5-YEAR DYNAMIC PREDICTION OF DEMENTIA USING REPEATED MEASURES OF COGNITIVE TESTS AND A DEPENDENCY SCALE

Céline Ben-Hassen[1], Catherine Helmer[1], Claudine Berr[2], Hélène Jacqmin-Gadda[1]

[1] INSERM, Bordeaux Population Health Research Center, Univ Bordeaux, ISPED, Bordeaux, France
[2] INM, Univ Montpellier, INSERM, Montpellier, France

Corresponding author :
Helene Jacqmin-Gadda
Univ Bordeaux, ISPED
INSERM UMR1219 - Bordeaux Population Health Research Center,
146 rue Léo Saignat, CS 61292,
33076 Bordeaux, France

Word count: 3997, 7 tables and figures and 47 references

Abstract: 190

Abstract

The progression of dementia prevalence over the years and the lack of efficient treatments to stop or reverse the cognitive decline make dementia a major public health challenge in the developed world. Identifying subjects at high risk of developing dementia could improve the management of these patients and help selecting the target population for preventive clinical trials. We used joint modeling to build a dynamic prediction tool of dementia based on the change over time of two neurocognitive tests (the Mini-Mental State Examination and the Isaacs Set Tests) as well as an autonomy scale (the Instrumental Activities of Daily Living). The model was estimated on the French cohort PAQUID and validated both internally in cross-validation and externally on the French cohort 3C. We evaluated its predictive abilities through AUCROC and Brier score accounting for right censoring and competing risk of death and obtained AUC values equal to 0.95 in average for the risk of dementia in the next 5 or 10 years. This tool is able to discriminate a high risk group of subjects from the rest of the population. This could be of great help in clinical practice and research.

Keywords: Alzheimer's disease, Cognition, Dementia, Dependency, Joint model, Prediction

Dementia is defined as a progressive cognition-altering disease that significantly impacts daily life, Alzheimer's disease being its most common form (70-80%) (1,2). By now 47 million of people are affected worldwide. This number is predicted to reach 131 million by 2050 due to population aging (3,4) while no efficient treatment is currently available (5,6).

Clinical signs of dementia leading to diagnosis are preceded by a long pre-diagnosis phase of cognitive decline (7,8). The general consensus is that to be efficient, future treatments should target subjects in the pre-diagnosis phase (10). Being able to identify early subjects with high risk of dementia is of major interest to design, evaluate and later apply such a therapeutic/prevention strategy.

Several prediction models have already been proposed using biomarkers, brain imaging, or cognitive tests (11,12,13). Those based on combinations of cognitive tests and/or autonomy scales are especially useful since such measures are noninvasive and not expensive and thus easily collected in epidemiologic studies and clinical practice (13–15). However, previous works proposed prediction tools based on a single measure or rarely two measures (16) of each predictor while cognitive tests and autonomy scales are time-dependent. Exploiting all the measures collected for a given individual until the current time and updating the prediction with each new measure could lead to more precise predictions since they would account for the individual change over time of the predictors. A few prediction scores for dementia have been previously proposed based on several repeated measures of a single time-dependent marker (17-19).

In this paper, using a joint model for the time to dementia and multiple longitudinal markers, we propose a tool to predict the risk of dementia in the next 5 or 10 years using repeated measures of several easy to perform cognitive tests and an autonomy scale.. The model is estimated on a large French population based cohort (PAQUID) and carefully evaluated by

cross-validation on PAQUID and on an external validation cohort (3C) accounting for censoring and competing risk of death.

METHODS

Data

PAQUID. The Personnes Agées QUID (PAQUID) cohort was used as learning sample. Created in 1988, it aimed to study brain aging and dependency in the elderly (20). The cohort included 3,777 subjects randomly selected on the electoral rolls aged 65 years and over living at home in two Southwest departments of France. Participants first received a visit at home by a trained psychologist to collect information regarding their social, demographic and medical situation. Subjects were also asked to perform several neurocognitive tests. Dementia diagnosis was established by a two-step procedure detailed in Supplementary data: a screening was performed by the psychologist, then subjects screened positive were visited by a neurologist who established the diagnosis according to the DSMIIIR criteria. Subjects were then followed for 25 years (with visits at t = 1, 3, 5, 8, 10, 13, 15, 17, 20, 22, and 25 years after inclusion) with the same set of cognitive tests and the same diagnosis procedure for dementia at each visit. For this study, the sample of analysis excluded subjects who were diagnosed with dementia, subjects with major hearing or visual impairment or confined to bed at inclusion, subjects without any follow-up visit and subjects who had no value during the follow-up for at least one of the four predictors considered (IST, MMSE, BVRT, IADL). The final sample included 2,880 subjects (Flow-chart in Figure 1). The data are available upon request from the 2nd author.

Three-City (3C). The Three-City cohort was used as validation sample. This cohort includes participants aged 65 years and over randomly selected on the electoral rolls of three French cities (Bordeaux, Dijon and Montpellier) at the baseline visit between 1999 and 2001 (21). The validation sample consists only of the participants of Montpellier and Bordeaux (N=4363)

where follow-up lasted up to 17 years after inclusion with visits at t = 2, 4, 7, 10, 12, 14 and 17 years (for Bordeaux only), whereas the follow-up was shorter in Dijon. The study design was similar to PAQUID. At baseline and at each follow-up visit participants completed several neurocognitive tests and were diagnosed with dementia according to the DSM-IV criteria as described in the Supplementary data.). The same exclusion criteria as for the PAQUID sample were applied, leading to a final validation sample of 3,953 subjects (Figure 2). The data are available upon request from e3c.coordinatingcenter@gmail.com.

Cognitive tests

For building the prediction model we considered repeated measures of three cognitive tests and the IADL scale whose values at inclusion were all previously found predictors of dementia (14,22) and were repeatedly measured in both cohorts.

Isaacs Set Test (IST). The IST measures verbal fluency (23). Participants must give as many words as possible belonging to four semantic categories (animals, cities, fruits, colors). In PAQUID, the test was stopped when the participant reached 10 words in each category leading to a strong ceiling effect. Given this truncation, we previously demonstrated that the score at 15 seconds was a better measure of cognitive decline and had a Gaussian distribution (24,25). Thus we used the score at 15 seconds truncated at 10 words in each category both in PAQUID and 3C. The score ranges from 0 to 40.

Mini-Mental State Examination (MMSE). The MMSE is a measure of global cognitive functioning (26). It measures orientation to space, orientation to time, immediate and delayed memory (repeating three words immediately and after a while), attention, language, and constructive praxis. The score ranges from 0 to 30 (higher is better).

Visual Retention Test of Benton (BVRT). The BVRT evaluates the visual memory: participants are shown a simple geometrical figure, and are then asked to find each among four proposals (27); this task is repeated for 15 different figures. The final score is the number of correct answers. It ranges from 0 to 15.

Instrumental Activities of Daily Living (IADL). IADL is a scale that measures dependency. It consists of eight items in total, each evaluating one aspect of daily life. For each item, the scale ranges from 0 (completely independent, no help required) to 4 (completely dependent, unable to perform the activity alone). We used a subscale named IADL4 which only takes into consideration four items that were collected for both men and women and were previously shown to be associated with the risk of dementia in the PAQUID study (28): using the telephone, using means of transport, taking medication as prescribed, managing money.

Statistical analysis

Joint models. The prediction model was a multivariate joint model (29) that fits simultaneously the trajectories over time of the four markers using mixed models and the time-to-dementia using a proportional hazard model with approximation of the baseline hazard by penalized splines. Cognitive scores changes over time were described by linear mixed model while IADL4 change was modeled by a log-linear Poisson mixed model. The proportional hazard model included as explanatory variables both the expected current value of the markers and the slope of the markers' trajectories (on the log scale for IADL4).

As the MMSE and BVRT have asymmetric distribution with floor and/or ceiling effects that can strongly biased linear mixed model estimates (25), they were normalized using a pre-transformation estimated with a latent process mixed model in the LCMM package (30). As MMSE is a widely used test, the normalizing transformation was previously estimated with LCMM and validated on several cohorts and a tool was developed for easy computation of the

normalized score (NormPsy, 31). The BVRT transformation was directly estimated on the PAQUID sample with LCMM. The normalized MMSE and BVRT and the IST scores were then centered and standardized using the mean and standard deviation at inclusion.

Model estimation. The strategy of analysis is detailed in the web appendix. Briefly "unimarker" joint models were first estimated on repeated measures of each marker and time-to-dementia using the PAQUID sample. The mixed-effect sub-models and the time-to-dementia sub-model were both adjusted on the age at inclusion (standardized as such: $(age - 65)/10$) for the mixed-effect sub-models), sex, and education status (1 if the subject had at least a primary school diploma, 0 if not). We selected the best unimarker models by comparing different modeling of the time-trend for the markers, . Second, the multimarkers model was estimated. Models were estimated using the R package JMbayes (32). Estimation of joint model with several longitudinal markers is challenging due to the large number of parameters and random effects. As convergence was not reached with the most complex 4-marker model, we removed the BVRT test which was by far the least predictive among the four unimarker models and slightly simplified the modeling of marker trajectories. In the final model, the time trend of the cognitive tests was modeled through cubic splines with 1 interior node, while that of IADL4 was assumed to be linear (on the log scale). The three mixed models included correlated random intercept and slope with an unconstraint correlation structure between the 6 random effects.

Predictions

Using parameter estimates of the joint models, we estimated for each subject i still at risk of dementia at a given time *s*, the probability to develop dementia between s and $s + t$, given the repeated measures of the markers collected until s using formula *[3]* in Supplementary data. As each visit can take place approximately within one year of the planned date, for PAQUID, predictions were computed for s = 0, 1, 2, 4, 6, 9, 11, 14, 16 and 18 years. For 3C, predictions

were computed for s = 0, 3, 5, 8 and 11 years. Every prediction was calculated with an horizon (t) equal to 5 years and 10 years.

Evaluation of predictive abilities

AUC and Brier Scores. Predictive abilities of the model were evaluated both by 5-fold cross-validation on the learning data set PAQUID, and on the external validation sample 3C. The predictive accuracy was assessed by the AUC under the ROC curve (which represents the capacity of a model to discriminate between subjects who underwent the event and subjects who did not) and the Brier Score (BS, which measures the error of prediction). The Inverse probability of Censoring Weighting (IPCW) estimators of AUC and BS were computed using the R package TimeRoc accounting for both the right censoring and the competing risk of death (17,33).

For subjects not diagnosed with dementia during the follow-up, their dementia status at death was unknown if they died several years after the last visit. As recommended by Jacqmin-Gadda et al (14) for computing AUC and BS for interval censored data, subjects who died without dementia diagnosis more than 2 years after the last visit were considered as right censored at the last visit. To account for competing risks, two definitions of AUC were estimated. In the first one, a control was defined as a subject who was free of any event (dementia or death) at time $s + t$ (Formula [*4a*], Supplementary data), and in the second one, a control was defined as a subject who was free of the main event (dementia) at time $s + t$ (Formula [*4b*], Supplementary data).

Estimation of cumulative incidence of dementia and calibration. The joint model was estimated without accounting for competing death and thus the predicted probability was the probability to develop dementia in the 5 or 10 years assuming a null death risk in the window. To quantify the dementia risk accounting for competing risk of death and interval censoring for dementia,

we computed a non-parametric estimate of the 5-year and 10-year cumulative incidence of dementia for four risk groups defined according to the quartiles of the predicted probabilities (low, medium, high and very high) at each prediction time. For each group and each time s, an illness-death model was estimated by penalized likelihood maximization accounting for interval censoring (34) using all the subjects still at risk of dementia at the closest visit preceding s. The t-year cumulative incidence is the probability to develop dementia before dying in the next t years. The estimation was performed using the package SmoothHazard (35).

To evaluate the calibration of the predicted probabilities for t=5 and 10 years and the impact of competing risks and interval censoring, we compared these estimates with the means of the t-year predicted probabilities from the joint model and with the t-year cumulative incidence estimated by a Nelson Aalen estimator (using the packages survival (36) and riskRegression (37)) which neglects the competing death and interval censoring.

RESULTS

Data description

The training (PAQUID) and validation samples (3C) are described in Table 1. Both cohorts have a mean age at inclusion around 73 years old and included approximately 60% of women. Subjects in 3C were more educated than subjects in PAQUID: 68.7% in PAQUID, 91.2% in 3C have at least a primary school diploma (PSD). As a result, the initial values of the cognitive tests were slightly higher in 3C. Similarly, there were more subjects who were considered as autonomous at inclusion according to the IADL4 scale in 3C than in PAQUID (89.9% vs 75.4%). Accordingly, the incidence of dementia in PAQUID was higher than in 3C (Supplementary Figure 1). Supplementary Figure 2 displays the spaghetti plots of observed values of MMSE, IST, BVRT and IADL for subsamples of PAQUID and 3C as well as smooth

mean curves estimated on whole samples; this highlights the differential evolution between subjects who developed dementia and the others.

Model estimation

Table 2 summarizes estimates of the four unimarker joint models and the final 3-marker joint model. An increase of the age at inclusion was associated with a significant increased risk of dementia (adjusted on sex, education status, current value and slope of the marker) except in models including IADL. After adjustment on current cognitive and IADL scores and their slopes, the risk of dementia significantly increased with the educational level. Indeed, when subjects with a higher educational status reach the same values and slopes of cognitive tests than subjects with a lower educational status, it is a consequence of a stronger cognitive decline, associated with a higher risk of dementia. After adjustment, women showed a significant lower risk of dementia compared to men in all models except in the unimarker models considering only IST or only BVRT.

According to the unimarker models, an increase in the current value of each cognitive test or its current slope (i.e. a less decreasing slope since slopes are negative) was significantly associated with a decreased risk of developing dementia. Opposite statements apply to the IADL4 scale as an increase of the IADL4 value means more dependency. In the fully adjusted multimarker joint model, the directions of the associations remained the same but the associations were much weaker for each marker.

Comparison of 5-year predictive abilities

Figure 3 compares the 5-year predictive abilities of the four unimarker models and the final 3-markers model by 5-fold cross-validation on the PAQUID cohort (left panel) and by external validation on the 3C cohort (right panel). The first AUC evaluates the discrimination between subjects who develop dementia and subjects alive and free of dementia at 5-years while the

second one evaluates discrimination between subjects who develop dementia and all the others (who died without dementia within 5 years or were alive and free of dementia 5 years later). These two estimates were very similar on Figure 3. Among the four unimarker models, MMSE and IADL4 had the best predictive abilities according to both AUC and BS. Their AUC values were above 0.9 on PAQUID (cross-validation), and above 0.85 on 3C. These AUCs are the probabilities that a subject randomly selected among those who develop dementia in the next 5 years has a higher predicted probabilities of dementia than a randomly selected subject free of dementia after five years. The IST exhibited only slightly lower predictive abilities while prediction performances of the BVRT were much lower.

More importantly, the joint model combining MMSE, IST and IADL4 exhibited better predictive abilities than each of the unimarker model, both in internal and external validation, whatever the evaluation criterion and whatever the time of predictions. The cross-validated AUC values on PAQUID were between 0.93 and 0.96 depending on the time of prediction and between 0.85 and 0.96 on the 3C cohort.

The predictive abilities of the models were further evaluated in the subsamples with “high cognitive level” (no IADL limitations and MMSE>28) and “no cognitive impairment” (no IADL limitations and MMSE>23) at baseline. Results were very similar with those of the full samples but the difference between the MMSE model and the 3-markers model was very small among the high cognitive level subjects (Supplementary Figures 3 and 4).

5-year cumulative incidence estimation and calibration

Supplementary Figure 5 displays the non-parametric estimates of the 5-year cumulative incidence of dementia in each risk group defined by the quartiles of the distribution of the predicted probabilities, according to the time of prediction accounting for competing death and interval censoring. In PAQUID as well as in 3C, the results show a much higher cumulative

incidence in the fourth quartile. In this group, the 5-year cumulative incidence ranges from 0.26 to 0.66 according to the time of prediction in PAQUID and from 0.23 to 0.32 in 3C. Subjects belonging to the other quartiles have a very low risk of dementia before death: below 0.1 in 3C at all timepoints, and below 0.2 at most timepoints in PAQUID with a clear gradient between the 3 quartiles.

Figure 4 (top panel) compares the above estimates of the cumulative incidence with the Nelson-Aalen estimates neglecting competing death and interval censoring and with the mean of the predicted probabilities. The latter fits well the Nelson Aalen estimates showing a good calibration of the joint model. The differences with the illness-death estimator are also quite small highlighting a modest impact of the competing risk of death and interval censoring on 5-year periods except for the medium and high risk group for PAQUID.

Overall, the results do show that the predictive tool is well calibrated and exhibits excellent discrimination.

10-year prediction

Figure 5 presents the predictive abilities of the 5 models for 10-year dementia prediction. The 3-markers model remained the best but the difference between the AUC and BS of the BVRT model and the other models was smaller. As expected, the main difference between 5-year and 10-year prediction was the stronger impact of competing death for 10-year prediction. Indeed the AUC considering deceased subjects as control (AUC2) was smaller than the AUC excluding subjects who died (AUC1) and the predicted probabilities assuming no death in the highest quartile were overestimated compared to the illness-death estimates accounting for competing risks and interval censoring as shown in the calibration plot (Figure 4 bottom panel).

DISCUSSION

We designed a predictive tool based on the evolution over time of two neurocognitive tests (IST and MMSE) and one dependency scale (IADL4). This tool is able to predict the risk of dementia in the following 5 or 10 years with great accuracy (AUC values close to 0.95). Especially, subjects in the highest quartile of predictive scores have a much higher risk of dementia than other subjects.

Existing predictive models in the literature show a lot of diversity depending on the variables and factors selected to compute the risk of dementia. Few of them have been validated (13,38). The main variables selected to build the predictive tools are demographic information (age, sex, education), cognitive tests scores (the MMSE being the most used), cardio-vascular risk factors and comorbidities and genetic factors (mostly the APOE allele status). Many prediction models are based on one or several cognitive scores (39–45), leading to AUC values ranging between 0.63 (44) and 0.89 (40,41). Combining information from both cognitive tests and the dependency scale IADL4 also led to good prediction accuracy (15). However, all these tools only use baseline evaluation to predict the risk of dementia. Some works have compared predictive abilities of models using repeated measures of a unique predictor, either cognitive tests or functional scales (17-19). A few tools based on repeated measures of several markers were proposed using either deep learning (46) or functional principal component analysis (47). Most of them used data from the Alzheimer Disease Neuroimaging Initiative (18,19,46,47) and focused on short-term prediction of conversion from MCI to dementia (6 to 18 months) reaching an AUC around 0.8. To our knowledge, no studies have been conducted to predict the risk of dementia in the general population using joint models with multiple longitudinal markers.

Our study has several strengths. First, this work was conducted on two large population-based cohorts (PAQUID and 3C) with a long follow-up, which allowed us to validate the tool both in cross-validation and in external validation. Additionally, unlike many cohorts that establish dementia based on algorithms only, both PAQUID and 3C use an accurate clinical diagnosis for dementia established by a neurologist according to validated criteria. Second, our predictive tool combines information from repeated measurements of two neurocognitive tests as well as a dependency scale,. Using repeated measurements rather than baseline values allows to detect a cognitive decline or loss of autonomy. Even if the MMSE and the IADL4 are already on their own good predictors of dementia, we succeeded in increasing the predictive abilities of the tool by combining information from these markers with the IST, reaching AUC values close to 0.95.

A limitation of this work is that the competing risk of death was not fully handled. As no software was available for estimating joint models and computing predictions with multiple markers accounting for both competing risks and interval censoring of the main event, the joint model was estimated by neglecting these two issues. However, predictive abilities of the joint model were evaluated by AUC and BS accounting for the competing risk of death. Moreover, to quantify the impact of competing risk and interval censoring, we estimated the cumulative incidence of dementia over the window of prediction in each risk group using a non-parametric approach that deals with these two issues and compared it with the Nelson-Aalen estimator and the mean of the predicted probabilities from the joint models. The joint model appeared well calibrated and its 5-year predictions were close to those obtained by the non-parametric illness-death model. This suggests that the impact of the competing risk and interval censoring remains modest over 5-year intervals of prediction. We also faced convergence issues when estimating multimarker models. Consequently, we were not able to include the BVRT in the final model. As it was by far the least predictive marker on its own, however, it is very likely that it would not have significantly increased the predictive abilities of the tool.

We estimated two discrimination measures. The first AUC is more useful for individuals who want to quantify their risk of developing dementia if they survive from other causes for the next 5 or 10-years while the second one is more useful in a Public Health perspective to discriminate all subjects who could need care for dementia.

Conclusion

Our results show that combining repeated measures of two cognitive tests (IST and MMSE) and a dependency scale (IADL4) increases the predictive abilities of a simple and inexpensive dementia prediction tool, reaching AUC values around 0.95. Using joint modeling makes it possible to compute predictions that can be updated at each new measurement. Numerous researches are currently ongoing to evaluate the performances of biomarkers for dementia and Alzheimer's disease prediction. However, our results show that, using two cognitive tests and a dependency scale very easy and quick to measure, we were able to obtain a highly powerful dementia prediction tool. This tool could be of great help to assist clinicians or clinical researchers in identifying high-risk subjects, possibly to be included in clinical prevention trial. Moreover, the prediction models using only repeated measures of MMSE, or to a lesser degree, IADL scale, also exhibited good behavior. This makes these tools interesting when the three markers are not collected or for use by general practitioners, for either reassuring their worried patients or referring them to specialists if needed. A new assessment of prediction performances in this framework would be useful.

References

1. Qiu C, Kivipelto M, von Strauss E. Epidemiology of Alzheimer's disease: occurrence, determinants, and strategies toward intervention. Dialogues *Clin. Neurosci*. 2009;11(2):111–128.

2. Prince M, Wimo A, Guerchet M, et al. World Alzheimer Report 2015. The Global

Impact of Dementia. An Analysis of Prevalence, Incidence, Cost and Trends. 2015.

3. Arvanitakis Z, Shah RC, Bennett DA. Diagnosis and Management of Dementia: Review. *JAMA*. 2019;322(16):1589–1599.

4. Alzheimer's Disease International. Dementia Statistics. (https://www.alz.co.uk/research/statistics). (Accessed November 6, 2020)

5. Cummings J. Lessons Learned from Alzheimer Disease: Clinical Trials with Negative Outcomes. *Clin. Transl. Sci*. 2018;11(2):147–152.

6. Cummings JL, Morstorf T, Zhong K. Alzheimer's disease drug-development pipeline: few candidates, frequent failures. *Alzheimers Res. Ther*. 2014;6(4):37.

7. Amieva H, Mokri H, Le Goff M, et al. Compensatory mechanisms in higher-educated subjects with Alzheimer's disease: a study of 20 years of cognitive decline. *Brain J. Neurol.* 2014;137(4):1167–1175.

8. Sperling RA, Aisen PS, Beckett LA, et al. Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement*. 2011;7(3):280–292.

9. Morris JC, Price JL. Pathologic correlates of nondemented aging, mild cognitive impairment, and early-stage alzheimer's disease. *J. Mol. Neurosci*. 2001;17(2):101-118.

10. Aisen PS, Andrieu S, Sampaio C, et al. Report of the task force on designing clinical trials in early (predementia) AD. *Neurology*. 2011;76(3):280–286.

11. Di Stefano F, Epelbaum S, Coley N, et al. Prediction of Alzheimer's Disease Dementia: Data from the GuidAge Prevention Trial. *J. Alzheimers Dis*. 2015;48(3):793–804.

12. Eckerström C, Olsson E, Bjerke M, et al. A Combination of Neuropsychological, Neuroimaging, and Cerebrospinal Fluid Markers Predicts Conversion from Mild Cognitive Impairment to Dementia. *J. Alzheimers Dis*. 2013;36(3):421–431.

13. Tang EYH, Harrison SL, Errington L, et al. Current Developments in Dementia Risk Prediction Modelling: An Updated Systematic Review. *Plos One.* 2015;10(9):e0136181.

14. Silva D, Guerreiro M, Santana I, et al. Prediction of Long-Term (5 Years) Conversion to Dementia Using Neuropsychological Tests in a Memory Clinic Setting. *J. Alzheimers Dis*. 2013;34(3):681–689.

15. Jacqmin-Gadda H, Blanche P, Chary E, et al. Prognostic score for predicting risk of dementia over 10 years while accounting for competing risk of death. *Am. J. Epidemiol.* 2014;180(8):790–798.

16. Chary, E., Amieva, H., Pérès, K., Orgogozo, J. M., Dartigues, J. F., & Jacqmin-Gadda, H. (2013). Short-versus long-term prediction of dementia among subjects with low and high educational levels. *Alzheimer's & Dementia*, 9(5), 562-571.

17. Blanche P, Proust-Lima C, Loubère L, et al. Quantifying and comparing dynamic predictive accuracy of joint models for longitudinal marker and time-to-event in presence of censoring and competing risks. *Biometrics*. 2015;71(1):102–113.

18. Li K, Chan W, Doody RS, et al. Prediction of Conversion to Alzheimer's Disease with Longitudinal Measures and Time-To-Event Data. *J. Alzheimers Dis*. 2017;58(2):361–371.

19. Wu Y, Zhang X, He Y, et al. Predicting Alzheimer's disease based on survival data and longitudinally measured performance on cognitive and functional scales. *Psychiatry Res*. 2020;291:113201.

20. Letenneur L, Commenges D, Dartigues JF, et al. Incidence of dementia and Alzheimer's

disease in elderly community residents of south-western France. *Int. J. Epidemiol.* 1994;23(6):1256–1261.

21 The 3C Study Group. Vascular factors and risk of dementia: design of the Three-City Study and baseline characteristics of the study population. *Neuroepidemiology.* 2003;22(6):316–325.

22 Tombaugh TN, McIntyre NJ. The Mini-Mental State Examination: a comprehensive review. *J. Am. Geriatr. Soc*. 1992;40(9):922–935.

23 Isaacs B, Kennie A. The Set Test as an aid to the detection of dementia in old people. *Br. J. Psychiat.* 1973;123:467–470.

24. Proust-Lima C, Amieva H, Dartigues J-F, et al. Sensitivity of Four Psychometric Tests to Measure Cognitive Changes in Brain Aging-Population–based Studies. *Am. J. Epidemiol.* 2006;165(3):344–350.

25. Proust-Lima C, Dartigues J-F, Jacqmin-Gadda H. Misuse of the Linear Mixed Model When Evaluating Risk Factors of Cognitive Decline. *Am. J. Epidemiol*. 2011;174(9):1077–1088.

26. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res*. 1975;12(3):189–198.

27. Benton AL. A visual retention test for clinical use. *Arch. Neurol. Psychiatry.* 1945;54(3):212–216.

28. Pérès, K., Helmer, C., Amieva, H., Orgogozo, J. M., Rouch, I., Dartigues, J. F., & Barberger-Gateau, P. (2008). Natural history of decline in instrumental activities of daily living performance over the 10 years preceding the clinical diagnosis of dementia: a prospective population-based study. *J Am Geriatr Soc*, 56(1), 37-44.

29. Rizopoulos, Dimitris, and Pulak Ghosh. "A Bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event." *Stat. Medicine* 2011;30(12):1366-1380.

.

30. Proust-Lima C, Philipps V, Liquet B. Estimation of Extended Mixed Models Using Latent Classes and Latent Processes: The R Package lcmm. *J. Stat. Softw.* 2017, 1(2) [electronic article]. (https://www.jstatsoft.org/v078/i02)

31. Philipps V, Amieva H, Andrieu S, et al. Normalized Mini-Mental State Examination for Assessing Cognitive Change in Population-Based Brain Aging Studies. *Neuroepidemiology*. 2014;43(1):15–25.

32. Rizopoulos D. The R Package JMbayes for Fitting Joint Models for Longitudinal and Time-to-Event Data Using MCMC. *J. Stat. Softw.* 2016; 1(7) [electronic article]. (https://www.jstatsoft.org/v072/i07)

33. Blanche P., Dartigues J. F., Jacqmin-Gadda H. (2013). Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat. Medicine* 2013;32(30):5381-5397.

34 Joly, P., Commenges, D., & Letenneur, L. A Penalized Likelihood Approach for Arbitrarily Censored and Truncated Data: Application to Age-Specific Incidence of Dementia. *Biometrics*, 1998; 54(1), 185-194

35. Touraine C, Gerds TA, Joly P. SmoothHazard: An R Package for Fitting Regression Models to Interval-Censored Observations of Illness-Death Models. *J. Stat. Softw.* 2017; 1(7) [electronic article]. (https://www.jstatsoft.org/v079/i07)

36. Therneau T, April. A Package for Survival Analysis in S. 2020

37. Ozenne B, Sørensen AL, Scheike T, et al. riskRegression: Predicting the Risk of an Event using Cox Regression Models. *The R Journal*. 2017;9:440–460.

38. Licher, S., Yilmaz, P., Leening, M. J., et al. External validation of four dementia prediction models for use in the general community-dwelling population: a comparative analysis from the Rotterdam Study. *European journal of epidemiology*. 2018; 33(7), 645-655.

39. Mura T, Baramova M, Gabelle A, et al. Predicting dementia using socio-demographic characteristics and the Free and Cued Selective Reminding Test in the general population. *Alzheimers Res. Ther*. 2017;9(1):21.

40. Wolfsgruber S, Jessen F, Wiese B, et al. The CERAD neuropsychological assessment battery total score detects and predicts Alzheimer disease dementia with high diagnostic accuracy. Am. J. Geriatr. Psychiatry Off. *J. Am. Assoc. Geriatr. Psychiatry*. 2014;22(10):1017–1028.

41. Derby CA, Burns LC, Wang C, et al. Screening for predementia AD: time-dependent operating characteristics of episodic memory tests. *Neurology*. 2013;80(14):1307–1314.

42. Grober E, Sanders AE, Hall C, et al. Free and cued selective reminding identifies very mild dementia in primary care. *Alzheimer Dis. Assoc. Disord*. 2010;24(3):284–290.

43. Restaino M, Matthews FE, Minett T, et al. Predicting risk of 2-year incident dementia using the CAMCOG total and subscale scores. *Age Ageing*. 2013;42(5):649–653.

44. Ehreke L, Luppa M, Konig H, et al. Does the Clock Drawing Test predict dementia? - Results of the Leipzig Longitudinal Study of the Aged (LEILA 75+). Z. *Gerontol. Geriatr*. 2011;43:60–60.

45. Jorm AF, Masaki KH, Petrovitch H, et al. Cognitive deficits 3 to 6 years before dementia onset in a population sample: the Honolulu-Asia aging study. *J. Am. Geriatr. Soc*.

2005;53(3):452–455.

46. El-Sappagh, S., Abuhmed, T., Islam, S. R., & Kwak, K. S. (2020). Multimodal multitask deep learning model for Alzheimer's disease progression detection based on time series data. *Neurocomputing* 2020, 412, 197-215.

47. Li K Luo S. Dynamic prediction of Alzheimer's disease progression using features of multiple longitudinal outcomes and time-to-event data. *Stat. Medicine*. 2019 ; 38(24) : 4804-4818.

Figure 1: Flowchart (PAQUID cohort).

Figure 2: Flowchart (3C cohort).

Figure 3: Estimation of the AUC and the Brier Scores for 5-year dementia prediction from time s on PAQUID (5-fold cross-validation) (A) and 3C (external validation) (B). To compute the AUC n°1, a control is defined as a subject who was free of any event (dementia or death) at time $\boldsymbol{s + 5}$. To compute the AUC n°2, a control is defined as a subject who was free of the main event (dementia) at time $\boldsymbol{s + 5}$. Predictions are computed from either the final multimarker joint model (solid line) or a unimarker joint model based on MMSE (short dashed line), IADL4 (dotted line), IST (dashed-dotted line) or BVRT (long dashed line).

Figure 4: Calibration of the predictive tool on PAQUID (A) and 3C (B). Comparison of the means of the predicted probabilities from the joint model with non parametric estimates of the 5-year (top panel) and 10-year (bottom panel) cumulative incidence computed by Nelson-Aalen estimators and illness-death models. Red, very high risk group; yellow, high risk group; blue, medium risk group; black, low risk group; solid line, Nelson-Aalen estimators; dashed line, illness-death models estimators; dotted line, means of the predicted probabilities from the joint model.

Figure 5: Estimation of the AUC and the Brier Scores for 10-year dementia prediction from time s on PAQUID (5-fold cross-validation) (A) and 3C (external validation) (B). Definitions of AUC n°1 and AUC n°2 are identical to Figure 3. Predictions are computed from either the final multimarker joint model (solid line) or a unimarker joint model based on MMSE (short dashed line), IADL4 (dotted line), IST (dashed-dotted line) or BVRT (long dashed line).

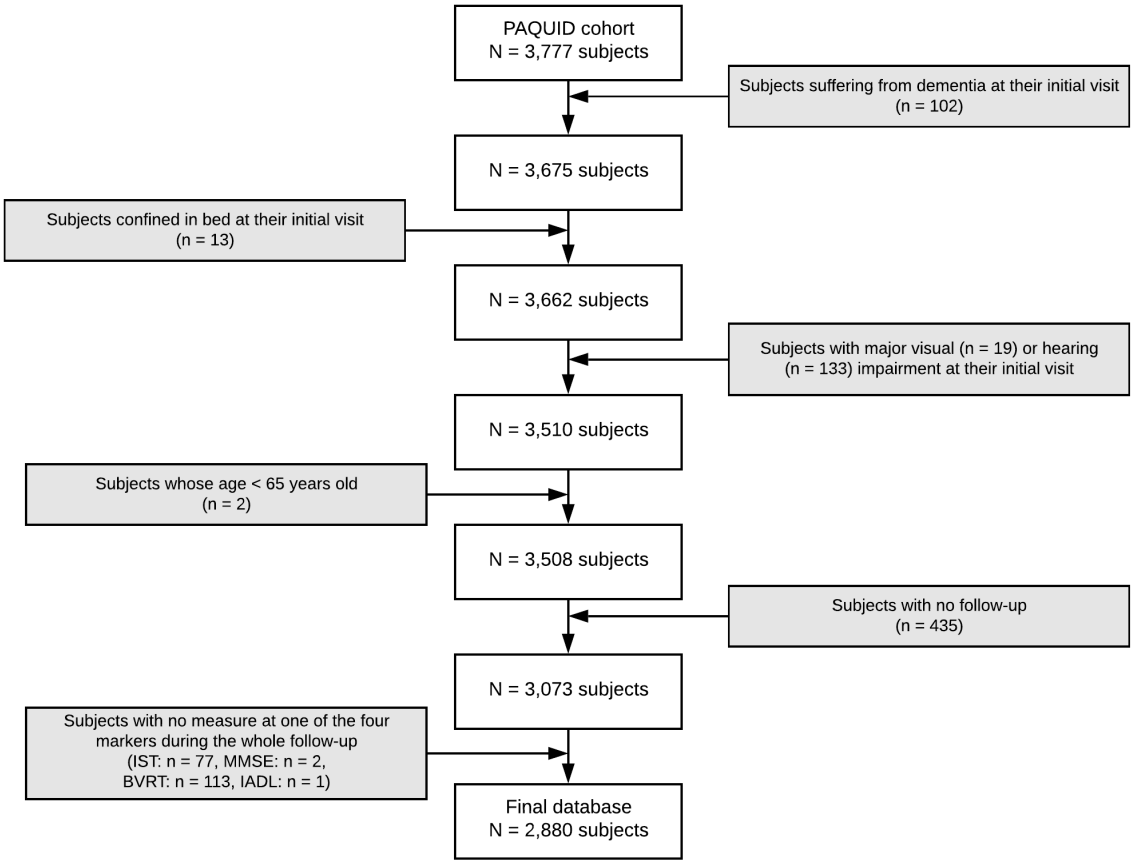
PAQUID cohort
N = 3,777 subjects
Subjects suffering from dementia at their initial visit
(n = 102)
N = 3,675 subjects
Subjects confined in bed at their initial visit
(n = 13)
N = 3,662 subjects
Subjects with major visual (n = 19) or hearing
(n = 133) impairment at their initial visit
N = 3,510 subjects
Subjects whose age < 65 years old
(n = 2)
N = 3,508 subjects
Subjects with no follow-up
(n = 435)
N = 3,073 subjects
Subjects with no measure at one of the four
markers during the whole follow-up
(IST: n = 77, MMSE: n = 2,
BVRT: n = 113, IADL: n = 1)
Final database
N = 2,880 subjects

Figure 1

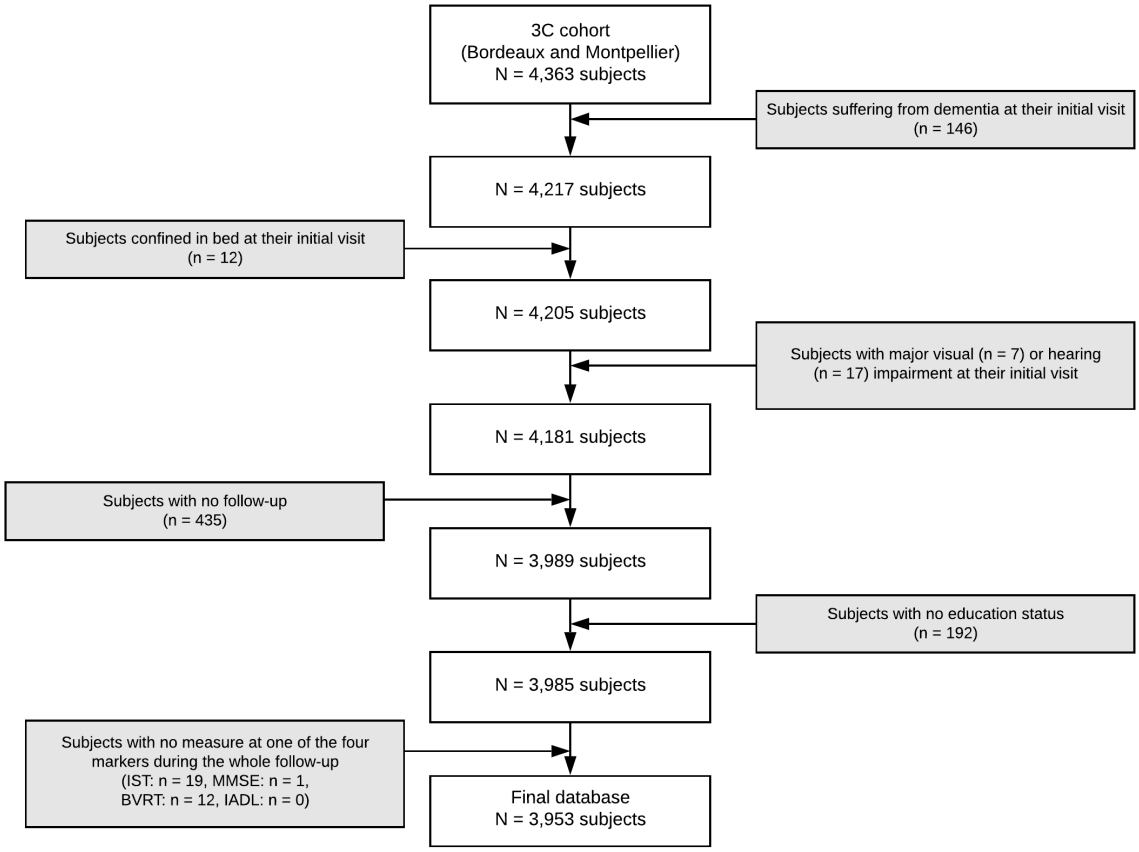
3C cohort
(Bordeaux and Montpellier)
N = 4,363 subjects
Subjects suffering from dementia at their initial visit
(n = 146)
N = 4,217 subjects
Subjects confined in bed at their initial visit
(n = 12)
N = 4,205 subjects
Subjects with major visual (n = 7) or hearing
(n = 17) impairment at their initial visit
N = 4,181 subjects
Subjects with no follow-up
(n = 435)
N = 3,989 subjects
Subjects with no education status
(n = 192)
N = 3,985 subjects
Subjects with no measure at one of the four
markers during the whole follow-up
(IST: n = 19, MMSE: n = 1,
BVRT: n = 12, IADL: n = 0)
Final database
N = 3,953 subjects

Figure 2

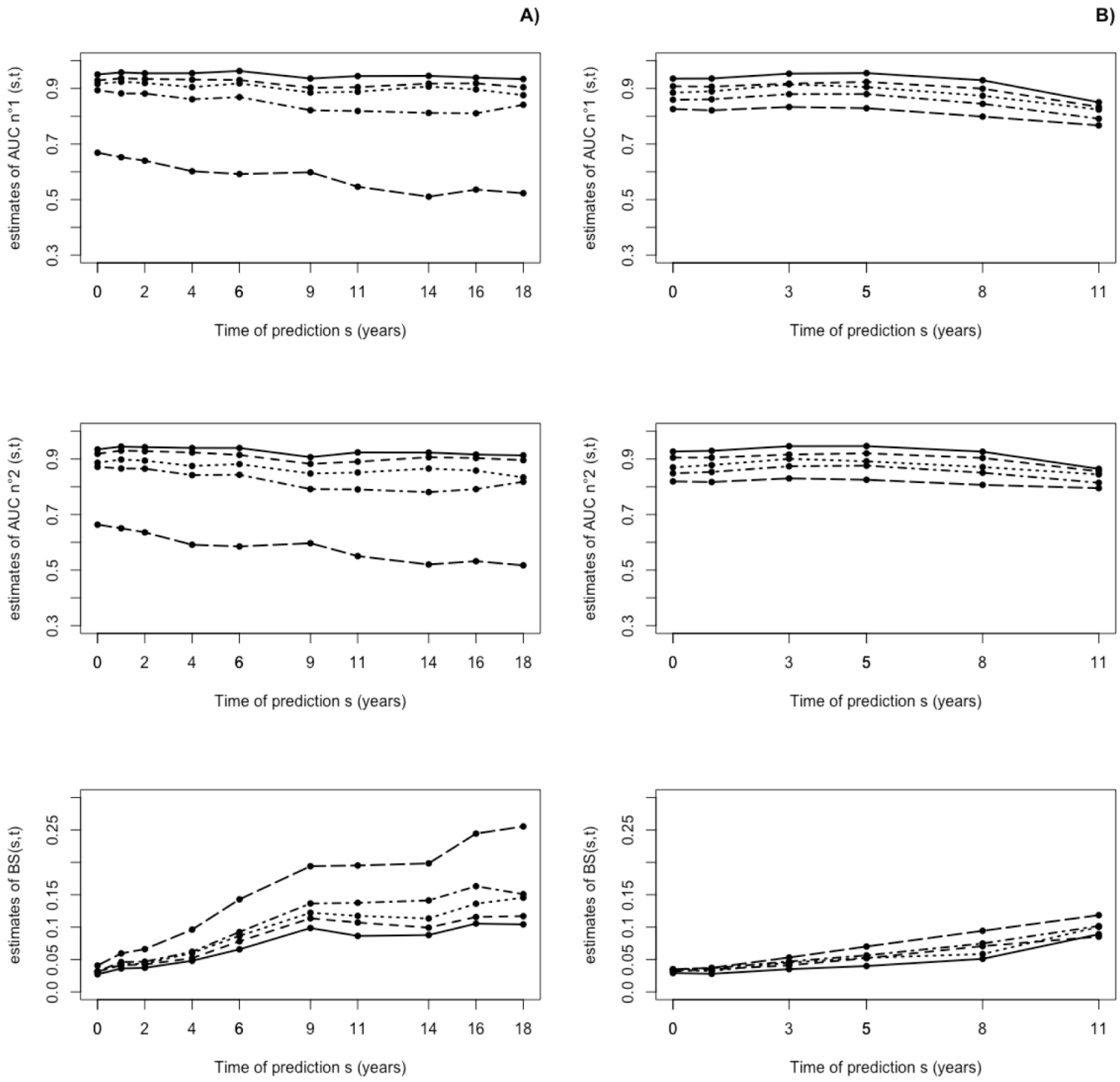
A)
B)
estimates of AUC n°1 (s,t)
estimates of AUC n°2 (s,t)
estimates of BS(s,t)
Time of prediction s (years)
0.3
0.5
0.7
0.9
0.0
0.05
0.1
0.15
0.25
0
2
4
6
9
11
14
16
18
3
5
8

Figure 3

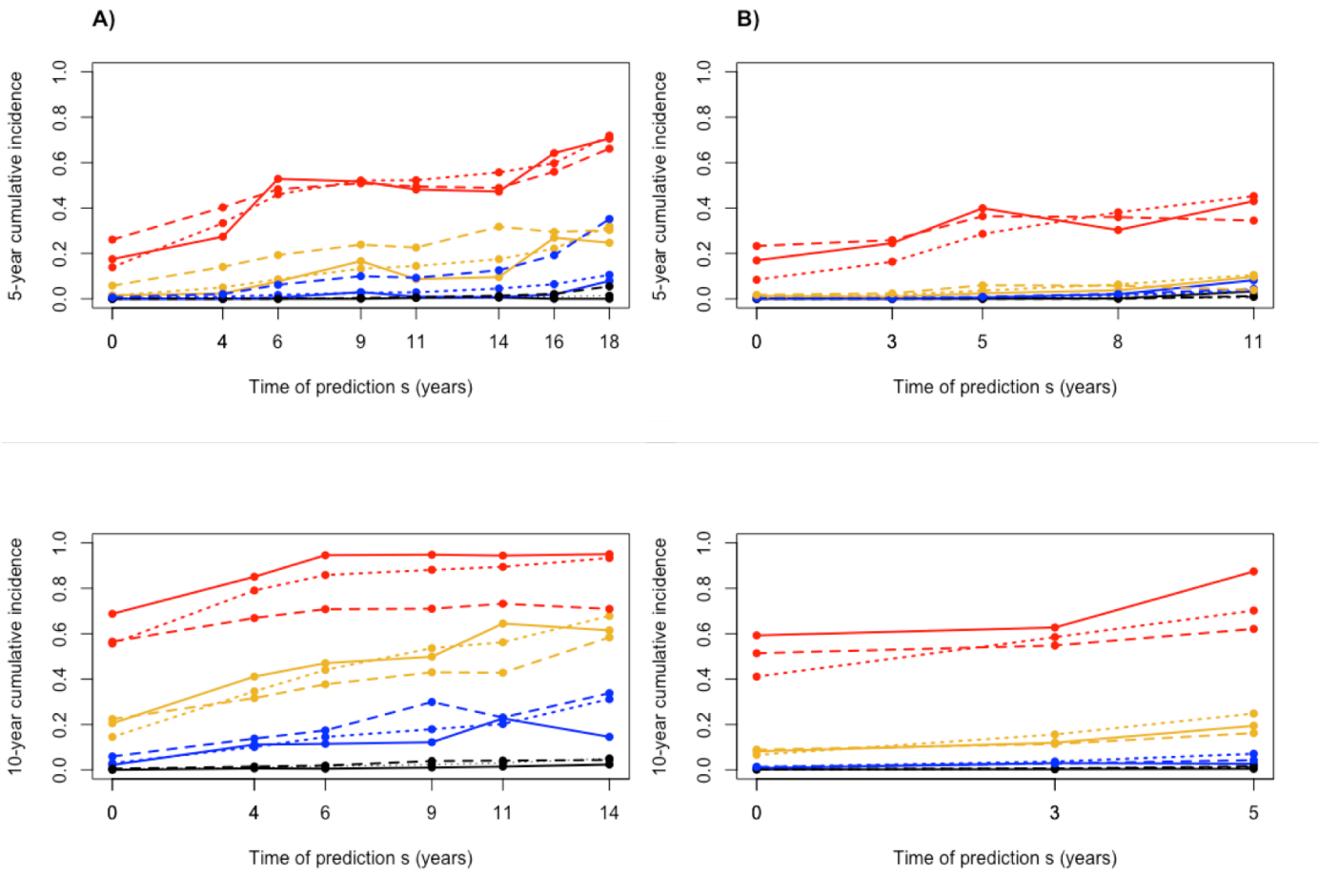
A)
B)
5-year cumulative incidence
10-year cumulative incidence
0.0
0.2
0.4
0.6
0.8
1.0
0
4
6
9
11
14
16
18
3
5
8
Time of prediction s (years)

Figure 4

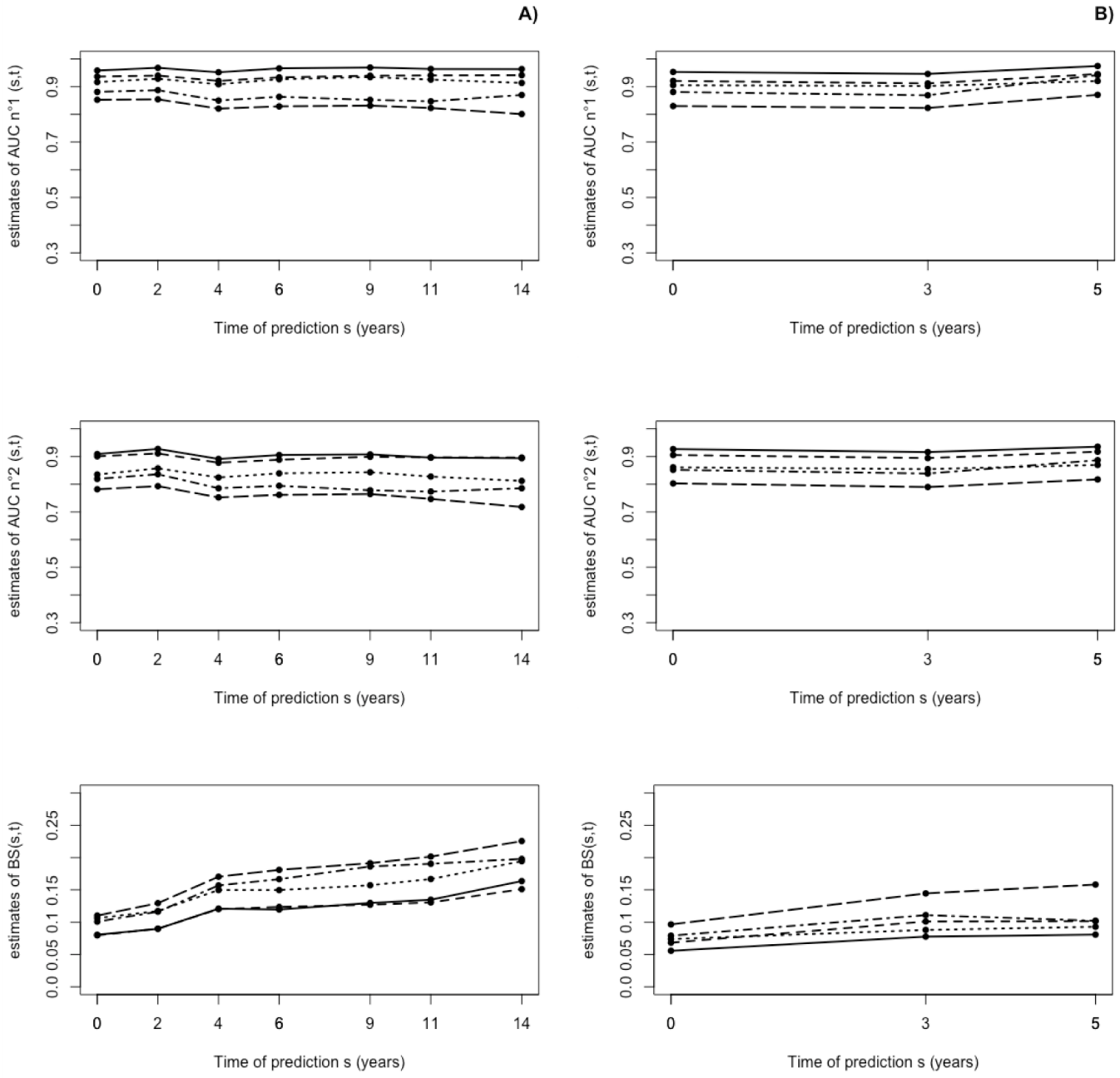
A)
B)
estimates of AUC n°1 (s,t)
estimates of AUC n°2 (s,t)
estimates of BS(s,t)
Time of prediction s (years)
0.3
0.5
0.7
0.9
0.0
0.05
0.1
0.15
0.25
0
2
4
6
9
11
14
3
5

Figure 5

Table 1. Description of the learning dataset (PAQUID) and the validation dataset (3C). IST: Isaacs Set Tests at 15 seconds; MMSE: Mini-Mental State Examination; BVRT: Benton Visual Retention Test; IADL: Instrumental Activities of Daily Living.

	PAQUID (N=2,880)	**3C (N=3,953)**
Sex		
Women	1,643 (57%)	2,354 (59.5%)
Education status		
With primary school diploma	1,979 (68.7%)	3,606 (91.2%)
Age at inclusion *(mean, SD)*	73.09 (5.89)	73.25 (5)
Cognitive scores at baseline *(mean, SD)*		
IST	27.25 (6.14)	29.48 (5.56)
MMSE	26.25 (2.97)	27.32 (1.97)
BVRT	10.27 (2.6)	11.59 (2.04)
Autonomy scale at baseline IADL4		
0	2,172 (75.4%)	3,552 (89.9%)
1	447 (15.5%)	313 (7.9%)
2	162 (5.6%)	53 (1.3%)
3	68 (2.4%)	20 (0.5%)
4	25 (0.9%)	1 (0.02%)
missing	6 (0.2%)	14 (0.4%)
Follow-up in years *		
Median and [IQR]	9.29 [3.68,14.95]	9.73 [5.64,15.4]
Number of repeated measures		
MMSE*:Median and [IQR]*	4 [3, 7]	5 [3, 5]
IADL*:Median and [IQR]*	4 [3, 7]	5 [3, 5]
IST*:Median and [IQR]*	4 [2, 6]	4 [3, 5]
BVRT*:Median and [IQR]*	4 [2, 6]	4 [3, 5]
Incident cases of dementia		
Number and (%)	813 (28.2%)	602 (15.2%)

* The follow-up time is defined as the time to dementia diagnosis or to the last cognitive assessment for censored subjects

Table 2. Estimates (with 95% credibility interval) of the time-to-event submodel of the joint models. PSD: primary school diploma; IST: Isaacs Set Tests; MMSE: Mini-Mental State Examination; BVRT: Visual Retention Test of Benton; IADL: Instrumental Activities of Daily Living.

	Unimarker				***3-markers***
	IST	***MMSE***	***BVRT***	***IADL4***	
Sex (Ref: men)	0.058 [-0.113, 0.242]	-0.35 [-0.592, -0.082]	-0.11 [-0.287, 0.068]	-0.61 [-0.894, -0.344]	-0.45 [-0.693, -0.204]
PSD vs no PSD	0.270 [0.091, 0.444]	0.874 [0.609, 1.15]	0.681 [0.477, 0.891]	0.454 [0.192, 0.716]	0.631 [0.382, 0.896]
Age in year	0.067 [0.053, 0.082]	0.042 [0.022, 0.059]	0.064 [0.051, 0.079]	-0.039 [-0.059, -0.020]	-0.024 [-0.043, -0.005]
IST_value	-1.437 [-1.598, -1.29]				-0.177 [-0.390, 0.038]
IST_slope	-0.870 [-1.18, -0.563]				-0.006 [-0.589, 0.643]
MMSE_value		-2.14 [-2.42, -1.91]			-0.832 [-1.10, -0.592]
MMSE_slope		-1.24 [-1.53, -0.97]			-0.175 [-0.704, 0.316]
BVRT_value			-1.59 [-1.81, -1.39]		
BVRT_slope			-0.458 [-0.797, -0.148]		
IADL4_value				2.83 [2.47, 3.27]	1.58 [1.27, 1.92]
IADL4_slope				2.24 [1.75, 2.83]	0.945 [0.560, 1.38]