial repetitions) we began also to perform serial LPs in the broader observational cohort. In 2017, consent for such LPs became an inclusion criterion for new participants. Over time, various other modalities were added. These included: i) in 2014, evaluation of central auditory processing, a neurosensory function; ii) in 2015, evaluation of subjective cognitive impairment (Everyday Cognition test - ECog); and iii) in 2016, a modified MRI protocol designed to investigate the integrity of hippocampal subfields and brain microstructure (iron deposition, myelination concentration) (Fig. 2). Telephone follow-ups (FU) were conducted between on-site annual visits to continue contact and update clinical information (Fig. 3).

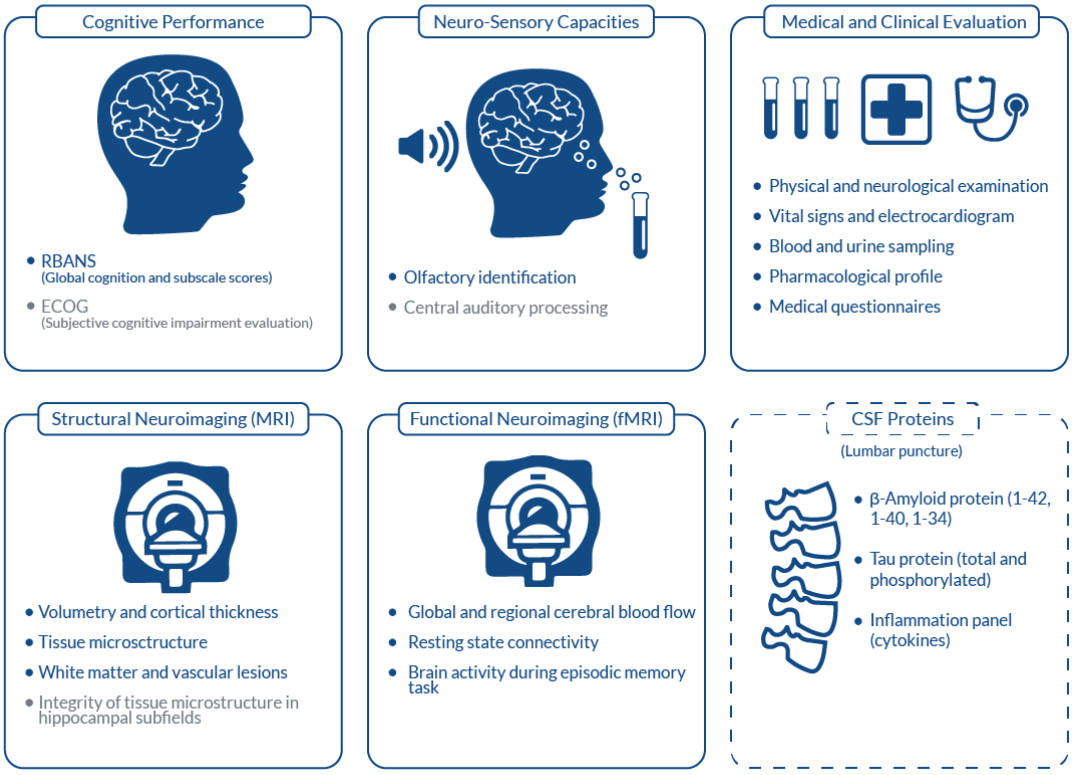

Figure 2: PREVENT-AD evaluations performed annually - November 2011 to November 2017. Most evaluations were conducted from the program's onset (blue items), while some were added later (gray items). Lumbar punctures (dotted line) were originally performed in clinical trials participants but then became optional in the observational cohort and most recently, an integral part of the program.

2. Preventive intervention trials

We conducted two clinical trials nested in PREVENT-AD to test potentially preventive pharmaceutical agents. Described below, these trials were INTREPAD, a randomized, placebo-controlled trial of low dose naproxen sodium, and DEPEND, a proof-of-concept trial of the lipid lowering agent probucol as a potential inducer of apolipoprotein E (apoE) protein availability (Fig. 3).

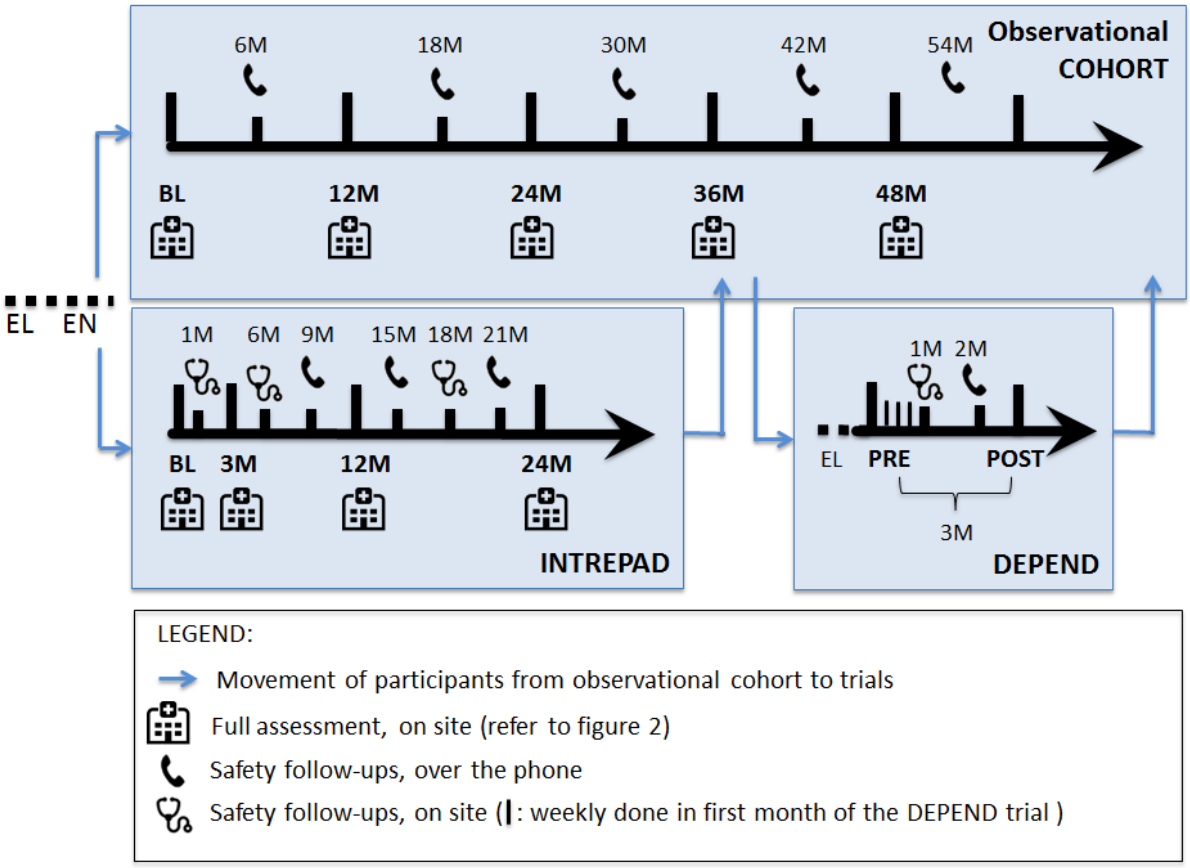

FIGURE 3: Timelines of observational cohort, INTREPAD & DEPEND trials. EL: eligibility visit; EN: enrolment visit; BL: Baseline visit; M: months;

a. INTREPAD (***I****nvestigation of* ***N****aproxen* ***TR****eatment* ***E****ffects in* ***P****re-symptomatic* ***A****lzheimer's* ***D****isease;* clinicaltrials.gov NCT02702817*)* was a two-year double-masked trial of the non-steroid anti-inflammatory drug (NSAID) naproxen sodium 220 mg b.i.d. vs. placebo in 195 PREVENT-AD participants. Recruitment for INTREPAD began in March 2012 and ended in March 2015. Standard annual PREVENT-AD evaluations (Fig. 2) were supplemented with an additional session three months after randomization. The 3-month assessment was intended to determine whether treatment-related changes, if any, occurred gradually or as a rapid response. LPs were optional, but were undertaken by over half of all participants. The primary outcome was a composite Alzheimer Progression Score (APS, described below) derived using item response theory from various

cognitive and biomarker measures.[17] Results of INTREPAD were published in Meyer *et. al.*, 2019.[18]

b. DEPEND (**D**osage and **E**fficacy of **P**robucol-Induced apo**E** to **N**egate cognitive **D**eterioration; clinical trials.gov NCT02707458) was a single-arm proof-of-concept trial planned as a 3-month dose-finding phase, followed by a 1-year validation and follow-up phase. As suggested by earlier pilot data, its principal outcome was change in concentration of apoE. Secondary outcomes were corresponding reduction in vascular biomarkers in CSF and blood.[19] LPs were therefore obligatory. Twenty-four participants enrolled in the first phase were given a standard dose of probucol (600 mg), with intention to develop an individualized dosing regimen. Data collection for the first 3-month phase occurred from June through December 2016. The follow-up phase aimed to treat participants with personalized doses of probucol over one year to observe the specified outcomes.

3. Data Sharing Initiatives

We are making PREVENT-AD data broadly available as an open science resource providing opportunities to reuse those data for additional research analysis, as well as potential collaborations. Because the original PREVENT-AD ethics approval and consent process did not fully address open science data sharing plans, several steps related to ethics were required (see 'Open Science Ethics' section B.6.).

The PREVENT-AD dataset is hosted by the Canadian Open Neuroscience Platform (CONP) in the context of the Tanenbaum Open Science Institute (TOSI) at the Montreal Neurological Institute (MNI). To our knowledge, the MNI is the first clinical research institute to announce open science as a core principle.[15] The Canadian Open Neuroscience Platform (CONP), a unified metadata sharing interface for the neuroscience community will contribute to the TOSI. The first open science release of PREVENT-AD data includes a portion of longitudinal

data from the observational cohort and from the main pharmaco-prevention trial, INTREPAD collected between November 2011 and November 2017.

B. DETAILS OF METHODS

1. Recruitment and eligibility determination

The main recruitment strategy was distribution of >250,000 flyers throughout Montreal and surrounding cities via a commercial coupon packet service (Publi-sac). Distribution was targeted based on demographic data provided by the vendor. Appearances in the media (TV, radio, newspapers) also provided some help in recruitment. However, like many researchers, PREVENT-AD investigators observed that mass distribution of invitation material (not talks at interest groups or civic organizations, or even local chapters of the Alzheimer's society and similar organizations) was the most effective method for recruitment of large numbers.

After the initial indication of interest, an efficient, participant-friendly recruitment process followed. A study nurse performed preliminary eligibility screening over the phone or via an online questionnaire. An on-site eligibility visit then included more specific questions on family history of AD dementia, medical and surgical history, pharmacological profile, and lifestyle habits, as well as physical and neurological examinations, blood and urine sampling and an electrocardiogram (EKG). Two cognitive screening tests assessed integrity of cognition: the Montreal Cognitive Assessment (MoCA) and the Clinical Dementia Rating (CDR),[20,21] including its brief cognitive test battery. When cognitive status was in doubt (MoCA ≤ 26/30 or CDR >0), a complete evaluation was obtained from a neuropsychologist was obtained. A ~30-minute MRI session was also run to rule out structural brain disease, while simultaneously ensuring participants' familiarity with the MRI environment. A practice session for an episodic memory task was undertaken as part of this session. Final determination of eligibility for PREVENT-AD (and, when at issue, for nested prevention trials) was made by clinical consensus of one or more study physicians, a research nurse, and a

neuropsychologist. Trial inclusion/exclusion criteria are specified in a publication describing results from INTREPAD[18] or, for DEPEND, by request to the Centre. See Figure 4 for the participant flow in the recruitment process.

All consent procedures fulfilled modern requirements for human subject's protection, while avoiding excess participant burden. Consent forms were carefully crafted to use simple but comprehensive language (typically at an 8th grade reading level).

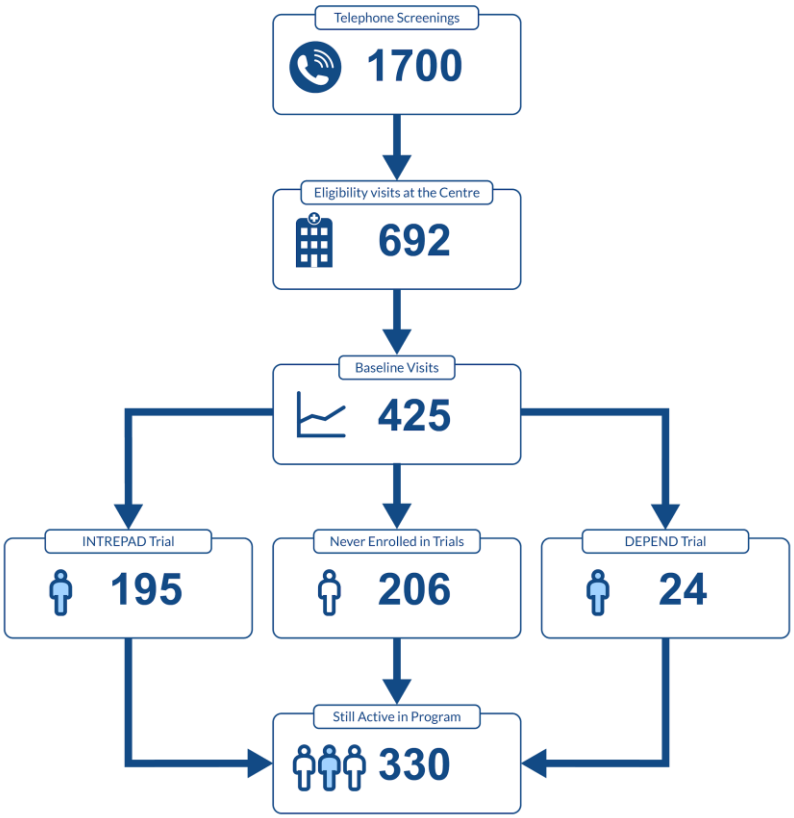

Figure 4: Flow diagram of the recruitment progress, baseline assessments distribution and retention. *We received about 1700 telephone calls from interested individuals and, after a short telephone screen, we invited 692 persons to the Centre for eligibility evaluation. A total of 425 participants completed their baseline visit (BL). Of these, 195 constituted BL visits of the INTREPAD trial. The remaining 230 BL assessments became part of the observational cohort. Twenty- four participants eventually took part in DEPEND. All trial participants were invited to continue their participation in the observational cohort after completion of their trial, resulting in a total of 330 participants still active in the program. These numbers do not represent the exact numbers with data available for analysis and sharing as explained in section B.2.*

2. Characteristics of the population

The range of data collected at recruitment and updated at each FU encounter is shown in Table 2.

Table 2: Demographics and Clinical Information

*Demographics**	Date of birth Gender Education level Mother tongue Marital status Professional status Language Ethnicity Residence Housing Handedness
Clinical data	Family history of Alzheimer-like dementia Pharmacological profile Blood hematology, biochemistry, coagulation, endocrinology Urine biochemistry Electrocardiogram Blood pressure / pulse Height* / Weight / body mass index Cardiovascular risk score (CAIDE score)* Medical and surgical history Physical activity level Physical and neurological examination Alcohol consumption Smoking habit Sleeping habits Review of Systems Adverse events / Compliance to study drug (when applicable)
Genetics*	Variants of: APOE, BDNF, HMGR intronM, BCHEK, TLR4

*collected once

Genotyping. All participants consented to blood acquisition for genotyping, but genetic analysis was performed only in those who were confirmed eligible. DNA was isolated from 200 µl whole blood using a QIASymphony apparatus and the DNA Blood Mini QIA Kit (Qiagen, Valencia, CA, USA). The standard

QIASymphony isolation program was used following the manufacturer's instructions. As indicated, allelic variants of six genes associated with AD were determined using pyrosequencing (PyroMArk96).[22-25]

From the 425 participants who underwent baseline visits, 385 were confirmed as appropriate for final data analysis. Among the 40 exclusions, 31 were judged unsuitable for continued participation because of cognitive deficits that had escaped detection but became apparent upon more detailed testing at baseline. Other reasons for exclusion included similar post-enrollment detection of stroke (2), anxiety and attention problems (2), refusal of further MRI (2), or discovery that their AD family history in fact failed to fulfill entry criteria (3). Table 3 summarizes key baseline characteristics of the remaining 385 members of the analysis pool. The somewhat smaller number available for data sharing reflects the ongoing process of re-consent to the open science sharing. As of January 2020, we contacted the remaining group of active participants (n=330). Information about participants who are lost to follow-up or who have withdrawn will be presented with the forthcoming data sharing plan as a second step. Out of the 330 who remained active, 310 accepted (94%), 15 refused (5%) and 5 did not answer (2%). In addition to this re-consent process, data curation, including quality controls may have excluded few participants.

TABLE 3: Key baseline characteristics of PREVENT-AD participants

Number in final analysis pool	**385**
Age (years) average ± SD	65.4 ± 5.2
Gender (M / F)	111 / 274
Education (years) average ± SD	16.5 ± 3.4
MoCA score (out of 30)	28.0 ± 2.1
APOEε4 status (%)	4-4 = 2.09% 4-3 = 31.85% 4-2 = 4.18%

3. Longitudinal follow-up

Data for the observational cohort and trial participants are described separately. The observational dataset included results of 200 baseline (BL) evaluations, 186 12-month FU, 142 24-month FU, 98 36-month FU and 50 48-month FU. INTREPAD trial data potentially available for analysis derived from 185 candidates at baseline, 160 at 3-month FU, 152 at 12-month FU and 147 at 24-month FU. Additional annual FU following completion of the trial are available from 134 candidates at 36-month FU and for 122 at 48-months FU. The DEPEND trial enrolled 24 participants of whom 19 completed the 3 month proof-of-concept trial (Fig. 5).

Note that the available dataset is built with data collected before the November 27th 2017 'end date' for construction of the available dataset, and do not include results of later data collection. Given the imposition of a fixed date, not all participants were able to complete 4 years of follow-up (Fig. 5). The withdrawal rate was only about 3 % per year, which testifies to the success of the program's retention efforts. The latter included deliberate attempts to strengthen participants sense of 'belonging' to the study cohort, including frequent and helpful telephone contacts with the study team. Medical staff offered participants pertinent health advice, support and referrals to other professionals when needed. The Centre provided an atmosphere that encouraged discussions, and sent participants regular newsletters reporting on the research progress and information on AD. An annual luncheon event celebrated the dedication of all participants, informed them of the most recent findings in AD research, and facilitated participant exchanges and interactions.

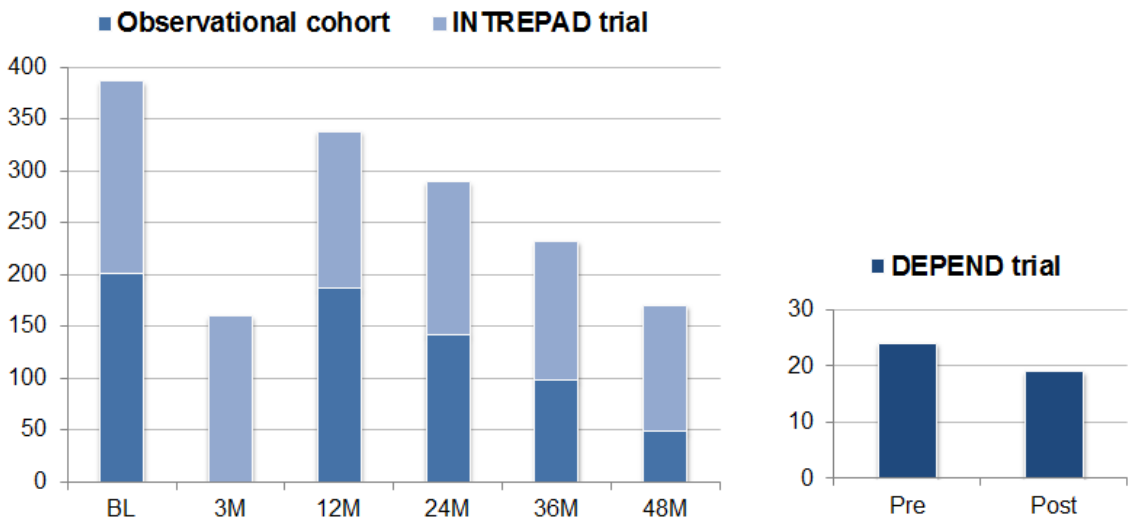

Figure 5: Progression of longitudinal follow-ups. *Observational cohort and INTREPAD trial are added together inasmuch as the participants underwent the same set of evaluations. Note that the 3 months follow-up (3M) was performed only for INTREPAD trial participants. DEPEND trial visits were undertaken outside the regular annual follow-up schedule.*

4. PREVENT-AD biomarkers

The PREVENT-AD research group measured not only classical AD biomarkers, but also emergent potential indicators of AD progression described below.

a. Cognition

Neuropsychological performance was measured using the RBANS,[26] which evaluates 5 cognitive domains: immediate memory, delayed memory, attention, language and visuospatial abilities. A global cognition score summarizes all these domains. This test was designed specifically to provide sensitive detection of cognitive decline in persons whose cognitive status is still within normal limits. It is therefore used frequently in prevention trials or studies of cognitively frail (but still “normal”) elderly. Its ~30-minute battery is available in both French and English in 4 equivalent versions to reduce practice effects in longitudinal assessment. Trained research assistants administered testing, and scores were calculated by a single PhD neuropsychologist. We developed correction factors

to improve version equivalence among the 4 French versions.[27] Apart from this adjustment, we used 'research' (not "corrected" for age) scores instead of published age 'norms' that are customary in clinical use. For more details about "age-corrected" scores, see 'Code Availability'.

At the same visit, we administered the structured Alzheimer Dementia 8 (AD8) interview to the study partner, who rated the participant on eight functional abilities intended to discriminate normal cognitive aging from very mild dementia. The AD8 was designed specifically for the detection of change over time.

Beginning in 2015, each participant was also asked to rate subjective change in memory abilities using the *Measurement of Everyday Cognition (ECog).* This instrument uses a four-point scale to ascertain perceived changes over the past year. Although administered annually, these ratings did not necessarily coincide with annual FU visits.

b. Cerebrospinal Fluid (CSF) proteins

Lumbar puncture (LP) was performed by a neurologist (PR-N) in a procedure that typically lasted less than 15 minutes. A large-bore introducer was inserted at the L3-L4 or L4-L5 intervertebral space, after which the atraumatic Sprotte 24 ga. spinal needle was inserted to puncture the dura. Up to 30 ml of CSF were withdrawn in 5.0 ml polypropylene syringes. These samples were centrifuged at room temperature for 10 minutes at ~2000g, and then aliquoted in 0.5 ml polypropylene cryotubes, and quick-frozen at -80°C for long-term storage. A video describing the LP procedure developed at the StoP-AD Centre is available at https://www.youtube.com/watch?v=9kckrlBIR2E. LPs were optional in INTREPAD, but required for the DEPEND proof-of-concept trial. They were first performed on PREVENT-AD participants in the observational cohort in 2016 and, effective in 2017, became a requirement for enrollment of new participants.

Biomarkers for amyloid, tau and neurodegeneration were analyzed in CSF samples. Typically, levels of Amyloid-beta 1-42 ($A\beta_{1-42}$), total tau (t-*Tau*) and phosphorylated tau ($_{181}$p-*Tau*) were determined by enzyme-linked immunosorbent assay (ELISA) using Innotest technology (Fujirebio) following the

European BIOMARK-APD standardized protocol.[28] Other novel or emergent CSF (and plasma) biomarkers were also examined but are not yet available for sharing, pending validation. These markers include inflammation, neuronal and vascular biomarkers, albumin and other amyloid-beta species such as $A\beta_{1-34}$. In the clinical trials, CSF concentrations of study drug was assayed using liquid chromatography and mass spectrometry for pharmacokinetic studies.[51]

To date, a total of 565 LPs have been performed. About 56% of INTREPAD participants (110 out of the 195 enrolled) initially agreed to undergo a series of four LPs. Seventy-five completed this series, and 45 participants have continued in the serial LP protocol to undergo a 5th LP and 30 a 6th LP. Among the observational cohort, 72 participants underwent at least one LP. Additionally, 24 pre- and 19 post-treatment LPs were performed as part of the DEPEND trial.

The success rate of CSF collection was 98%. The major post-procedure complication was headache, mostly transient and mild, observed in 26% of cases. Typical, more significant mild post-dural puncture headaches (PDPH) occurred following 4% of LPs, a figure comparable with the literature.[29] Another frequent discomfort was nerve irritation, a sensation that can be felt in the legs and feet, when the needle was touching fine nerves of the *cauda equina*. This sensation was typically transient as the physician rotated the needle to relieve any discomfort reported (depicted in the LP video presented above). Although it was available if needed, none of the volunteers required a blood-patch or other interventions to relieve post-LP headache or other adverse event.

c. Neuroimaging

To detect longitudinal changes, participants were scanned on a Siemens TIM Trio 3 Tesla Magnetic Resonance Imaging (MRI) scanner using a Siemens standard 12 or 32-channel coil, as specified by protocol (Siemens Medical Solutions, Erlangen, Germany). Eighteen claustrophobic participants were excluded during the recruitment process. The duration of MRI sessions varied by experimental protocol but included up to six functional acquisitions and four

structural modalities. See Table 4 for parameters and Figure 6-9 for acquisition protocols.

Table 4: PREVENT-AD MRI parameters

Modality	TR (ms)	TE (ms)	TI (ms)	∝	In-plane resolution (mm)	Slice thickness (mm)	Matrix Size	Number of volumes
Structural modalities								
t1w	2300	30	-	9°	1 x 1	1	256 x 256	-
FLAIR	5000	388	1800	-	1 x 1	1	256 x 256	-
T2*	650	20	-	20°	0.8 x 0.8	2	256 x 256	-
qT2*	44	2.84-39.8	-	-	1 x 1	1	192 x 192	-
t2w	2500	198	-	-	0.64 x 0.64	0.64	320 x 320	-
MP2RAGE	5000	2.91	700/2500	4/5°	1 x 1	1	256 x 256	-
dwi65	9300	92	-	-	2 x 2	2	96 x 96	65
Functional modalities								
single-echo ASL	4000	10	-	90°	4 x 4	7	64 x 64	80
bold	2000	30	-	90°	4 x 4	4	64 x 64	150
encoding-BOLD/ retrieval-BOLD	2000	30	-	90°	4 x 4	4	64 x 64	183/453
multi-echo ASL	4120	8.4	30	90°	4.5 x 4.5	5	64 x 64	264

TR = Repetition Time; TE = Echo Time; TI = Inversion Time; ∝ = Flip Angle; t1w = MPRAGE (Magnetization Prepared Rapid Acquisition Gradient Echo); FLAIR = FLuid Attenuated Inversion Recovery; T2star = GRadient Echo T2*; qT2star = 12-Echo T2* (2.84, 6.2, 9.56, 12.92, 16.28, 19.64, 23, 26.36, 29.72, 33.08, 36.44 and 39.8 ms); dwi65 = Diffusion Weighted Imaging with 65 directions. The 3 protocols are acquired twice, in the A-P and P-A phase encoding directions, to compensate for image distortions post-processing; ASL = Pseudo-Continuous Arterial Spin Labeling; BOLD = Resting State Blood Oxygen Level Determination.

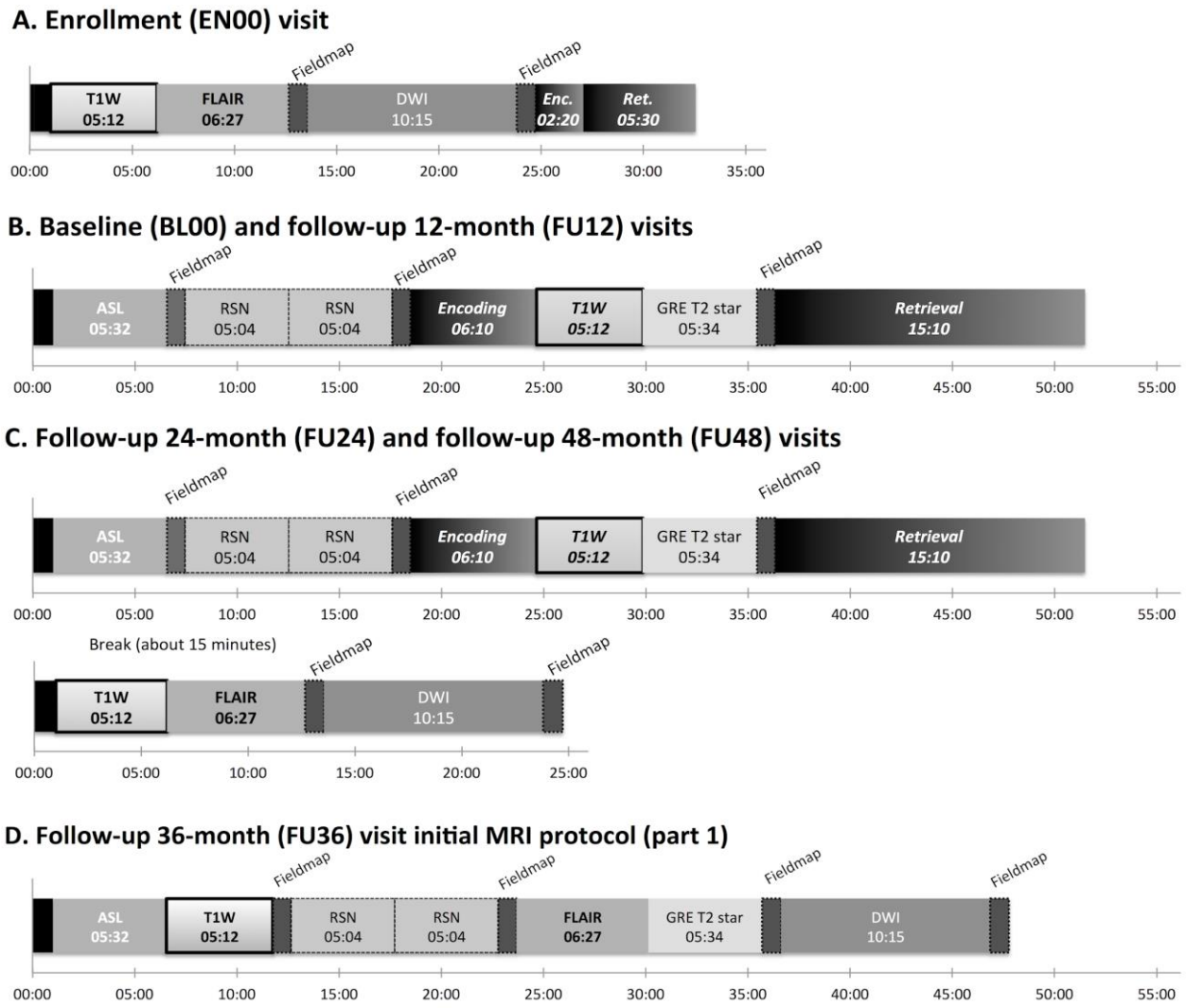

Figure 6: Observational Cohort 12-channel coil MRI protocol.

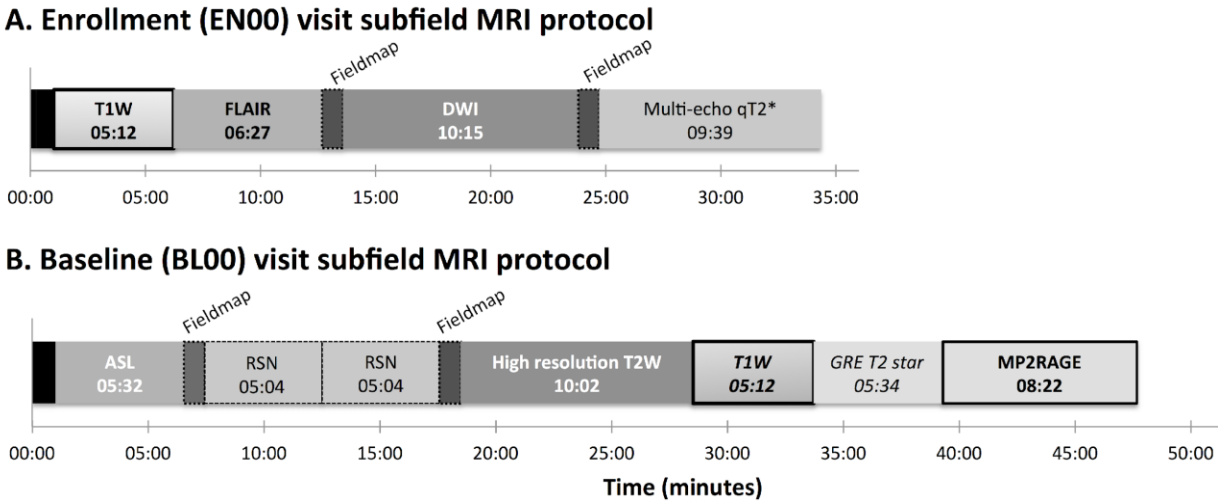

Figure 7: Observational Cohort 32-channel coil MRI protocol (started in June 2016, for new enrollees only)

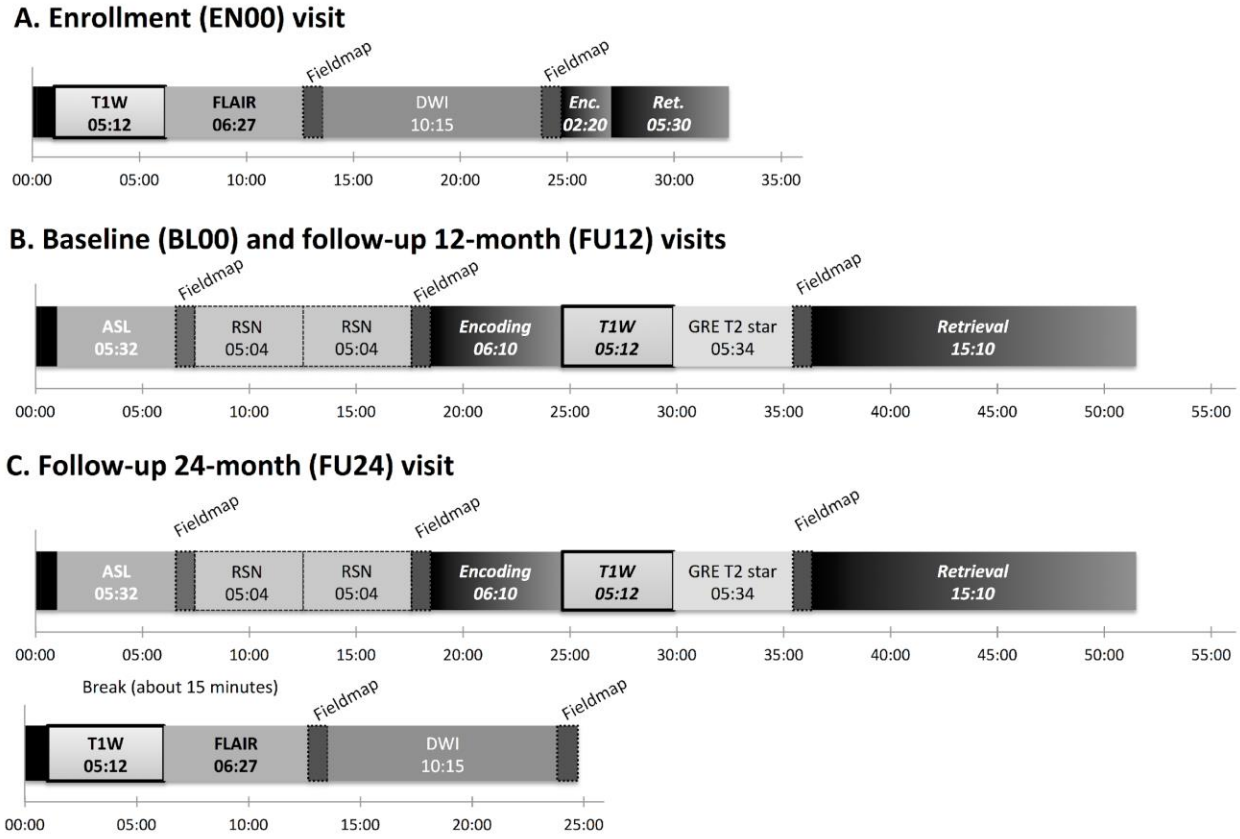

Figure 8: INTREPAD Trial MRI protocol (12-channel coil).

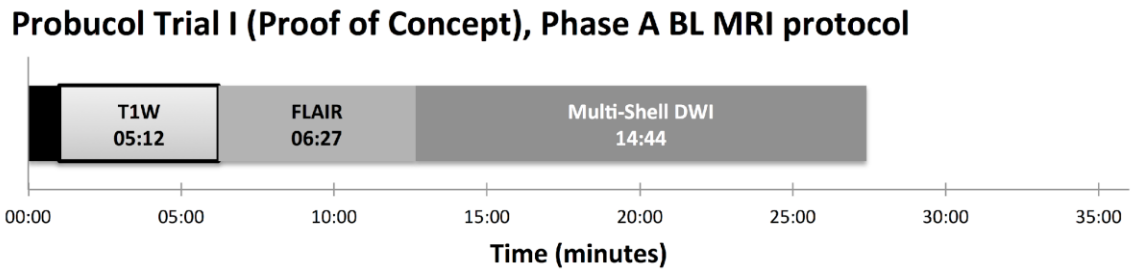

Figure 9: Probucol Proof of Concept MRI protocol (32-channel coil).

Because heart rate and respiration correlate with functional Blood Oxygen Level Dependent (BOLD) signal, physiological monitoring was performed during all functional acquisitions, allowing physiological correction of the fMRI data (heart rate and breathing).[30-32]

During these functional acquisitions, heart rate, chest wall motion and a logic pulse (marking the time of each fMRI volume) were recorded using a BIOPAC MP150 system at a sampling rate of 400Hz (BIOPAC Sytems, Inc., Goleta, CA). Chest wall motion was monitored using a respiratory belt transducer (TSD201) connected to a RSP100C Respiration Amplifier module, part of the BIOPAC system, while EKG traces were recorded using the ECG100C amplifier of the BIOPAC to monitor participant's heart rate.

Episodic memory task fMRI

An episodic memory task for object-location associations was performed by participants longitudinally. The study design is similar to that published previously.[33,34] Participants were scanned as they encoded an object and it's left/right spatial location on the screen. Forty-eight encoding stimuli were presented one at a time for 2000 msec with a variable inter-trial interval (ITI). A twenty minute break followed encoding, during which time structural MRIs were acquired. After this break participants were presented with the associative retrieval task in which they were presented with 96 objects (48 "old"-previously encoded objects; 48 "new" objects) and were asked to make a forced-choice between four-alternative answers: i) *"The object is FAMILIAR but you don't*

remember the location"; ii) *"You remember the object and it was previously on the LEFT"*; iii) *"You remember the object and it was previously on the RIGHT"*; and iv) *"The object is NEW"*. The E-Prime program was used to run the experimental protocol and collect behavioural data (Psychology Software Tools Inc., Pittsburgh, PA, USA). Text files were exported from the E-Prime software and are shared within the open PREVENT-AD dataset. Details are provided in the Data Record section. Different stimuli were employed during each testing time allowing longitudinal data collection.

During the episodic memory task fMRI acquisitions, the visual stimuli were generated by a PC laptop computer and projected by a LCD projector onto a screen visible to participants *via* a mirror mounted within the standard head coil. Plastic optical corrective glasses were provided for participants who required correction for visual acuity. Participants used a fibre optic four-button response box to perform the experimental tasks.

d. Neuro-sensory markers

Abilities in odor identification (OI) were tested in a 30-minute session in a well-ventilated room, using the standardized University of Pennsylvania Smell Identification Test (UPSIT).[35] This test uses "scratch-and-sniff" stimuli of 40 items (4 randomized booklets of 10 odorants each). Although the test can be self-administered, a trained examiner administered the test to improve reliability given that we would use the results in clinical trials. "Both francophone and anglophone participants were presented with odors from the US version of the UPSIT. The francophone test used an in-house French translation. In a leave-one-out analysis, we assessed the reliability of the UPSIT in the PREVENT-AD cohort among an initial sample of 159 participants, obtaining a Cronbach α of 0.821, which suggests high internal consistency.[36,37]

We tested central auditory processing (CAP) using the Synthetic Sentence Identification with Ipsilateral Competing Message (SSI-ICM) test and the Dichotic Stimulus Identification (DSI) test. After having first been assessed for simple

auditory acuity (with monosyllabic words), participants were asked to identify spoken "pseudo-sentences," either with various sound levels of a distracting background narrative (SSI-ICM) or with dichotic binaural presentation (DSI). The latter test was available only in French. CAP testing was introduced in 2014.[38,39]

In the SSI-ICM test, one pseudo-sentence is heard while a story is recited in the background. Both the sentence and story are played in the same (ipsilateral) ear. The participant is asked to identify the target sentence among 10 choices offered. Participants performed this task a minimum of 10 and a maximum of 30 times, with designated score-dependent stopping points.[40] The other ear was then tested using the same protocol. SSI-ICM testing can typically be completed in less than 30 minutes.

The DSI task tests dichotic listening capability. For this task different pseudo-sentences are played simultaneously in the two ears. Participants a asked to identify the two target sentences from a list of 10. Participants performed this task a minimum of 5 and a maximum of 10 times, with designated score-dependent stopping points,[40] in a session requiring less than 15 minutes.

e. Alzheimer progression score

Although it is a derived variable, we describe here the nature of the composite Alzheimer Progression Score (APS) used as the primary outcome measure in INTREPAD. The scores are available within the INTREPAD shared dataset. The APS was developed in conjunction with colleagues at Johns Hopkins University. A composite such as the APS was envisioned at the time INTREPAD was designed, but its development and validation relied on data from parallel assessments in the longitudinal observational (non-trial) cohort. The APS is based on an Item Response Theory latent-variable approach to the many potentially informative data points collected longitudinally in PREVENT-AD. It relies on an assumption that informative changes in any AD marker arise from a single underlying latent process, *viz.,* AD pathogenesis.[17] The construct validity of the APS approach was first demonstrated in the BIOCARD study, in which a version that incorporated data from that study showed ability to conjoin four

different imaging and CSF markers as predictors of subsequent change in clinical diagnosis to MCI or AD dementia.[17] We used non-trial PREVENT-AD participant data both to estimate parameters for the APS scoring algorithm (using baseline data) and then to demonstrate measurement invariance (*i.e.,* showing that the same set of parametric "weights" estimated at baseline served well to estimate disease progression at later time points). We further demonstrated "portability" of the score to the trial cohort by comparing the performance of observational cohort-derived "weighting" parameters performance using "weights" derived *de novo* in trial participants. Variables for inclusion were selected if their longitudinal measure of change more than offset item variance (hence, they were likely to contribute positively to statistical power of the method to detect change). Included measures were several cognitive test results, neurosensory abilities (olfactory identification), total brain volume, grey matter cortical thickness and density, cerebral blood-flow and CSF biomarkers. Missing data were accommodated in an "averaging over" of interpolating data points assessed by a subsequent iterative validation of the resulting values. The value of the composite APS is evidenced by the fact that this score appears to provide more information on subtle brain changes than any one of its constituent biomarkers.[18]

5. Data Management & Open Science Plans

Management, quality control (QC), validation and distribution of PREVENT-AD data were performed in LORIS, a system designed for linking heterogeneous data (e.g. behavioral, clinical, imaging, genomic) within a longitudinal context.[12] Numerous LORIS modules were used to facilitate the curation process, including the Participant Status, Family Information, Family History, Acknowledgements, Document Repository, Drug Compliance and Data Release modules (Table 5). In addition, behavioral forms included customized algorithms developed for aggregating various pieces of data in a user-friendly manner. Data selection and dissemination was done through the Data Query Tool, or via specialized scripts that prepared large amounts of data via releases in spreadsheet-ready formats.[12]

Table 5: LORIS modules

LORIS Modules	**Brief Description**
Participant Status	Overall status categories “Active”, “Stop Medication Active”, “Ineligible”, “Withdrawn”, “Excluded”, etc for each participant and their respective statuses for the individual drug trials.
Family Information	Identifying and characterizing relatives of participants.
Family History	Family history of clinical and memory problems.
Acknowledgements	Reference for study collaborators that dynamically generates a publication acknowledgement list.
Document Repository	Centralized location for managing study documents (original forms, publications, manual of operations, etc)
Drug Compliance	Compliance rate for drug trials for each participant.
Data Release	Distribute and manage permissions for packaged datasets for analysis.

To ensure consistency in research publications, presentations and reports based on the PREVENT-AD dataset, a cycle of internal releases was implemented, typically once a year, from 2012 to 2017. A total of 5 data releases were prepared and shared with collaborators via the Data Release Module in LORIS (DR 1.0, 2.0, 3.0, 4.0 and 5.0). For each such data release, a “freeze” was placed on all data as of a specified date. Between the time of the “freeze” and eventual release, data were edited, validated and cleaned. While data were continuously queryable in the Data Query Tool, data releases were also tagged as a series of spreadsheets (including a data dictionary for all instrument variables), archived imaging datasets, and a document summarizing the project and available variables. To gain access to PREVENT-AD data releases, collaborators were required to sign a Data Use Agreement that included a publication policy.

6. Development of Open Science data sharing

As an initial instance of Open Science data sharing, we released a subset of the INTREPAD clinical trial data (resting-state scans) in 2015 to the Consortium for Reliability and Reproducibility (CoRR, sample called UM1).[41] That subset included 80 participants enrolled in the trial having two available time points (baseline and 3-month follow-up). Each time point included two resting-state fMRI acquisitions and a defaced t1-weighted structural image.

To expand our open science initiative, acquisition and preparation of data in a structured and standardized fashion were key elements. The numerous steps required to prepare the dataset before its dissemination on an open science platform included: agreement with collaborating investigators, ethical considerations (discussed below), data selection, completion of data entry/analysis, quality control and dataset documentation.

7. Open Science Ethics

Making the PREVENT-AD dataset available for open science was contingent on several ethical considerations. First, a number of points in the original PREVENT-AD participant information consent forms and ethics approval were incompatible with broad open science sharing plans. These included: 1) the description of how program data would only be shared within the collaboration or with external researchers on a case-by-case basis; and 2) a commitment to retain sole ownership of all research results. It was therefore clear that important aspects of sharing data for open science, including its risks, required explanation and consent for all research participants, most of whom had remained actively involved in StoP-AD Centre activities. Full ethics approval was obtained for the open science plans and re-consent process.

To date, 310 out of 330 participants still active in the program have agreed to share their data for open science. Fifteen participants preferred not to participate

in this type of data sharing, while we are still waiting for the answer of five individuals to complete the re-consent process of active participants. As a follow-up to the re-consent process, for those participants not willing to participate in open science, we sought to better understand both the reasons for this choice as well as whether additional measures could be taken that might affect participants' decision or perception of the risks. The reasons expressed included: 1) general concerns about the security, confidentiality, and anonymity of data (n = 4); 2) explicit concerns about the potential consequences for themselves or family members with regards to the genetic information in the study and its potential misuse by insurers and pharmaceutical companies (n = 3); and 3) a broad lack of trust in researchers (n = 2). Another two persons refused to participate in the open science initiative because they insisted that they should have access to the study data themselves, either directly or through their doctor. We have not formulated an adequate response for these individuals who would not necessarily participate in the open science plans, and list this reason separately from the other reasons of refusal (described above) involving appropriate re-use, data security, or confidentiality, all of which were items specifically anticipated by our ethics considerations. Several other participants either declined to comment (n = 2) or were unspecific about their reasons for not participating in our open science initiative (n = 2).

Second, even when partially de-identified data had been prepared for sharing with collaborating research teams, additional dataset de-identification steps were required to share data with a much larger community of researchers. All PREVENT-AD participant names had already been assigned an internal study code. These codes were now assigned a new "public" alphanumeric code, to which the participant's identity cannot directly linked, with the sole exception of an ability to do this retained exclusively by the StoP-AD team. All brain images to be shared were "scrubbed" to remove all potentially identifying fields from their header (e.g., date of birth or acquisition dates) and structural modalities were defaced to prevent re-identification using 3D rendering of the face.[42] For ethnicity

data, only large categories remained for sharing (e.g., Caucasian, Asian). Neither *date of birth* nor *dates of visits* are available, but age (in months) at each time point, is shared. Postal code information is not shared, and company names were removed from information on participants' profession.

When accessing the 'open' data, researchers agree to a standard set of good data use practices, such as meeting ethics requirements and keeping the data secure. To further reduce the potential risk of data misuse and protect participants' privacy, a large part of the open science dataset is shared through a Registered Data Access model, whereby data are only available to *bona fide* researchers and clinical care professionals[43]. PREVENT-AD data shared through Registered Access includes, for example, cognitive tests results, genotyping data, medical information, family history of AD, and certain demographic data such as ethnicity and profession/occupation (see Table 7). Furthermore, some of these sensitive data may be shared with additional precautions, following an 'on-request' procedure. Permitted uses of data are communicated using Consent Codes, a structured way of presenting consent-based permissions and restrictions on the use of Open Science data. [44,45] In our case, PREVENT-AD data must be used for neuroscience research as stipulated in the consent forms and in the terms of use.

7. Ongoing and future efforts.

AD research is at last regarded as a critical priority for Canada and the world.[4-6] Accordingly, it is now essential to collect data from large cohorts, and subsequently to make such data available to the global research community. Government initiatives at a national level, like the Canadian Consortium of Neurodegeneration in Aging were implemented to advance and concentrate research and development towards prevention of Alzheimer's disease dementia and other neurodegenerative disorders.[7] These initiatives represent Canada's efforts toward the development of data-rich consortia, like the Alzheimer's

Disease Neuroimaging Initiative (ADNI) in the United States. These consortia proved to be especially fruitful in aggregating and sharing data from multiple research centres. As an example, the (US) ADNI listed 1488 publications in PubMed as of 20 February 2020. Although initiated much more recently, the shared PREVENT-AD dataset has already become a compelling resource for the neuroscience community, with 75 downloads of the open dataset so far, and will undoubtedly increase the rate of scientific discovery in dementia research.

As an encouragement to other projects contemplating the conversion of their data to open science access, we note that this transition in PREVENT-AD required substantial resource availability but was relatively smoothly achieved over the course of approximately six months. In that time, enormous efforts were required to obtain additional ethics review and a re-consent process for all participants whose data are to be shared. Additional data preparation and "cleaning" was also required. Although costly, we expect these efforts and resources to yield much incremental value to the project in years to come.

The STOP-AD Centre continues to collect data. In time, additional data-gathering modalities along with an expanded set of longitudinal observations in PREVENT-AD will become available to the greater research community. Newer methods will include other neuroimaging techniques, such as positron emission tomography (PET), magnetoencephalography (MEG), as well as genomic information from large-scale methods, such as GWAS. Additional information will also become available on lifestyle, comparison groups data, and data from participants who develop mild cognitive impairment (MCI). To further complement the existing cohort, novel sensitive blood-based biomarker assays are being developed to bypass the CSF (or PET) requirement now needed to monitor disease progression in both the pre-symptomatic and symptomatic phases of AD. Given that some of these new acquisitions will differ from pre-existing PREVENT-AD modalities, they may yield less longitudinal information, but will nonetheless enhance the information value of the evolving data resource. With these new

data continually being acquired in this "at-risk" population, PREVENT-AD is poised to become a marquee study in dementia research, akin to other data sharing initiatives such as the Alzheimer's Disease Neuroimaging Initiative (ADNI). The emerging field of machine learning and deep learning methods are bound to change the analytic landscape in the years to come and as PREVENT-AD will release more data, our dissemination methods will have to adapt to better reflect this new reality.

Code Availability:

Generation of neuropsychological dataset: RBANS specificities

The RBANS French versions are known not to be fully equivalent. The author of the RBANS (Dr. Christopher Randolph) recommended systematic correction of +4 for the semantic fluency section of version B. Our group compared performance at the 3M visit with BL scores in the INTREPAD trial, assuming no treatment effects and comparable abilities at the two timepoints. It was determined that additional corrections were needed to control version differences. Suspecting part of this problem could be traced back to the non-equivalence of English and French tests, we developed adjustment factors that brought the several versions into approximate equivalence. This procedure is described in Lafaille-Magnan et al., 2018.[27] The script used is as followed https://github.com/marieleyse/RBANS-correction.

Whereas clinical testing may call for scoring criteria that vary by age to compare individual performance to "normative" data, we avoided correcting for age in scoring the RBANS for research purposes. We scored all participants using norms for individuals aged 60—69 years. This method revealed an actual decline in performance with age, whether or not this decline was related to disease. Both scores (clinical; adjusted for age and research; using 60-69 norms) are available.

Generation of episodic memory task fMRI dataset: software and stimuli

Psychology Software Tools E-Prime (version 2) were used to design the experiment, collect the data and perform analysis. Data were saved in.edat2 format readable by the program only. Data were also saved as text files to facilitate data sharing (https://pstnet.com/products/e-prime-legacy-versions/).

For de-identification before data sharing, the text files containing the behavioral task fMRI related data were scrubbed to remove dates and PREVENT-AD study ID using a script available on Github: https://github.com/cmadjar/Loris-MRI/blob/open_preventad_v20.1.0/tools/scrub_and_relabel_task_events.pl.

Images used for the task were from a bank of standardized stimuli.[46,47]

Generation of de-identified MR images for data sharing:

Anatomical images (T1W, T2W, FLAIR, MP2RAGE, quantitative T2*, GRE T2*, fieldmap magnitude files) were defaced using the defacing algorithm developed by Fonov and Collins (2018) that were shown to not significantly affect data processing outcome.[42] A slight modification of the code was done so that it could be integrated into the LORIS platform. The version of the script used to run the defacing on the PREVENT-AD datasets is available in Github (https://github.com/cmadjar/Loris-MRI/blob/open_preventad_v20.1.0/uploadNeuroDB/bin/deface_minipipe.pl). Identifying fields (such as PREVENT-AD participant's ID, date of birth, date of MRI, etc) were scrubbed from the DICOM headers using the DICOM Anonymization Tool (DICAT; https://github.com/aces/DICAT).

DATA RECORDS

Available data

Since April 2019, MRI raw data and basic demographics (age at MRI, sex, language, handedness) are available for sharing using the LORIS platform (https://openpreventad.loris.ca).

Users are able to access the OPEN data listed in Table 6, for a group of participants who had agreed to data sharing at the time of the release (n=232 as of April 2019). The URL provided leads to the PREVENT-AD OPEN LORIS instance, since the CONP portal was not entirely functional at the time of the release. More data from these 232 participants will be accessible via a Registered Access model and will be eventually released (Table 7). Other waves of data sharing with increased number of participants and other type of data are also planned to enrich this important research resource.

TABLE 6: PREVENT-AD OPEN data (data from observational cohort and INTREPAD trial), as of April 29th, 2017 (n=232).

Data type	Level of access	Data files / format / notes	URL
MRI / fMRI -Anatomical -Diffusion -task fMRI -resting state fMRI -ASL	OPEN	MR images: .minc task fMRI behavioral data: .txt notes: -QC data are available for anatomical and diffusion imaging **Data related to subjects with incidental findings can only be shared via Registered access**	https://openpreventad.loris.ca/
Basic demographics -Age at MRI -Gender -Study Language -Handedness	OPEN	CSV files notes: -Age at MRI : longitudinal data -Other data: EL / EN	https://openpreventad.loris.ca/

Table 7: PREVENT-AD data available for sharing in a REGISTERED access mode. Note that few data marked with an asterix (*) are only available 'on request'.

Data type	Level of access	Data files / format / notes
Neuropsychological tests · MoCA · RBANS · AD8	REGISTERED	CSV files notes: -MoCA and CDR: EL -RBANS and AD8: longitudinal
CSF proteins* · amyloid * · total tau* · phosphorylated tau*	*on request only	CSV files notes: -longitudinal LPs (up to 6 LP)

Neuro-Sensory tests · smell identification test · central auditory processing (CAP)	REGISTERED	CSV files notes: - olfaction: longitudinal - CAP: longitudinal, but after 2014
Genetics · BDNF · BCHEK · APOE*	REGISTERED *on request for APOE	CSV files notes: -analysis done with EL blood sample
Clinical Data · Blood Pressure / Pulse · CAIDE score · Blood laboratory results - Level of education - Profession / Retirement - Ethnicity - Height / Weight - Pharmacological Profile	REGISTERED	CSV files
INTREPAD related data · Treatment allocation · APS · AE – SAE*	REGISTERED *on request	CSV files

MoCA: Montreal Cognitive assessment, RBANS: repeatable battery for the assessment of neuropsychological status, CAIDE score: cardiovascular risk factor, APS: Alzheimer progression score, AE-SAE: Adverse Event – Serious Adverse Event

Published manuscripts using this dataset

Major findings using PREVENT-AD data form data releases (1.0 to 5.0) can be found in 22 published articles and in more than 75 abstracts (http://prevent-alzheimer.net/). In brief, several analyses demonstrated novel association, correlation or prediction among various direct and derived measures of AD pathology, including MRI, CSF biomarkers of AD, protein mediators of innate immune activity, and neurosensory faculties.[18,36,38,48-50] In addition to the INTREPAD trial derived results publications associations were established between measures of AD pathology (revealed by MRI and/or CSF proteins) and subjective cognitive decline, and proximity to age at onset of parental AD symptoms.[18,51-54] Novel MRI techniques and disease progression modelling were

also validated in this dataset.[17,41,55]. Articles looking at the association between PET and CSF amyloid and tau and looking at the relation with vascular risk factors were recently published (Add Melissa and Theresa Here). Several additional papers are in preparation.

We expect that sharing the PREVENT-AD data with the larger community under the open science principle will result in many additional publications in the coming years.

Additional information about the PREVENT-AD program can be found at http://prevent-alzheimer.net/.

TECHNICAL VALIDATION

Homogeneity of procedure: We decided to acquire data for this study at a single site to limit the need for harmonization. Other elements were part of the protocol with the intention to reduce the risk of external variability: i) cognitive testing was scored by a single neuropsychologist and the final scores were automatically computed by LORIS; ii) MRI acquisitions were performed on the same scanner and quality controlled by the same individual; iii) lumbar punctures were performed by the same physician (more than 95% of the time) following an internationally accepted, published protocol; iv) clinical information was reviewed by the same two physicians; v) clinical evaluation was performed by the same team of nurses and psychometrists.

Data entry: Data was entered in LORIS in duplicate. The 'conflict resolvers' feature of LORIS allowed detection of discrepancies between two entries of the same information and systematic corrections of mistakes by the data entry personnel. In case of doubt, the proximity of the team of nurses and psychometrists facilitated the process of information verification from the source documentation, at the time of data entry. Before data releases, the clinical team, headed by the study physicians and the research coordinator, reviewed special cases and determined their potential for data analysis by a pass or fail decision. Any failed cases were then flagged in LORIS and made unavailable for analysis.

Quality control (QC): LORIS has several internal QC in place to avoid missing data in required fields, out of range values, etc. Weekly, 26 automatic checkpoints were run to detect any abnormalities. At every cycle of data freeze and release, the data entry personnel completed data entry and performed data verification via additional automatic checkpoints. MRI acquisitions were harmonized across sessions by setting up saving each session's protocol directly on the scanner console so that the same protocol could be used for a given MRI visit. In addition, acquisition parameters were automatically checked to ensure

acquisition harmonization at the time of insertion into LORIS. Finally, manual quality control was performed on all structural modalities of the original dataset by the same individual and the result of that quality assessment was imported into the open dataset along with the shared images. QC status, predefined comments and text comments were saved directly in LORIS at the time of QC.

Additional QC for data sharing on CONP: 1 out of 10 research files were manually reviewed to compare the source documents with data entered in LORIS. We choose to review data acquired in 2013 and 2017 and reviewed all data from the following 'instruments': RBANS tests, smell identification tests, laboratory blood results and central auditory processing tests (at BL and all the annual FU). We selected handedness, family history of AD, medical and surgical history and laboratory blood results to review at eligibility and enrolment visit. Out of 140 instruments reviewed, only 5 mistakes were found and corrected (wrong visit date, repetition not entered for RBANS test (3) and 1 information missing in the family history of AD). We added 5 more automated QC checks to make sure no names were part of the dataset, no duplicated visit label were created, no data entry 'in progress' were pending and checking for '0' values in laboratory results. After the de-identification process of the anatomical MR images, every single image was visually reviewed to insure proper defacing. Any problem with the defacing process was detected, fixed and reviewed. A final review to detect the presence of dates and any potentially identifiable information was also performed in the whole shared dataset.

Uniformity and comparability of data collection methods and analysis: The aforementioned attention to design and quality control contributed to the reproducibility of findings by different investigators (thus avoiding concerns about reproducibility).[56,57] Similarly, rigorous methods were employed to ensure proper data curation.[58] Overarching strategies such as the FAIR data principles are the current guideposts to success in this era of Big Data, as data sharing becomes a necessity.[14] Leveraging this data to create larger pools of data for conjoint analysis is also becoming the norm, but, this requires proper documentation, full provenance and reliable organization. The most thoroughly documented and harmonized data sets are of little use without ease of access to these resources. Thus user-friendly data systems like LORIS also provide a substantial advantage for such data-sharing efforts.

USAGE NOTES

For reuse of the PREVENT-AD data, users must respect PREVENT-AD terms and publication policy. They can be found when requesting an account on https://openpreventad.loris.ca/. In brief, data must be used for neuroscience purposes, and users must commit to properly cite the dataset and to follow good data use practices (e.g. not attempt to re-identify participants). The PREVENT-AD research group request that anyone using PREVENT-AD data for publication state in the methods section of their publication that PREVENT-AD is the (or one of the) source of data, and cite 2 PREVENT-AD scientific papers (including this one).[8]

For reuse of the PREVENT-AD data, we suggest that researchers carefully read and understand the context of the data collection described in this paper and in the documentation available with the data. For re-analysis of the INTREPAD trial data, please refer to our results paper recently published in Neurology and contact Dr. John Breitner for possible collaboration and further details about the study design[18].

Label convention used in the PREVENT-AD dataset is presented in Table 8. Data collected at a specific time point (under a specific visit label) are regrouped within instruments containing numerous variables (ex.: data for total RBANS scores, 12 months after baseline in the observational cohort, will be associated with PREFU12 under the instrument named 'RBANS').

Table 8: PREVENT-AD data label convention

PROJECT	VISIT TYPE	TIMEPOINT (in months from BL visit)
NAP: INTREPAD trial* PRE: Observational cohort *treatment allocation (naproxen vs placebo) is provided via the registered access only	EL : Eligibility visit EN : Enrolment visit BL : Baseline visit FU : Follow-up LP : lumbar puncture AP : Auditory Processing	00 03 12 24 26 48

IMPORTANT NOTES about the label convention: The first visit in the program is always labelled as PREEL00. INTREPAD trial participants are identified at enrolment visit (NAPEN00) and the following (ex.: NAPBL00, NAPFU03, etc). Even after the termination of the treatment and trial protocol (24 months; FU24), INTREPAD participants remain named as NAP for all the following annual FU (NAPFU36, NAPFU48, for example). However, if a participant was not able to follow study protocol until NAPFU03, he/she was excluded from the trial and 'switched back' to the observational cohort (PRE) to continue to be followed annually (PREFU12 and following). Central auditory processing (AP) also has its own label, even if this test was performed in concordance with BL and or FU. Lumbar puncture (LP) stand alone as the procedure was done on a separate day, close to the annual FU.

ACKNOWLEDGMENTS

PREVENT-AD was launched in 2011 as a $13.5 million, 7-year public-private partnership using funds provided by McGill University, the Fonds de Recherche du Québec – Santé (FRQ-S), an unrestricted research grant from Pfizer Canada, the J.L. Levesque Foundation, the Lemaire Foundation, the Douglas Hospital Research Centre and Foundation, the Government of Canada, and the Canada Fund for Innovation. Private sector contributions are facilitated by the Development Office of the McGill University Faculty of Medicine and by the Douglas Hospital Research Centre Foundation (http://www.douglas.qc.ca/). CONP was launched in 2018 and is funded, in part, by Brain Canada. Thanks to all the financial resources.

PREVENT-AD is the result of efforts of many other co-investigators from a range of academic institutions and private corporations, as well as an extraordinarily dedicated and talented clinical and technical assistant staff, students, and post-doctoral fellows. Here is listed the entire PREVENT-AD Research Group: https://preventad.loris.ca/acknowledgements/acknowledgements.php?date=[2017-12-01]. The authors thank David Fontaine, PhD, who oversaw cognitive testing at enrollment, scored the RBANS cognitive evaluations, and administered additional cognitive batteries when indicated. In Dr. Judes Poirier's laboratory we wish to mention the good work that has been carried out by Anne Labonté, Doris Dea, Louise Théroux, and Cynthia Picard for CSF biomarker analyses, genotyping and more. Melissa Appleby, Laura Mahar, Miranda Tuwaig, Marie-Elyse Lafaille-Magnan, Christina Kazazian, Tanya Lee, Galina Pogossova, Renuka Giles, and Karen Wan collected cognitive and neurosensory data, assisted to the MRI sessions, entered data in LORIS and worked hard to obtain a

high-quality dataset, thanks to them. Drs. Tharick Ali Pascoal, Marina Tedeshi Dauar, and Laksanun Cheewakriengkrai, dedicated time and energy to assist during LPs and performed neurological assessments with the research participants. Special thanks go to Marianne Dufour, administrative assistant, and to the clinical team; Ginette Mayrand, Joanne Frenette, Valérie Gervais, Isabelle Vallée, Rana El-Khoury, Leslie-Ann Daoust and Fabiola Ferdinand, all nurses who met with participants and were devoted to our participant's well-being. Not to forget, everyone working on data analysis, including Jeannie-Marie Leoutsakos for the development of the APS and Melissa Savard, the PREVENT-AD data analyst. We thank Benoit Jutras, PhD, from *Université de Montreal*, for gifting us equipment to test central auditory processing. The figures in this manuscript have been generously created by Derek Lo. We give an additional thanks to the LORIS team, at the Montreal Neurological Institute, who recently accelerated our involvement towards boarder data sharing, for the benefit of the community. The authors acknowledge the generosity and commitment of all research participants who volunteers for this work and put their trusts and hope in this research program.

AUTHORS CONTRIBUTIONS:

Jennifer Tremblay-Mercier	Supervision of data collection efforts and the clinical team. Participating in Data Sharing policies. Drafting and revising the manuscript.
Cécile Madjar	Data Management, conceptualization of database. Drafting and revising the manuscript.
Samir Das	Database development, Open Science Initiative, drafting and revising the manuscript
Stephanie O.M. Dyke	Led the ethics and policy of the Open Science initiative. Drafting and revising the manuscript.
Pierre Étienne	Former co-director of StoP-AD Centre. Leader of the clinical unit, major role in data acquisition. Study design. Participating in Data Sharing policies. Revising the manuscript.
Marie-Elyse Lafaille-Magnan	Leader of the odor identification data. Revising the manuscript.
Serge Gauthier	Study design counseling, Participating in Data Sharing policies. Revising the manuscript.
Pierre Bellec	Neuroimaging - Leader of Resting State fMRI data. Participation in Data sharing policies.
Louis Collins	Neuroimaging - Leader of Volumetry MRI data. Participation in Data sharing policies.
Natasha M. Rajah	Neuroimaging - Leader of Episodic memory

	task fMRI data.
Veronique Bohbot	Neuroimaging - Spatial Navigation
Yasser Iturria-Medina	Co-leader of the analytic unit. Designing and validating statistical/computational methods for analyzing the data.
Jeannie Leoutsakos	Co-leader of the analytic unit. Designing and validating statistical/computational methods for analyzing the data.
Justin Kat	Development of the database. Support to data entry and data sharing efforts.
Rick Hoge	Neuroimaging - Leader of Blood Flow fMRI data
Mallar Chakravarty	Neuroimaging - Leader of hippocampal subfield MRI data.
Jean-Baptiste Poline	CONP technical chair, meta data description and standardization, revision of the manuscript
Pedro Rosa-Neto	Leader of the lumbar puncture program. Major role in data acquisition. Editing the manuscript.
Sylvia Villeneuve	Current StoP-AD Centre Associate Director. Major role in data acquisition. Study design. Participating in Data Sharing policies. Editing the manuscript
Alan C. Evans	Database Development. Open Science Initiatives and Neuroimaging - Leader of cortical thickness data.
Judes Poirier	Current Director of StoP-AD. Major role in data acquisition and laboratory methods. Study design. Participating in Data Sharing policies. Editing the manuscript.
John C. S. Breitner	Founding StoP-AD Centre Director. Study conceptualization and design. Data sharing policies. Editing the manuscript. Study funding.

COMPETING INTEREST

No competing interest was disclosed.

REFERENCES

1 Villemagne, V. L. *et al.* Amyloid β deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer's disease: a prospective cohort study. *The Lancet Neurology* **12**, 357-367 (2013).

2 Iturria-Medina, Y. *et al.* Early role of vascular dysregulation on late-onset Alzheimer's disease based on multifactorial data-driven analysis. *Nature communications* **7**, 11934 (2016).

3 Dubois, B. *et al.* Preclinical Alzheimer's disease: definition, natural history, and diagnostic criteria. *Alzheimer's & Dementia* **12**, 292-323 (2016).

4 *Canada to Become 30th Country with National Dementia Strategy.*, <AlzheimerSocietyBlog.ca, > (2017).

5 Your guide to Canada's national dementia strategy. (2019).

6 Canada, A. s. S. *Dementia-numbers*, <http://alzheimer.ca/fr/Home/About-dementia/What-is-dementia/Dementia-numbers.> (

7 Mohaddes, Z. *et al.* National neuroinformatics framework for canadian consortium on neurodegeneration in aging (CCNA). *Frontiers in neuroinformatics* **12** (2018).

8 Breitner, J., Poirier, J., Etienne, P. & Leoutsakos, J. Rationale and Structure for a New Center for Studies on Prevention of Alzheimer's Disease (StoP-AD). *The journal of prevention of Alzheimer's disease* **3**, 236-242 (2016).

9 Breitner, J. C. *et al.* APOE-ε4 count predicts age when prevalence of AD increases, then declines: the Cache County Study. *Neurology* **53**, 321-321 (1999).

10 Huang, W., Qiu, C., von Strauss, E., Winblad, B. & Fratiglioni, L. APOE genotype, family history of dementia, and Alzheimer disease risk: a 6-year follow-up study. *Archives of neurology* **61**, 1930-1934 (2004).

11 Das, S., Zijdenbos, A. P., Vins, D., Harlap, J. & Evans, A. C. LORIS: a web-based data management system for multi-center studies. *Frontiers in neuroinformatics* **5**, 37 (2012).

12 Das, S. *et al.* The MNI data-sharing and processing ecosystem. *NeuroImage* **124**, 1188-1195 (2016).

13 McKiernan, E. C. *et al.* Point of view: How open science helps researchers succeed. *Elife* **5**, e16800 (2016).

14 Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data* **3** (2016).

15 Das, S. *et al.* Cyberinfrastructure for open science at the montreal neurological institute. *Frontiers in neuroinformatics* **10**, 53 (2017).

16 Poupon, V., Seyller, A. & Rouleau, G. A. The Tanenbaum open science Institute: leading a paradigm shift at the montreal neurological institute. *Neuron* **95**, 1002-1006 (2017).

17 Leoutsakos, J.-M., Gross, A., Jones, R., Albert, M. & Breitner, J. 'Alzheimer's Progression Score': Development of a Biomarker Summary Outcome for AD Prevention Trials. *The journal of prevention of Alzheimer's disease* **3**, 229 (2016).

18 Meyer, P.-F. *et al.* INTREPAD: A randomized trial of naproxen to slow progress of presymptomatic Alzheimer disease. *Neurology* **92**, e2070-e2080 (2019).
19 Poirier, J. *et al.* Apolipoprotein E and lipid homeostasis in the etiology and treatment of sporadic Alzheimer's disease. *Neurobiology of aging* **35**, S3-S10 (2014).
20 Morris, J. C. The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology* (1993).
21 Nasreddine, Z. S. *et al.* The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society* **53**, 695-699 (2005).
22 Leduc, V. *et al.* HMGCR is a genetic modifier for risk, age of onset and MCI conversion to Alzheimer's disease in a three cohorts study. *Molecular psychiatry* **20**, 867 (2015).
23 De Beaumont, L., Pelleieux, S., Lamarre-Theroux, L., Dea, D. & Poirier, J. Butyrylcholinesterase K and Apolipoprotein E-ε4 Reduce the Age of Onset of Alzheimer's Disease, Accelerate Cognitive Decline, and Modulate Donepezil Response in Mild Cognitively Impaired Subjects. *Journal of Alzheimer's Disease* **54**, 913-922 (2016).
24 Gosselin, N. *et al.* BDNF Val66Met polymorphism interacts with sleep consolidation to predict ability to create new declarative memories. *Journal of Neuroscience* **36**, 8390-8398 (2016).
25 Miron, J. *et al.* TLR4 Gene Expression and Pro-Inflammatory Cytokines in Alzheimer's Disease and in Response to Hippocampal Deafferentation in Rodents. *Journal of Alzheimer's Disease*, 1-10 (2018).
26 Randolph, C., Tierney, M. C., Mohr, E. & Chase, T. N. The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS): preliminary clinical validity. *Journal of clinical and experimental neuropsychology* **20**, 310-319 (1998).
27 Lafaille-Magnan, M.-E. *et al.* CROSSTALK BETWEEN ENGLISH AND FRENCH FORMS OF THE RBANS IN AN ALZHEIMER'S DISEASE PREVENTION TRIAL. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* **14**, P559-P560 (2018).
28 Lelental, N. *et al.* Comparison of different matrices as potential quality control samples for neurochemical dementia diagnostics. *Journal of Alzheimer's Disease* **52**, 51-64 (2016).
29 Bertolotto, A. *et al.* The use of the 25 Sprotte needle markedly reduces post-dural puncture headache in routine neurological practice. *Cephalalgia* **36**, 131-138 (2016).
30 Madjar, C. *et al.* Task-related BOLD responses and resting-state functional connectivity during physiological clamping of end-tidal CO2. *NeuroImage* **61**, 41-49, doi:https://doi.org/10.1016/j.neuroimage.2012.02.080 (2012).
31 Birn, R. M., Diamond, J. B., Smith, M. A. & Bandettini, P. A. Separating respiratory-variation-related fluctuations from neuronal-activity-related

fluctuations in fMRI. *NeuroImage* **31**, 1536-1548, doi:https://doi.org/10.1016/j.neuroimage.2006.02.048 (2006).

32 Glover, G. H., Li, T. Q. & Ress, D. Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* **44**, 162-167 (2000).

33 Ankudowich, E., Pasvanis, S. & Rajah, M. Changes in the correlation between spatial and temporal source memory performance and BOLD activity across the adult lifespan. *Cortex* **91**, 234-249 (2017).

34 Rajah, M. N. *et al.* Family history and APOE4 risk for Alzheimer's disease impact the neural correlates of episodic memory by early midlife. *NeuroImage: Clinical* **14**, 760-774 (2017).

35 Doty, R. L., Shaman, P. & Dann, M. Development of the University of Pennsylvania Smell Identification Test: a standardized microencapsulated test of olfactory function. *Physiology & behavior* **32**, 489-502 (1984).

36 Lafaille-Magnan, M.-E. *et al.* Odor identification as a biomarker of preclinical AD in older adults at risk. *Neurology* **89**, 327-335 (2017).

37 Lafaille-Magnan, M.-E. *Odor Identification as a Biomarker for Progression of Pre-symptomatic Alzheimer's Disease*, McGill University Libraries, (2018).

38 Tuwaig, M. *et al.* Deficit in central auditory processing as a biomarker of pre-clinical Alzheimer's disease. *Journal of Alzheimer's Disease* **60**, 1589-1600 (2017).

39 Tuwaig, M. *Central Auditory Processing as a Marker for Disease Progression in Individuals at Risk for Alzheimer's Dementia*, McGill University Libraries, (2016).

40 Gates, G. A., Anderson, M. L., Feeney, M. P., McCurry, S. M. & Larson, E. B. Central auditory dysfunction in older persons with memory impairment or Alzheimer dementia. *Archives of Otolaryngology–Head & Neck Surgery* **134**, 771-777 (2008).

41 Orban, P. *et al.* Test-retest resting-state fMRI in healthy elderly persons with a family history of Alzheimer's disease. *Scientific data* **2**, 150043 (2015).

42 Fonov, V. & Collins, D. L. BIC Defacing Algorithm. *BioRxiv*, 275453 (2018).

43 Dyke, S. O. *et al.* Registered access: authorizing data access. *European Journal of Human Genetics* **26**, 1721 (2018).

44 Dyke, S. O. *et al.* Consent codes: upholding standard data use conditions. *PLoS genetics* **12**, e1005772 (2016).

45 Corpas, M., Kovalevskaya, N. V., McMurray, A. & Nielsen, F. G. A FAIR guide for data providers to maximise sharing of human genomic data. *PLoS computational biology* **14**, e1005873 (2018).

46 Brodeur, M. B., Dionne-Dostie, E., Montreuil, T. & Lepage, M. The Bank of Standardized Stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. *PloS one* **5**, e10773 (2010).

47 Brodeur, M. B., Guérard, K. & Bouras, M. Bank of standardized stimuli (BOSS) phase II: 930 new normative photos. *PLoS One* **9**, e106953 (2014).

48 Tardif, C. L. *et al.* Regionally specific changes in the hippocampal circuitry accompany progression of cerebrospinal fluid biomarkers in preclinical Alzheimer's disease. *Human brain mapping* **39**, 971-984 (2018).

49 Meyer, P.-F. *et al.* Bi-directional association of cerebrospinal fluid immune markers with stage of Alzheimer's disease pathogenesis. *Journal of Alzheimer's Disease* **63**, 577-590 (2018).

50 Miron, J. *et al.* Association of TLR4 with Alzheimer's disease risk and presymptomatic biomarkers of inflammation. *Alzheimer's & Dementia* (2019).

51 Meyer, P. F. *et al.* No apparent effect of naproxen on CSF markers of innate immune activation. *Annals of clinical and translational neurology* (2019).

52 Vogel, J. W. *et al.* Brain properties predict proximity to symptom onset in sporadic Alzheimer's disease. *Brain* **141**, 1871-1883 (2018).

53 Villeneuve, S. *et al.* Proximity to parental symptom onset and amyloid-β burden in sporadic Alzheimer disease. *JAMA neurology* **75**, 608-619 (2018).

54 Verfaillie, S. C. *et al.* Subjective cognitive decline is associated with altered default mode network connectivity in individuals with a family history of Alzheimer's disease. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* **3**, 463-472 (2018).

55 Dadar, M. *et al.* Validation of a regression technique for segmentation of white matter hyperintensities in Alzheimer's disease. *IEEE transactions on medical imaging* **36**, 1758-1768 (2017).

56 Baker, M. 1,500 scientists lift the lid on reproducibility. *Nature News* **533**, 452 (2016).

57 Peng, R. The reproducibility crisis in science: A statistical counterattack. *Significance* **12**, 30-32 (2015).

58 Nichols, T. E. *et al.* Best practices in data analysis and sharing in neuroimaging using MRI. *Nature neuroscience* **20**, 299 (2017).

59 CONP. *Publication policy*, <https://conp.ca/community/governance/publication-policy/>

60 Sherif, T. *et al.* CBRAIN: a web-based, distributed computing platform for collaborative neuroimaging research. *Frontiers in neuroinformatics* **8**, 54 (2014).