frontiers

Model-free prognosis and decision in personalized medicine: A case study on the conversion from Mild Cognitive Impairment to Alzheimer's Disease

P.G.L. Porta Mana [1,2,*], **Ingrid Rye** [3], **Alexandra Vik** [1,2], **Marek Kocinski** [2,4], **Arvid Lundervold** [2,4], **Astri J. Lundervold** [3], **and Alexander S. Lundervold** [1,2]

[1] *Department of Computer Science, Electrical Engineering and Mathematical Sciences, Western Norway University of Applied Sciences, Bergen, Norway*
[2] *Mohn Medical Imaging and Visualization Centre (MMIV), Department of Radiology, Haukeland University Hospital, Bergen, Norway*
[3] *Department of Biological and Medical Psychology, University of Bergen, Norway*
[4] *Department of Biomedicine, University of Bergen, Norway*

Correspondence*:
Corresponding Author
pgl@portamana.org

ABSTRACT

[Luca & Astri, draft]

Patients with Mild Cognitive Impairment (MCI) have an increased risk of a trajectory toward Alzheimer's disease (AD). Early identification of patients with a high risk of AD is essential to provide treatment before the disease is well-established in the brain. It is, therefore, of great importance to study how well different kinds of predictors allow us to estimate a trajectory from MCI towards AD in an individual patient.

But more is needed for a personalized approach to prognosis, prevention, and treatment, than just the obvious requirement that prognoses be as best as they can be for each patient. Several situational elements that can be different from patient to patient must be accounted for:

- the *kinds* of clinical data and evidence available for prognosis;
- the *outcomes* of the same kind of clinical data and evidence;
- the kinds of treatment or prevention strategies available, owing to different attitudes toward life, different family networks and possibilities of familial support, different additional medical factors such as physical disabilities, and different economic means;
- the advantages and disadvantages, benefits and costs of the same kinds of treatment or prevention strategies; the patient has a major role in the quantification of such benefits and costs;
- finally, the initial evaluation by the clinician – which often relies on too subtle clues (family history, regional history, previous case experience) to be considered as measurable data.

Statistical decision theory is the normative quantification framework that takes into account these fundamental differences. Medicine has the distinction of having been one of the first fields to adopt this framework, exemplified in brilliant old and new textbooks on clinical decision-making.

Clinical decision-making makes allowance for these differences among patients through two requirements. First, the quantification of prognostic evidence on one side, and of benefits and costs of treatments and prevention strategies on the other, must be clearly separated and handled in a modular way. Two patients can have the same prognostic evidence and yet very different prevention options. Second, the quantification of independent prognostic evidence ought to be in the form of *likelihoods about the health condition* (or equivalently of likelihood ratios, in a binary case), that is, of the probabilities of the observed test outcomes given the hypothesized health conditions. Likelihoods from independent clinical tests and predictors can then be combined with a simple multiplication; for one patient, we could have three kinds of predictor available; for another, we could have five. The clinician's pre-test assessment is included in the form of a probability. These patient-dependent probabilities are combined with the patient-dependent costs and benefits of treatment or prevention to arrive at the best course of action for that patient. The main result underlying statistical decision theory is that decision-making *must* take this particular mathematical form in order to be optimal and logically consistent.

The present work investigates the prognostic power of a set of neuropsychological and Magnetic Resonance Imaging examinations, demographic data, and genetic information about Apolipoprotein-E4 (APOE) status, for the prediction of the onset of Alzheimer's disease in patients defined as mildly cognitively impaired at a baseline examination. The longitudinal data used come from the ADNI database.

The prognostic power of these predictors is quantified in the form of a combined likelihood for the onset of Alzheimer's disease. As a hypothetical example application of personalized clinical decision making, three patient cases are considered where a clinician starts with prognostic uncertainties, possibly coming from other tests, of 50%/50%, 25%/75%, 75%/25%. It is shown how these pre-test probabilities are changed by the predictors.

This quantification also allows us to rank the relative prognostic power of the predictors. It is found that several neuropsychological examinations have the highest prognostic power, much higher than the genetic and imaging-derived predictors included in the present set.

Several additional advantages of this quantification framework are also exemplified and discussed in the present work:

- missing data are automatically handled, and results having partial data are not discarded; this quantification, therefore, also accounts for patient-dependent availability of *non-independent* predictors;
- no modelling assumptions (e.g., linearity, gaussianity, functional dependence) are made;
- the prognostic power obtained is intrinsic to the predictors, that is, it is a bound for *any* prognostic algorithm;
- variability ranges of the results owing to the finite size of the sample data are automatically quantified.
- the values obtained, being probabilities, are more easily interpretable than scores of various kinds.

Keywords: keyword, keyword, keyword, keyword, keyword, keyword, keyword, keyword

1 EACH PATIENT IS UNIQUE

Meet Olivia, Ariel, Bianca, Curtis.[1] These four persons don't know each other, but they have something in common: they all suffer from a mild form of cognitive impairment, and are afraid that their impairment will turn into Alzheimer's disease within a couple of years. In fact, this is why they recently underwent some clinical analyses and cognitive tests. Today they got the results of their analyses. From these results and other demographic factors their clinician is assessing their risk of developing Alzheimer, and will then decide on possible preventive treatments together with the patients.

Besides this shared condition and worry, these patients have other things in common – but also some differences. Let's take Olivia as reference and list the similarities and difference between her and the other three:

- Olivia and Ariel turn out to have exactly identical clinical results and age. They would also get similar benefits from the available preventive-treatment options. Ariel, however, comes from a different geographical region from Olivia, where the conversion rate from Mild Cognitive Impairment to Alzheimer is lower: about 30%, whereas in Olivia's it is around 45% **?** Moreover there is no history of Alzheimer in Ariel's family.
- Olivia and Bianca also have exactly the same clinical results and age. They come from the same geographical region and have very similar family histories. In fact we shall see that they have the same probability of developing Alzheimer's disease. Bianca, however, suffer from several allergies that would render some of the preventive options less beneficial to her.
- Olivia and Curtis have different clinical results and age – in particular, Olivia has the risky APOE4 allele whereas Curtis hasn't, and Curtis is more than 20 years younger. But they otherwise come from the same geographical region, have very similar family histories, and would get similar benefits from the preventive options

We can categorize these differences as "difference in auxiliary information" (Olivia and Ariel), "difference in preventive benefits" (Olivia and Bianca), "difference in clinical factors" (Olivia and Curtis). Figure *** summarizes the similarity and differences between Olivia and the other three patients. Table *** reports the clinical results and demographic data common to Olivia, Ariel, Bianca, as well as those of Curtis.

Table 1. Clinical results & demographic data

	Age	Sex	HC	APOE	ANART	CFT	GDS	RAVLT-im	RAVLT-del	RAVLT-rec	TMTA	TMTB
Olivia, Ariel, Bianca	83.5	F	0.00277	Y	14	18	0	29	12	3	126.5	117.0
Curtis	61.5	M	0.00417	N	6	13	1	16	2	0	24.8	83.4

We will soon see that despite the many factors in common among these four patients – even despite the identical clinical results for Olivia, Ariel, Bianca, and despite the identical risk for Olivia and Bianca – *the optimal preventive option for each patient is different from those of the other three*. This difference arises from the difference in clinical results, or auxiliary information, or preventive benefits.

[1] Fictive characters; any reference to real persons is purely coincidental

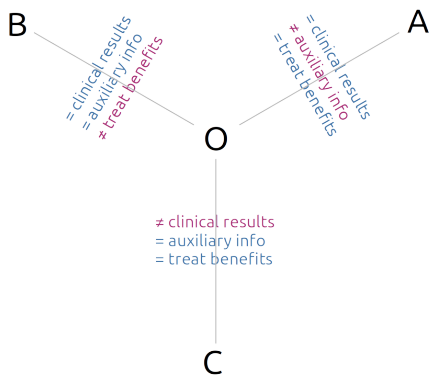

Figure 1. draft

— Luca, old pieces of text —

Personalized diagnosis, prognosis, treatment, and prevention strategies must make allowance for several fundamental differences among patients:

- the *kinds* of clinical data and evidence available for diagnosis or prognosis can be different;
- the *values* of the same kind of clinical data and evidence can be different;
- the kinds of treatment or prevention options can be different;
- the advantages and disadvantages, benefits and costs of the same kinds of treatment or prevention can be different;
- finally, the evaluation of the clinician – which often relies on too subtle clues (family history, regional history, case experience) to be considered as measurable data – can be different.

Is there really a methodological framework that can take all these differences into account? Yes, there is, and Medicine has the distinction of having been one of the first fields to adopt it (**?**): Statistical Decision Theory. Its application in Medicine is explained and exemplified in several, brilliant, old and new textbooks (**???**). This theory has mathematical and logical foundations and its principles constitute indeed the foundations for the definition and realization of Artificial Intelligence (**?**)

The basics of clinical decision making ..basics: each piece of evidence contributes with a likelihood or odds; they combine together and together with the clinician's pre-data evaluation. Then they are combined with the different benefits/costs of treatments or prevention strategies to find the optimal one. Decision trees can be necessary (but don't change this framework). Costs & benefits are evaluated by clinician & patient together.

$$\overbrace{\mathrm{p}(\text{health condition} \mid \text{results of all tests},\ \text{prior info})}^{\textit{post-test probability}} \propto \overbrace{\mathrm{p}(\text{health condition} \mid \text{prior info})}^{\textit{pre-test probability by clinician}} \times \textit{likelihoods of tests} \left\{ \begin{array}{l} \mathrm{p}(\text{result of 1st test} \mid \text{health condition},\ \text{prior info}) \times \\ \mathrm{p}(\text{result of 2nd test} \mid \text{health condition},\ \text{prior info}) \times \\ \ldots \end{array} \right. \quad (1)$$

For Original Research Articles (**?**), Clinical Trial Articles (**?**), and Technology Reports (**?**), the introduction should be succinct, with no subheadings (**?**). For Case Reports the Introduction should include symptoms at presentation (**?**), physical exams and lab results (**?**).

2 ARTICLE TYPES

For requirements for a specific article type please refer to the Article Types on any Frontiers journal page. Please also refer to Author Guidelines for further information on how to organize your manuscript in the required sections or their equivalents for your field

3 MANUSCRIPT FORMATTING

3.1 Heading Levels

3.2 Level 2

3.2.1 Level 3

3.2.1.1 Level 4

3.2.1.1.1 Level 5

3.3 Equations

Equations should be inserted in editable format from the equation editor.

$$\sum x + y = Z \quad (2)$$

3.4 Figures

Frontiers requires figures to be submitted individually, in the same order as they are referred to in the manuscript. Figures will then be automatically embedded at the bottom of the submitted manuscript. Kindly ensure that each table and figure is mentioned in the text and in numerical order. Figures must be of sufficient resolution for publication see here for examples and minimum requirements. Figures which are

not according to the guidelines will cause substantial delay during the production process. Please see here for full figure guidelines. Cite figures with subfigures as figure **??** and **??**.

3.4.1 Permission to Reuse and Copyright

Figures, tables, and images will be published under a Creative Commons CC-BY licence and permission must be obtained for use of copyrighted material from other sources (including re-published/adapted/modified/partial figures and images from the internet). It is the responsibility of the authors to acquire the licenses, to follow any citation instructions requested by third-party rights holders, and cover any supplementary charges.

3.5 Tables

Tables should be inserted at the end of the manuscript. Please build your table directly in LaTeX.Tables provided as jpeg/tiff files will not be accepted. Please note that very large tables (covering several pages) cannot be included in the final PDF for reasons of space. These tables will be published as Supplementary Material on the online article page at the time of acceptance. The author will be notified during the typesetting of the final article if this is the case.

4 NOMENCLATURE

4.1 Resource Identification Initiative

To take part in the Resource Identification Initiative, please use the corresponding catalog number and RRID in your current manuscript. For more information about the project and for steps on how to search for an RRID, please click here.

4.2 Life Science Identifiers

Life Science Identifiers (LSIDs) for ZOOBANK registered names or nomenclatural acts should be listed in the manuscript before the keywords. For more information on LSIDs please see Inclusion of Zoological Nomenclature section of the guidelines.

5 ADDITIONAL REQUIREMENTS

For additional requirements for specific article types and further information please refer to Author Guidelines.

CONFLICT OF INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

The Author Contributions section is mandatory for all articles, including articles by sole authors. If an appropriate statement is not provided on submission, a standard one will be inserted during the production process. The Author Contributions statement must describe the contributions of individual authors referred

to by their initials and, in doing so, all authors agree to be accountable for the content of the work. Please see here for full authorship criteria.

FUNDING

Details of all funding sources should be provided, including grant numbers if applicable. Please ensure to add all necessary funding information, as after publication this is no longer possible.

ACKNOWLEDGMENTS

This is a short text to acknowledge the contributions of specific colleagues, institutions, or agencies that aided the efforts of the authors.

SUPPLEMENTAL DATA

Supplementary Material should be uploaded separately on submission, if there are Supplementary Figures, please include the caption in the same file as the figure. LaTeX Supplementary Material templates can be found in the Frontiers LaTeX folder.

DATA AVAILABILITY STATEMENT

The datasets [GENERATED/ANALYZED] for this study can be found in the [NAME OF REPOSITORY] [LINK].

FIGURE CAPTIONS

Figure 2. Enter the caption for your figure here. Repeat as necessary for each of your figures

”