SUPPLEMENTARY MATERIALS

RQS Scoring System

The radiomics quality score (RQS) score consisted of 16 components. The reviewers performed RQS evaluation according to six domains as previously reported (1, 2). Prior to the evaluation, a research meeting was held to educate the reviewers on the RQS system.

Two reviewers (with 6 and 9 years of experience in radiology, respectively) independently scored the articles for each of the six domains using RQS (Supplementary Material). If disagreement occurred between the two reviewers, a final decision was made through a consensus. Domain 1 represented image protocol quality and stability of image and segmentation (0 to 5 points). It consisted of image protocol quality (+2: well-documented protocol [+1], usage of public image protocol [+1]), multiple segmentations (+1), phantom study on all scanners (+1), and imaging at multiple time points (+1). Domain 2 represented feature selection and validation (-8 to 8 points). It consisted of feature reduction or adjustment for multiple testing (-3 or +3 points) and validation (-5 to +5 points). Domain 3 represented biologic/clinical validation and utility (0 to 6 points). It consisted of multivariable analysis with non-radiomics feature (+1), detection or discussion of biologic correlates (+1), comparison to 'gold standard' (+2), and report of potential clinical utility (+2). Domain 4 represented model performance index (0 to 5 points). It consisted of cut-off analyses (+1), discrimination statistics (+2: report of discrimination statics [+1] application of resampling method [+1]) and calibration statistics (+2: calibration statics [+1], application of resampling method [+1]). Domain 5 represented high level of evidence (0 to 8 points). It consisted of prospective validation (+7) or cost-effectiveness analysis (+1). Domain 6 represented open science and data (0 to 4 points). It consisted of open source scans, segmentations, code, and data (+4).

Consensus Reached for RQS Scoring

The additional topics were discussed by the two reviewers and a consensus was reached for evaluation. RQS was scored according to the consensus for the following topics:

1) Image protocol quality (domain 1): As Alzheimer's Disease Neuroimaging Initiative (ADNI) is a widely used multi-center open source database (3), several radiomics studies which used the ADNI database did not document the image protocol in the manuscript. However, an additional point for well-documented protocol was still earned in case the ADNI database was used with a reference for the database depository (adni.loni.usc.edu or www.adni-info.org) as the ADNI imaging protocol is widely known.

2) Multiple segmentation (domain 1): In quantitative neurodegenerative studies, brain region segmentation is usually performed automatically rather than manually or semi-automatically by single or multiple readers. Thus, if well-known and validated automatic segmentation software for brain imaging such as FreeSurfer, FMRIB Software Library, Statistical Parametric Mapping, or volBrain were used to perform segmentation of brain regions (4-6), an additional point was earned as it was considered to have better segmentation reproducibility.

3) Validation (domain 2): If cross-validation or nested cross-validation was performed only within the training set, it was considered to be missing validation and scored -5 points, as previously described (1). When the study was validated by randomly splitting a multicenter open source data such as the ADNI database into training and test sets, a scoring for a validation based on a dataset from the another institute (+4) was applied rather than applying the full scoring for a validation based on three or more datasets from distinct institutes (+5). This scoring was applied because although MRIs from ADNI were obtained across different sites and different MRI vendors, the ADNI study uses a strictly controlled MRI protocol, resulting in less heterogeneity of radiomics features that is acquired from a true external validation (7). Also, because the data was randomly split without knowledge of the acquired primary site, there is a possibility that data from the identical site may have been included in both the training and test sets.

4) Comparison with the gold standard (domain 3): A notable difference between oncology and AD studies is that the histologic confirmation of the disease diagnosis is feasible ex vivo for oncology, but not for AD (8). Thus, for diagnostic studies the pathologic gold standard according to the amyloid, tau, and neurodegeneration (ATN) guideline for AD diagnosis recommended by the National Institute on Aging and Alzheimer's Association Research Framework was applied

(9); cerebrospinal fluid amyloid beta (Aβ) or amyloid PET are specific to aggregated Aβ, and tau PET reflects aggregated tau, whereas hippocampal atrophy reflects neurodegeneration. For prognosis studies, well-known predictors of cognitive conversion such as apolipoprotein E allele of the apolipoprotein gene, clinical severity, brain atrophy, cerebrospinal fluid biomarkers were applied (10-14).

5) Potential clinical utility (domain 3): The statement of the Food and Drug Administration and National Institutes of Health Working Group (15) suggested 'clinical utility' is thought to be accomplished when a biomarker guide to net improvement of health outcomes or give information valuable for prevention, diagnosis, management or treatment of a disease (15-17). If only the potential utility was discussed without proper analysis, no additional point was scored.

REFERENCES

1. Park JE, Kim HS, Kim D, Park SY, Kim JY, Cho SJ, et al. A systematic review reporting quality of radiomics research in neuro-oncology: toward clinical utility and quality improvement using high-dimensional imaging features. *BMC Cancer* 2020;20:29
2. Park JE, Kim D, Kim HS, Park SY, Kim JY, Cho SJ, et al. Quality of science and reporting of radiomics in oncologic studies: room for improvement according to radiomics quality score and TRIPOD statement. *Eur Radiol* 2020;30:523-536
3. Jones-Davis DM, Buckholtz N. The impact of the Alzheimer's disease neuroimaging initiative 2: what role do public-private partnerships have in pushing the boundaries of clinical and basic science research on Alzheimer's disease? *Alzheimers Dement* 2015;11:860-864
4. Fischl B. Freesurfer. *Neuroimage* 2012;62:774-781
5. Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM. FSL. *Neuroimage* 2012;62:782-790
6. Manjón JV, Coupé P. volBrain: an online MRI brain volumetry system. *Front Neuroinform* 2016;10:30
7. Wyman BT, Harvey DJ, Crawford K, Bernstein MA, Carmichael O, Cole PE, et al. Standardization of analysis sets for reporting results from ADNI MRI data. *Alzheimers Dement* 2013;9:332-337
8. Boccardi M, Gallo V, Yasui Y, Vineis P, Padovani A, Mosimann U, et al. The biomarker-based diagnosis of Alzheimer's disease. 2-lessons from oncology. *Neurobiol Aging* 2017;52:141-152
9. Jack CR Jr, Bennett DA, Blennow K, Carrillo MC, Dunn B, Haeberlein SB, et al. NIA-AA research framework: toward a biological definition of Alzheimer's disease. *Alzheimers Dement* 2018;14:535-562
10. Andreasen N, Minthon L, Vanmechelen E, Vanderstichele H, Davidsson P, Winblad B, et al. Cerebrospinal fluid tau and Aβ42 as predictors of development of Alzheimer's disease in patients with mild cognitive impairment. *Neurosci Lett* 1999;273:5-8
11. Hsiung GY, Sadovnick AD, Feldman H. Apolipoprotein E ε4 genotype as a risk factor for cognitive decline and dementia: data from the Canadian Study of Health and Aging. *CMAJ* 2004;171:863-867
12. Karas G, Sluimer J, Goekoop R, Van Der Flier W, Rombouts S, Vrenken H, et al. Amnestic mild cognitive impairment: structural MR imaging findings predictive of conversion to Alzheimer disease. *AJNR Am J Neuroradiol* 2008;29:944-949
13. Landau SM, Harvey D, Madison CM, Reiman EM, Foster NL, Aisen PS, et al. Comparing predictors of conversion and decline in mild cognitive impairment. *Neurology* 2010;75:230-238
14. Riemenschneider M, Lautenschlager N, Wagenpfeil S, Diehl J, Drzezga A, Kurz A. Cerebrospinal fluid tau and β-amyloid 42 proteins identify Alzheimer disease in subjects with mild cognitive impairment. *Arch Neurol* 2002;59:1729-1734
15. O'Connor JP, Aboagye EO, Adams JE, Aerts HJ, Barrington SF, Beer AJ, et al. Imaging biomarker roadmap for cancer studies. *Nat Rev Clin Oncol* 2017;14:169-186
16. Hayes DF, Allen J, Compton C, Gustavsen G, Leonard DG, McCormack R, et al. Breaking a vicious cycle. *Sci Transl Med* 2013;5:196cm6
17. McShane LM, Hayes DF. Publication of tumor marker research results: the necessity for complete and transparent reporting. *J Clin Oncol* 2012;30:4223-4232