ELSEVIER

Alzheimer's & Dementia ■ (2018) 1-3

Alzheimer's & Dementia

Letter

Magnetic resonance imaging-based hippocampus volume for prediction of dementia in mild cognitive impairment: Why does the measurement method matter so little?

Magnetic resonance imaging (MRI)-based hippocampus volume (HV) is the best established imaging marker to support the prediction of AD dementia (ADD) in mild cognitive impairment (MCI), although its utility in clinical patient care has not yet been fully demonstrated [1,2]. HV can be scored on an ordinal scale based on visual inspection of MRI [3], or it can be estimated quantitatively by manual or automatic delineation of the hippocampus in MRI. While visual scoring tends to perform worse in MCI-to-ADD prediction, manual delineation and automatic methods show very similar performance [4]. Furthermore, there is hardly any difference among the numerous automatic methods with respect to predictive power in MCI. In the head-to-head comparison of four HV measurement methods in MCI subjects of the Alzheimer's Disease Neuroimaging Initiative by the European Medicines Agency, the area (AUC) under the receiver operating characteristic (ROC) curve for 2-year prediction of ADD ranged between 0.7290 and 0.7565, and among three of the four methods, the AUC ranged between 0.7516 and 0.7565 [5]. This appears surprising at first sight given that the quantitative methods differ strongly in accuracy and precision with respect to the anatomical delineation of the hippocampus. Time spent by the rater and computer processing time also differ strongly [4]. Here, we aim to provide a simple mathematical explanation of the stability of the performance of MRI-based HV in MCI with respect to the HV measurement method.

Let us assume that the true, error-free HV follows a Gaussian distribution in both MCI stable subjects and in MCI-to-ADD progressors:

$$\begin{aligned} \underline{S}\text{tables : trueHV} &\sim \mathrm{N}\left(\mu_S, \sigma_S^2\right) \\ \underline{P}\text{rogressors : trueHV} &\sim \mathrm{N}\left(\mu_P, \sigma_P^2\right) \end{aligned} \tag{1}$$

where $N(\mu,\sigma^2)$ is the Gaussian with mean μ and variance σ^2. The difference $\mu_S - \mu_P$ describes the biological difference of the true HV between MCI stables and MCI-to-ADD progressors, the variance σ^2 describes the (patho)physiological variability of the true HV between subjects.

The ROC curve for prediction of MCI-to-ADD progression is obtained by plotting the true positive rate q versus the false positive rate p for all possible thresholds c, where

$$q = \Phi\left(\frac{\mu_S - c}{\sigma_S}\right) \text{and } p = \Phi\left(\frac{\mu_P - c}{\sigma_P}\right) \tag{2}$$

and Φ is the standard normal cumulative distribution function [6]. The AUC is given by [6].

$$\mathrm{AUC} = \Phi\left(\frac{\mu_S - \mu_P}{\sqrt{\sigma_S^2 + \sigma_P^2}}\right) \tag{3}$$

To start with, let us assume

$$\begin{aligned} \mu_S &= 3.0 \text{ ml} \\ \mu_P &= 2.7 \text{ ml} \\ \sigma_S = \sigma_P = \sigma_{true} &= 0.3 \text{ ml} \end{aligned} \tag{4}$$

The AUC of MCI-to-ADD prediction by the true HV is 0.7602 in this scenario (Fig. 1A), close to the European Medicines Agency results cited previously.

Now let us consider HV estimates obtained by a given measurement method. The measurement process causes measurement errors that result in additional intersubject variability (a systematic offset of the HV estimates can be accounted for by linear transformation). Let us model this additional variability by a Gaussian with mean zero and variance σ^2_{meas}. The measured HV estimates than follow the same Gaussian distribution as the true HV but with increased variance

$$\sigma^2_{est} = \sigma^2_{true} + \sigma^2_{meas} \tag{5}$$

The red curve in Fig. 1B shows the AUC for prediction of MCI-to-ADD progression by HV estimates as function of σ_{meas} (scaled to $\mu_S - \mu_P$). The additional variability by the measurement error causes only mild AUC decrease. Even when the standard deviation of the measurement error reaches the difference $\mu_S - \mu_P$ of mean true HV between MCI stables and MCI-to-ADD progressors, the AUC only slightly decreases to 0.6915 (from AUC = 0.7602 with true, error-free HV), which still lies within the typically observed performance range.

*Corresponding author. Tel.: +4940741054347; Fax: +4940741040265.
E-mail address: r.buchert@uke.de

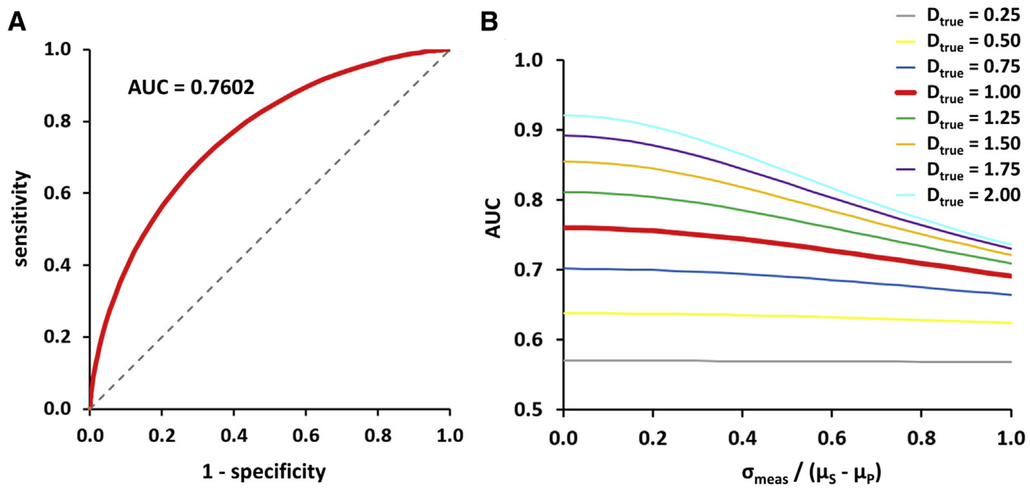

Fig. 1. Part (A) shows the ROC curve for MCI-to-ADD prediction by the true, error-free HV assumed to follow Gaussian distributions according to (4). Part (B) illustrates the deterioration of predictive performance (as measured by the decline of the area AUC under the ROC curve) by additional intersubject variability due to measurement errors σ_{meas} of HV estimates. The red curve shows the impact of the additional variability on the performance of the true HV according to part (A). The other curves show the impact of measurement error when the true, error-free HV would perform better (actual biological group difference $D_{true} > 1$) or worse ($D_{true} < 1$) than assumed in part (A). Abbreviations: HV, hippocampus volume; ADD, AD dementia; AUC, area under the ROC curve; MCI, mild cognitive imapairment.

To assess the impact of additional variance σ^2_{meas} on the prognostic performance also for other biological scenarios, σ_{true} was varied whereas μ_S and μ_P were kept constant at the values in (4). Biological scenarios were labeled by the actual biological group difference scaled to the actual (patho)physiological variance $D_{true} = (\mu_S - \mu_P)/\sigma_{true}$ (Fig. 1B). It is evident from Fig. 1B that the impact of the additional variance caused by the measurement process decreases with decreasing biological group difference. The lower the initial AUC, the flatter the decline of the AUC with increasing measurement error. This suggests that the stability of the predictive power of MRI-based HV in MCI with respect to the measurement method is a floor effect: even with the best measurement method, the power of MRI-based HV to predict MCI-to-ADD progression is inherently limited by the predictive properties of hippocampal atrophy. Thus, further decreasing HV measurement error compared with existing methods will have only very little impact on the predictive accuracy of hippocampus volumetry. To make HV widely available for routine clinical use, the measurement method should combine ease of use and short computation time with acceptable accuracy and precision. Efforts to harmonize HV measurement in the context of AD might account for this [7]. Furthermore, integrating existing HV volumetry methods in multivariable models rather than increasing HV measurement accuracy will be most efficient to make the best use of MRI-based HV as prognostic marker in MCI. Finally, hippocampal atrophy is not homogeneous across hippocampal subfields, suggesting that MRI-based volume measures of specific hippocampal subfields might provide better predictive power compared with the entire hippocampus [8].

Ralph Buchert*
Department of Diagnostic and Interventional Radiology and Nuclear Medicine
University Hospital Hamburg-Eppendorf
Hamburg, Germany

Catharina Lange
Department of Nuclear Medicine
Charité - Universitätsmedizin Berlin
Berlin, Germany

Per Suppa
jung diagnostics GmbH
Hamburg, Germany

Ivayla Apostolova
Department of Diagnostic and Interventional Radiology and Nuclear Medicine
University Hospital Hamburg-Eppendorf
Hamburg, Germany

Lothar Spies
jung diagnostics GmbH
Hamburg, Germany

Stefan Teipel
German Center for Neurodegenerative Diseases (DZNE)
Rostock, Germany

Bruno Dubois
Harald Hampel
AXA Research Fund and UPMC Chair
Sorbonne Universités

Université Pierre et Marie Curie (UPMC)
Inserm, CNRS
Institut du cerveau et de la moelle (ICM)
Département de Neurologie
Institut de la Mémoire et de la Maladie d'Alzheimer (IM2A)
Hôpital Pitié-Salpêtrière
Paris, France

Michel J. Grothe
German Center for Neurodegenerative Diseases (DZNE)
Rostock, Germany

References

[1] Bosco P, Redolfi A, Bocchetta M, Ferrari C, Mega A, Galluzzi S, et al. The impact of automated hippocampal volumetry on diagnostic confidence in patients with suspected Alzheimer's disease: a European Alzheimer's Disease Consortium study. Alzheimers Dement 2017; 13:1013–23.

[2] Ten Kate M, Barkhof F, Boccardi M, Visser PJ, Jack CR Jr, Lovblad KO, et al. Geneva Task Force for the Roadmap of Alzheimer's Biomarkers. Clinical validity of medial temporal atrophy as a biomarker for Alzheimer's disease in the context of a structured 5-phase development framework. Neurobiol Aging 2017;52:167–182.e1.

[3] Scheltens P, Leys D, Barkhof F, Huglo D, Weinstein HC, Vermersch P, et al. Atrophy of medial temporal lobes on MRI in "probable" Alzheimer's disease and normal ageing: diagnostic value and neuropsychological correlates. J Neurol Neurosurg Psychiatry 1992;55:967–72.

[4] Clerx L, van Rossum IA, Burns L, Knol DL, Scheltens P, Verhey F, et al. Measurements of medial temporal lobe atrophy for prediction of Alzheimer's disease in subjects with mild cognitive impairment. Neurobiol Aging 2013;34:2003–13.

[5] Hill DLG, Schwarz AJ, Isaac M, Pani L, Vamvakas S, Hemmings R, et al. Coalition Against Major Diseases/European Medicines Agency biomarker qualification of hippocampal volume for enrichment of clinical trials in predementia stages of Alzheimer's disease. Alzheimers Dement 2014;10:421–9.

[6] Faraggi D, Reiser B. Estimation of the area under the ROC curve. Stat Med 2002;21:3093–106.

[7] Wolf D, Bocchetta M, Preboske GM, Boccardi M, Grothe MJ, for the Alzheimer's Disease Neuroimaging Initiative. Reference standard space hippocampus labels according to the European Alzheimer's Disease Consortium-Alzheimer's Disease Neuroimaging Initiative harmonized protocol: utility in automated volumetry. Alzheimers Dement 2017;13:893–902.

[8] Carlesimo GA, Piras F, Orfei MD, Iorio M, Caltagirone C, Spalletta G. Atrophy of presubiculum and subiculum is the earliest hippocampal anatomical marker of Alzheimer's disease. Alzheimers Dement (Amst) 2015;1:24–32.

https://doi.org/10.1016/j.jalz.2018.03.006