



HAL
open science

A log-logistic survival model from multimodal data for prediction of Alzheimer's Disease

Pascal Lu, Olivier Colliot

► **To cite this version:**

Pascal Lu, Olivier Colliot. A log-logistic survival model from multimodal data for prediction of Alzheimer's Disease. SAFJR 2019 - Survival Analysis for Junior Researchers, Apr 2019, Copenhagen, Denmark. hal-02430943

HAL Id: hal-02430943

<https://inria.hal.science/hal-02430943>

Submitted on 7 Jan 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A log-logistic survival model from multimodal data for prediction of Alzheimer's Disease

Pascal Lu^{1,2} and Olivier Colliot^{1,2},

for the Alzheimer's Disease Neuroimaging Initiative

¹ Sorbonne Université, Inserm, CNRS, Institut du cerveau et la moelle (ICM), AP-HP - Hôpital Pitié-Salpêtrière, Boulevard de l'hôpital, 75013, Paris, France

² INRIA Paris, ARAMIS project-team, 75013, Paris, France

The early diagnosis of Alzheimer's disease (AD) is important for providing adequate care and is currently the topic of very active research. In particular, a current challenge is to predict the future occurrence of AD in patients with mild cognitive impairment. Multimodal data (such as cognitive/clinical, imaging and genetic) can provide complementary information for the prediction. Whereas clinical data, such as cognitive scores, provide an accurate estimate of the current subject's state, genetic variants help to identify whether a subject would develop Alzheimer's disease faster than another. In the state of the art, most papers put clinical and genetic variables on the same level in order to predict the current or future subject's state, although they do not provide the same type of information.

In this work, we propose a new survival model based on multimodal data to estimate the conversion date to AD from genetics and clinical data. We chose the log-logistic model which provides a parametric framework where the parameters depends on both clinical and genetic data. The hazard function is unimodal, which seems well-suited to our model. In our proposed formulation, genetic data \mathbf{x}_G only influences the speed $v(\mathbf{x}_G)$ at which the conversion would happen, whereas clinical data \mathbf{x}_C influences the initial state of the subject $p(\mathbf{x}_C)$. If S denotes the survival function and $t_{1/2}$ the median survival time, we set $t_{1/2} = p(\mathbf{x}_C)/v(\mathbf{x}_G)$, and $S'(t_{1/2}) = -v(\mathbf{x}_G)$. By determining $v(\mathbf{x}_G)$ and $p(\mathbf{x}_C)$, we are able to determine the associated survival function S .

We tested our model on the ADNI-1 dataset (from the Alzheimer's Disease Neuroimaging Initiative), using cognitive scores (such as MMSE, ADAS13, RAVLT) and genetic information (APOE, gender) and compared it to the Cox-regression model using the Kaplan-Meier estimate, and the classical parametric log-logistic model where the parameters depends on the covariates using a log-linear model.

E-mail: pascal.lu@inria.fr

Keywords: Log-logistic model, multimodal data, Alzheimer's Disease

References:

R. Marinescu, et al, TADPOLE Challenge: Prediction of Longitudinal Evolution in Alzheimer's Disease, <https://arxiv.org/abs/1805.03909>

A log-logistic survival model from multimodal data for prediction of Alzheimer's Disease

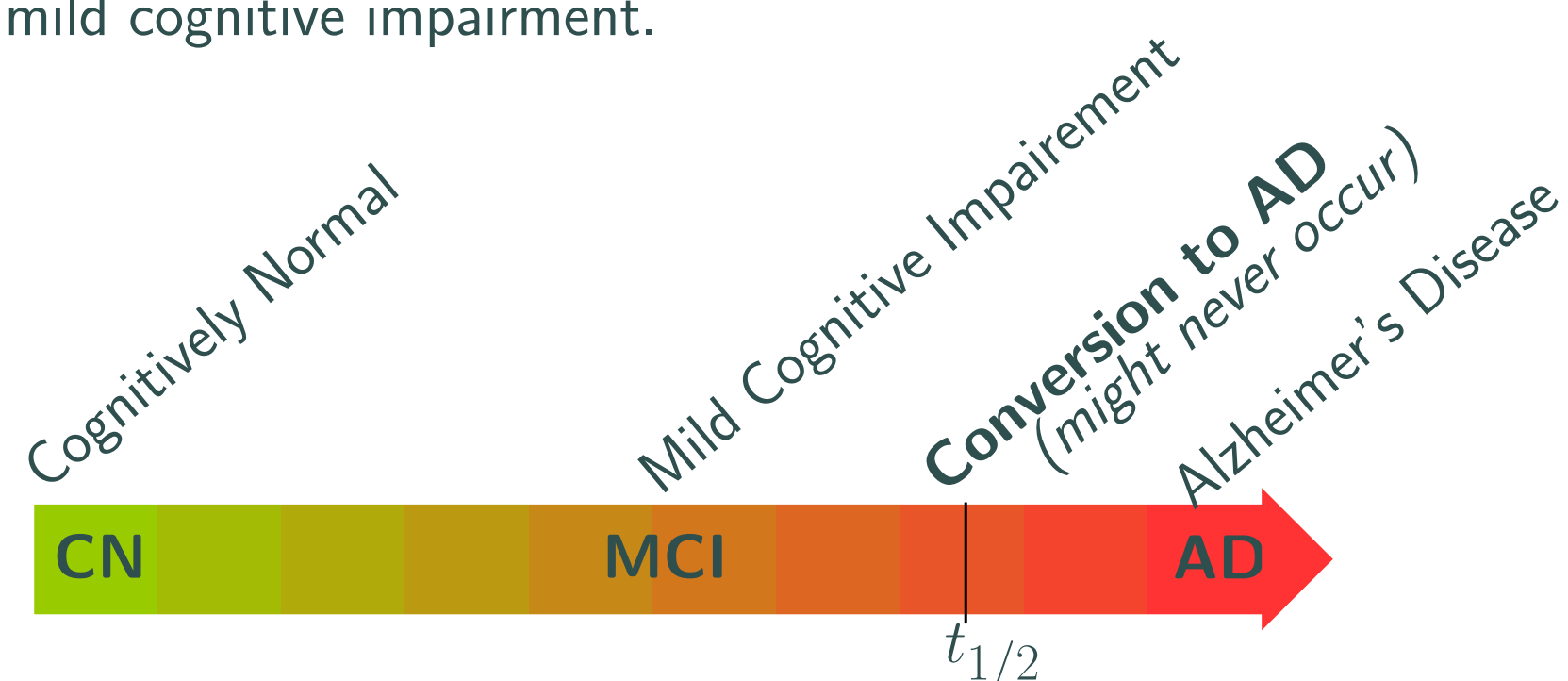
Pascal LU¹, Olivier COLLIOT¹, the Alzheimer's Disease Neuroimaging Initiative

pascal.lu@inria.fr, olivier.colliot@upmc.fr

¹ ARAMIS project-team, Sorbonne Université, INSERM, CNRS, INRIA Paris, Institut du cerveau et la moëlle épinière (ICM), Hôpital Pitié-Salpêtrière, Boulevard de l'hôpital, F-75013 Paris, France

Motivation

The early diagnosis of Alzheimer's disease (AD) is important for providing adequate care and is currently the topic of very active research. In particular, a current challenge is to predict the future occurrence of AD in patients with mild cognitive impairment.



Multimodal data (such as clinical and genetic) provides complementary information for the prediction. Whereas clinical data, such as cognitive scores, provide an accurate estimate of the current subject's state, genetic variants help to identify whether a subject would develop AD faster than another. In the state of the art, most papers put clinical and genetic variables on the same level in order to predict the current or future subject's state, although they do not provide the same type of information.

Survival function depending on alleles of gene APOE We ran the Kaplan-Meier estimator on different groups of subject depending of their APOE status. Median survival time (when the survival function is 0.5) is 66 months for APOE 0, 36 months for APOE 1, and 24 months for APOE 0.

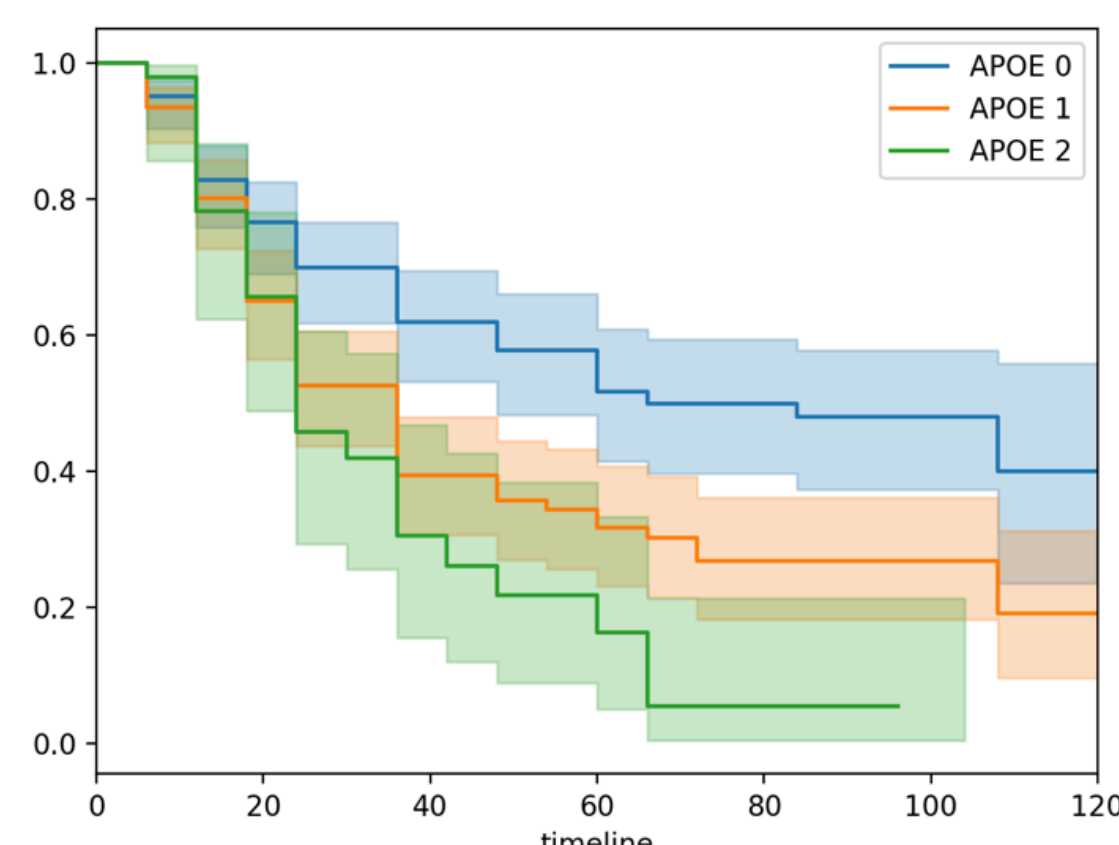


Fig 1. Estimator of Kaplan-Meier of the survival function $S(t) = \mathbb{P}\{T > t | \text{APOE}\}$

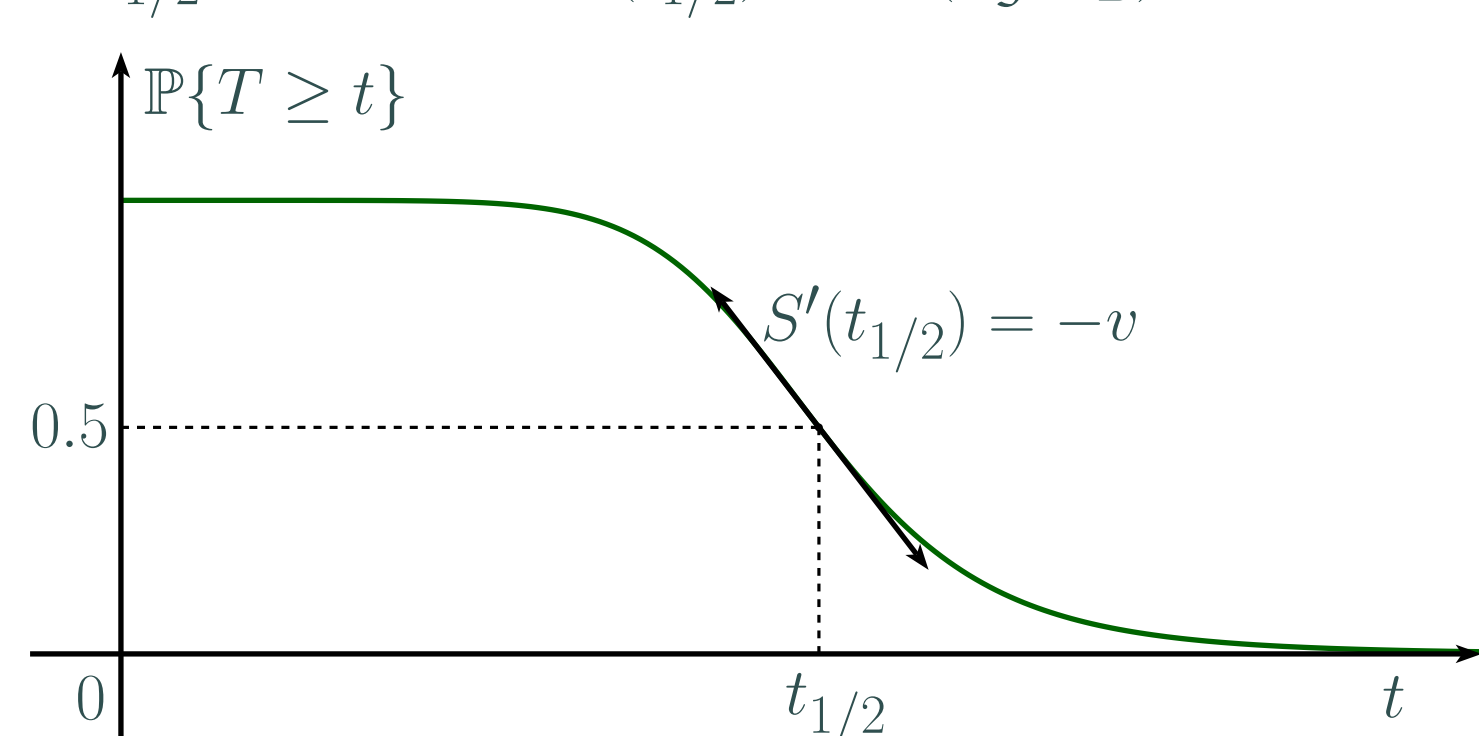
Model set-up

The log-logistic survival function has the following form, for covariates \mathbf{x} :

$$S(t|\mathbf{x}) = \frac{1}{1 + \lambda(\mathbf{x})t^{\alpha(\mathbf{x})}} \text{ with } \lambda(\mathbf{x}) > 0 \text{ and } \alpha(\mathbf{x}) > 0$$

Let \mathbf{x}_G be the genetic covariates and \mathbf{x}_I be the clinical covariates. We define the following variables of speed $v(\mathbf{x}_G, \mathbf{x}_I)$ and of position $p(\mathbf{x}_G, \mathbf{x}_I)$.

- We chose to set the median time $t_{1/2}$ (the time when $S(t) = \frac{1}{2}$) such that $t_{1/2} = \frac{p(\mathbf{x}_G, \mathbf{x}_I)}{v(\mathbf{x}_G, \mathbf{x}_I)}$.
- and at the median time $t_{1/2}$, the slope is $S'(t_{1/2}) = -v(\mathbf{x}_G, \mathbf{x}_I)$



Since $\alpha = -4t_{1/2}S'(t_{1/2})$ et $\lambda = \frac{1}{t_{1/2}^{\alpha}}$, then the survival function is given by:

$$S(t|\mathbf{x}_G, \mathbf{x}_I) = \mathbb{P}\{T \geq t | \mathbf{x}_G, \mathbf{x}_I\} = \frac{1}{1 + \exp\left(4p(\mathbf{x}_G, \mathbf{x}_I)\left(\ln t - \ln\left(\frac{p(\mathbf{x}_G, \mathbf{x}_I)}{v(\mathbf{x}_G, \mathbf{x}_I)}\right)\right)\right)}$$

under the conditions $v > 0$ (all subjects convert) and $p > 0$ (conversion date is non-negative).

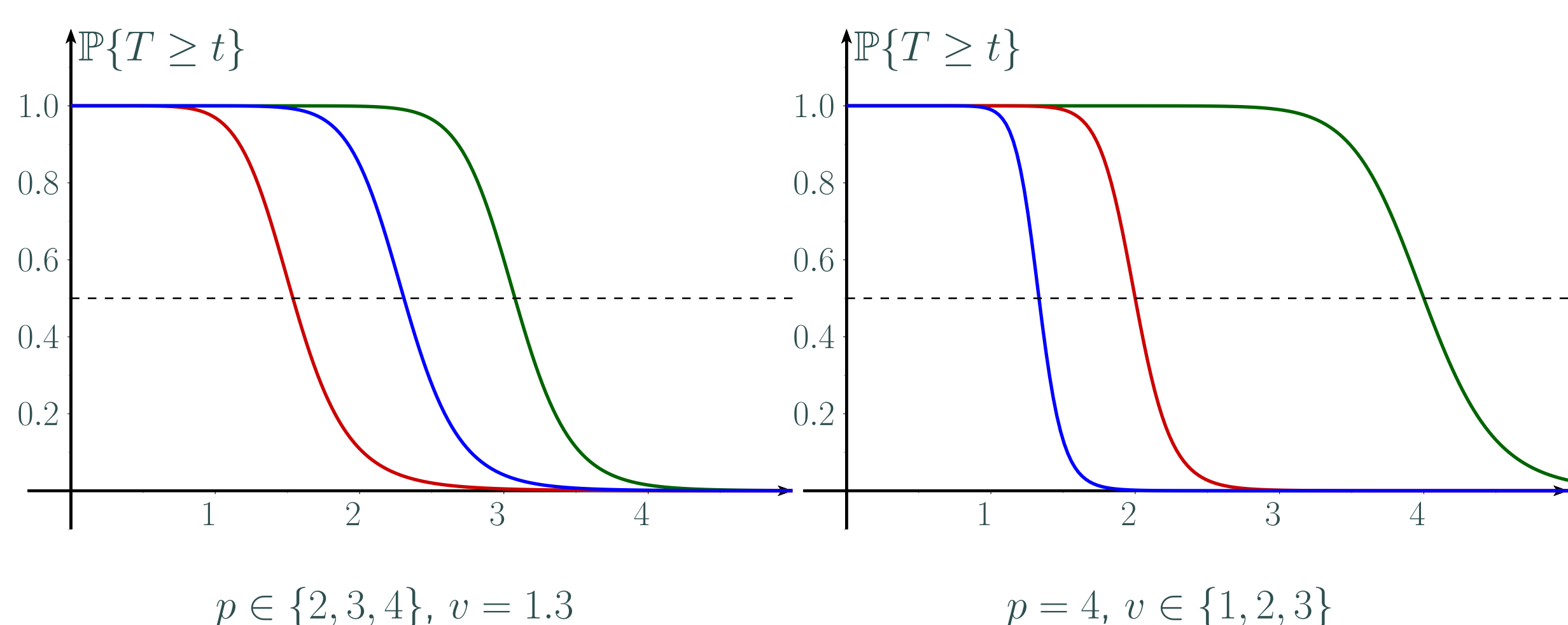


Fig 2. Effect of p and v on $S(t)$. It can be seen that p only translates the curve of $S(t)$ (for fixed v), whereas v changes both the shape and translation of the curve of $S(t)$

We made the assumptions that the speed depends on genetic and clinical data, while the position depends only on clinical data:

$$p(\mathbf{x}_G, \mathbf{x}_I) = \beta_I^T \mathbf{x}_I \text{ and } v(\mathbf{x}_G, \mathbf{x}_I) = \gamma_G^T \mathbf{x}_G$$

Parameters $\beta_I, \gamma_I, \gamma_G$ are learnt by minimizing the negative log-likelihood using the SLSQP algorithm under the conditions $p(\mathbf{x}_G^i, \mathbf{x}_I^i) > 0$ and $v(\mathbf{x}_G^i, \mathbf{x}_I^i) > 0$ for all subjects i .

Results

We selected the following covariates:

- for genetic data: the allele for gene APOE and the gender (man=0, woman=1),
- for clinical data: the education, MMSE (at baseline), ADAS 13 (at baseline), RAVLT (at baseline), age (at baseline).

from the ADNI1 dataset (<http://adni.loni.usc.edu>), for 325 MCI patients.

We compared the model with the Cox PH model, where the hazard and survival functions are resp. given by:

$$h(t) = h_0(t)e^{\beta_I^T \mathbf{x}_I + \beta_G^T \mathbf{x}_G} \text{ and } S(t) = (S_0(t))^{e^{\beta_I^T \mathbf{x}_I + \beta_G^T \mathbf{x}_G}}$$

Effects of covariates

- For the Cox PH model, $t_{1/2}$ decreases only if $\text{PI}(\mathbf{x}_I, \mathbf{x}_G) = \beta_I^T \mathbf{x}_I + \beta_G^T \mathbf{x}_G$ increases.
- For our model, $t_{1/2}$ decreases only if $p(\mathbf{x}_G, \mathbf{x}_I)$ decreases or if $v(\mathbf{x}_G, \mathbf{x}_I)$ increases.

This following table shows the *normalized* coefficients learnt by a Cox model and by our model. \nearrow means that the higher this covariates is, the risk to have AD increases; whereas \searrow means that the higher this covariates is, the risk to have AD decreases.

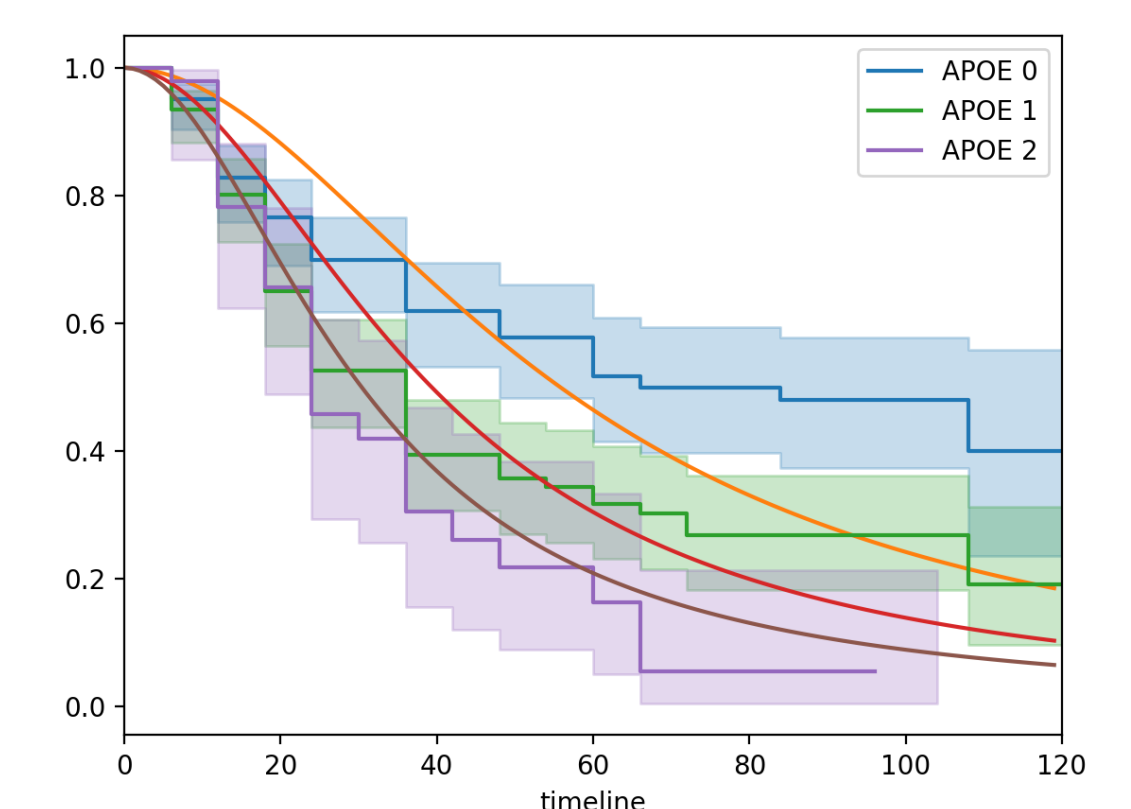
Covariate	Our model		Cox PH Model
	Coeff β	Coeff γ	Coeff β_I, β_G
MMSE (at baseline) \searrow	0.0080		-0.12
ADAS13 (at baseline) \nearrow	-0.059		0.35
Age \nearrow	0.0059		-0.020
Education	-0.0074		0.16
RAVLT (at baseline) \searrow	0.092		-0.49
Intercept (clinics)	0.38		
APOE4 \nearrow		0.0018	0.20
Gender \nearrow		0.00044	0.18
Intercept (genetics)		0.010	

Assessments on metrics We performed a 10-fold cross-validation to assess models under the following metrics.

	Our Model	Cox PH Model
Median survival time (months)	39	36
C-index	0.637	0.648
iAUC	0.81	0.84
Brier	0.14	0.14

Evaluation on group based on APOE status We assessed the model on groups of subject depending of their APOE status.

	Median survival time	
	Our Model	Kaplan-Meier
All (months)	39	36
APOE 0 (months)	56	66
APOE 1 (months)	39	36
APOE 2 (months)	30	24



Truncation Instead of learning the model on the original dataset, subjects with long conversion time can be truncated. More specifically, let τ be the follow-up date. If the subject i 's conversion is $T_i > \tau$, then we set his observed date is τ , and the conversion did not occur.

The effect of truncation makes the prediction much more accurate for small conversion date, but more inaccurate for long conversion date.

Conclusion and future work

It's a novel approach to **integrate genetic and clinical data** for prediction of conversion date to AD. Although it shows slightly lower performances compared to the Cox PH model, further work on speed by adding clinical covariates could improve the model. Clinical covariates could be added for instance, in an additive framework ($v(\mathbf{x}_G, \mathbf{x}_I) = \gamma_I^T \mathbf{x}_I + \gamma_G^T \mathbf{x}_G$).

References

- R. Marinescu, et al, TADPOLE Challenge: Prediction of Longitudinal Evolution in Alzheimer's Disease, <https://arxiv.org/abs/1805.03909>