AMYPAD – Centiloid measure of Amyloid PET to quantify brain amyloid burden

Summary of EMA SAWP discussion meeting

This document summarizes the new information and additional analyses provided by the AMYPAD consortium to the SAWP at the discussion meeting on June 7, 2023 and in the meeting minutes. The proposed refinements of the Context of Use are also included.

1) Clarifications on the revised Context of Use (CoU)

The general aim for the BQO procedure request was to demonstrate the robustness of the Centiloid methodology for the clinical use. The assessment of the prognostic risk for future cognitive decline was no longer in the scope of the qualification opinion.

A summary of the proposed context of uses for the Centiloid metric, and how they apply to different tracer and use of analysis pipeline is shown below:

CoU	Item /threshold	Can be used with Different tracers	Different pipelines could be applied
Cross-sectional (rule out amyloid)	Cut-off <10 CL	Yes	Yes
Cross-sectional (confirms pathological amyloid levels)	Cut-off >30 CL	Yes	Yes
Adjunct to visual inspection	Used as guidance for visual read. If CL is obtained in grayzone (>10CL and <30CL), it need to be checked if there is only focal uptake, atrophy or other scan issue (QC)	Yes	Yes
Longitudinal evaluation of amyloid pathology	Baseline CL	No (i.e., use the same tracer if repeat scans are performed)	No (i.e., use the same pipeline for analysis of repeat scans)
Evolving amyloid levels	If greater than >3 CL/annum observed	No (see above)	No (see above)

Various analyses provided meanwhile sufficient evidence that a amyloid PET quantification of larger than 30 CL shows presence of pathological amyloid levels, i.e. it allows for reliable inclusion of amyloid-ß levels. A PET scan quantification result below 10 CL allows reliable exclusion of pathological amyloid-ß levels.

The results obtained in AMYPAD multi-centre studies can be generalized. There was not only “one” center who did the analysis. Comparable results were obtained with the independent IDEAS dataset (real world data) using different pipelines.

Moreover, CL-based cut-off points are comparable when derived from PET-only pipelines and MR-based PET pipelines. In addition, CSF-derived cut-off points provide comparable results to CL-based cut-off points, when both PET and CSF data are available from patients. More detailed analyses can be provided as amendment to these meeting minutes.

2) Additional analyses to provide estimates of variability/uncertainty within a pipeline and between pipelines on the basis of data from work package A1

The between pipeline variability was evaluated using two different pipeline designs that represent two extremes of pipeline design with the assessments being performed on both visual read (VR) negative and positive images.

- For the subjects with negative VR, a mean difference below 1.8 CL was observed for all three tracers.
- For VR positive PET scans, higher levels of difference were observed where FBP had the minimum difference (mean±SE=0.96±0.66 CL) and FMM and FBB showed somewhat higher levels of differences (-5.42±0.82 CL and -8.77±1.03 CL respectively). Please note that the average CL value in these VR positive individuals is over 70 CL, therefore rendering these differences inconsequential.

The degree of flexibility between tracers depends on the context of use. Between pipeline variability is deemed acceptable to determine CL values for cross sectional use (as long as adhering to the recommended use, see summary below). However, for longitudinal studies, AMYPAD recommends consistency of pipeline and tracer (see summary table above)

The AMYPAD technical recommendations for utilizing the CL method comprise:

- Use consistent acquisition and reconstruction settings over time for comparative use
- Consider harmonizing (e.g., using phantoms) the data when comparing CL data obtained across different scanners
- Use software packages that are validated using the QC criteria recommended by Klunk et al. 2015 and approved for the intended use according your local regulation (e.g., CE-marked)
- CL quantification should be quality controlled to assess factors that may bias CL values (e.g., wrong positioning of the ROIs, atrophy, etc.)
- Use consistent reference region, preferably whole cerebellum
- Use the right CL threshold for the right setting

3) Recommendations for standardization of the "pipelines" to generate Centiloid Units based on the AMYPAD consortium work

Different scanners and reconstruction parameters have a minimal impact on the CL quantification when using the whole cerebellum and whole cerebellum+ brainstem as reference region. Brain anatomy mapping is needed only in the case of defining subject-based cortical targets and reference regions. Brain atrophy created a difference of difference of ~ 4CL, on average, in the subjects who are visual read positive, diagnosed with dementia (average CL ~90) and show the highest level of brain atrophy in the DPMS sample. Such a difference should not impact assessment of the image as 'positive' for the presence of composite amyloid burden in the brain.

Thus, no impact is expected from on cross-sectional studies but there may be an impact in longitudinal analyses, which is reflected in the recommendations in the summary table above.

4) Derivation of thresholds for the CL scale for determining tracer positivity

CL cut-offs have been derived against neuropathology as the standard of truth (SoT). A CL = 12.2 was found to be optimal to discard amyloid pathology (CERAD none vs. any; Youden criterion) in 4 cohorts (UCSF/UCD, UPitt., Mayo, AIBL). A CL = of 32.4 was found to be optimal to detect established amyloid pathology (CERAD any vs. frequent) in the same study. Similar cut-offs have been reported in independent studies using a multiplicity of statistical criteria and methodological approaches (e.g. Salvadó 2019; Amadoru 2020; Bullich 2021). It can be concluded that CL values < 10 CL are validated to rule out the presence of amyloid pathology and CL > 30 to predict established amyloid pathology.

Furthermore, CL thresholds are supported by visual read, specifically threshold differences in experienced and unexperienced readers. In the large IDEAS cohort across all radiotracers, agreement between majority expert read and local readers of a random sample was very good with above 85% agreement for positive scans and 91% agreement for negative scans, which indicates that non-expert readers are very competent and the cut-offs are very similar among all kinds of readers.

Centiloid quantification was performed in an independent clinical cohort (IDEAS) and presented by the IDEAS study team (Zeltzer et al 2022). Centiloid analyses showed a bimodal distribution with a minority surrounding the 24.4 CL threshold. There was an 86.5% agreement between the local readers and quantification-based positivity. 53.3% of the visual reads were positive by both quantification and visual inspection whilst 33.2% were both negative. From a discordance perspective there were approximately equal amounts of visual negative/quantification positive (6%) and 7.5% being visual positive/quantification negative. The average CL value of the visually negative scans was approaching the zero CL value as expected (3 CL ± 27) whilst the CL value of the positive scans was (72 CL ± 41).

An additional analysis was performed to derive CL cut-offs for the 95% specificity classification. Scans from the AMYPAD-DPMS cohort (n=729) were compared with results obtained from scans of the independent IDEAS cohort (n=983).

The 95% confidence intervals (95%CI) of the Mu1 and Sigma1 parameters obtained from the GMM can be compared between both studies. Other parameters such as Mu2, Sigma2 and Proportion depend

on the recruitment criteria of each study and are thus not directly comparable. The following results were obtained from the AMYPAD-DPMS:

- Mu1 (95%CI): 0.46 (-0.91, 1.88); Sigma1 (95%CI): 13.05 (11.52, 14.44)

Similar results were obtained with GMM from the IDEAS scans:

- Mu1 (95%CI): -1.56 (-3.31, 1.11); Sigma1 (95%CI): 12.90 (10.23, 17.96)

For the 95% specificity classification the following CL cut-off values were obtained:

AMYPAD-DPMS (CL (95%CL): 21.69 (18.49, 24.65)

IDEAS (CL (95%CL): 24.42 (20.15, 27.25

In conclusion, the Centiloid metric was consistent across two large real-world studies to provide a harmonized metric across 3 tracers and to compare to local visual image interpretation. These results were submitted and accepted for presentation at the CTAD 2023 conference. Overall, these results demonstrate the consistency of the CL metrics across tracers and against image interpretation to a high degree.

5) Validity of Centiloid procedure proposed by Klunk et al 2015

The original Centiloid procedure, based upon calibration to PIB, use of GAAIN cortical mask, and the use of whole cerebellum as preferred reference region, is widely employed in clinical research studies across the world without modification and is still considered valid.

AMYPAD advises to:

1) Record time of PET acquisition to account for kinetic properties of tracers,
2) Use standard pipeline and whole cerebellum for robust CL values in multi-center settings
3) Perform visual QC of PET images and visual QC of ROI placement prior to automated quantification.

AMYPAD studied the bias introduced by subject selection on level 2 calibration and the results indicated minimal bias (>2 CL) across tracers. Simulations assessed the maximum potential of systemic bias and showed that there is a likelihood of only minor systemic errors happening.

An analysis was performed to estimate the impact of error propagation in the initial generation of the CL conversion equations for FBB and FMM using a simulation-based approach. The errors (95% CI) in CL estimates with respect to the theoretical values were associated to the sample size of the dataset used to develop the CL calibration equation: ± 3 CL (CL=0), ± 7 CL (CL=100), for N=35; ± 1 CL (CL=0), ± 5 CL (CL=100), for N=74.

The maximum difference (95% CI) between FBB and FMM CL values due to error propagation in the CL equations was small for low CL values: ± 3.5 CL (CL=0 and 25 CL), ± 4.5 CL (CL=40), and it grew at larger CL values: ± 10.5 CL (CL=100)

The simulations have been repeated with florbetapir using the open access GAAIN dataset. The simulations of potential systematic bias due to the propagation of errors for florbetapir showed similar results as with FBB and FMM.

This potential systematic bias will not affect (i) the classification of amyloid-beta status, (ii) the estimates of amyloid-load changes over time when using a single radioligand (e.g. in a treatment monitoring setting) and (iii) the estimates of amyloid-beta load in cross-sectional clinical trials using one single ligand.

In conclusion, the propagation of errors in the generation of FBB and FMM CL calibration equations is expected to have a small influence (<3.5 CL) in the lower CL range and is likely to be below ±5 CL (80% confidence interval) in the region of transition between amyloid-beta negative and positive subjects (0<CL<50). It increases to approximately 10% at the higher end of the Centiloid Scale. The CL approach by Klunk et al. 2015 is considered appropriate to derive CL calibration equations. The users of these equations should be aware of their limitations and potential bias to avoid overinterpreting subtle CL differences across tracers. The use of large sample sizes in the development of the CL calibration equations would reduce the likelihood of errors due to propagation of uncertainties.

6) Additional analyses to demonstrate generalizability of Work Package A4

Amyloid PET images were quantified in Centiloids through different process pipelines used in the AMYPAD studies and the outcomes were compared. The mean CL for subjects processed through the BBRC pipeline was 13.2±24.9 CL (range -11.8 to 114.4). The AmyPype Centiloid output gave a mean of 11.9±26.7 CL (range -17.7 to 115.2). Analysis of the data gave a mean absolute deviation of 5.4±4.0. Correlation of the data was strong, with an R^2 of 0.94.

In conclusion, the BBRC pipeline employs the GAAIN whole cerebellum reference region, with the GAAIN Cortical mask VOI. PET and MRI images are co-registered and processed in MNI space. AmyPype employs the same GAAIN Whole cerebellum reference region and GAAIN Cortical mask VOI however the pipeline is PET-Only, with the images co-registered to the MNI152 T1-weighted average structural MRI template. From the data a strong correlation (R^2 0.94) was observed between the CL values generated from the BBRC and AmyPype Centiloid process pipelines. The low mean absolute deviation (5.4±4.0) is consistent with the noise levels found for Centiloid longitudinal/test-retest conclusions. Therefore, from this analysis it can be concluded that data generated from either pipeline is comparable and both pipelines are robust in their utility and output of Centiloid values for amyloid PET images.

The following analysis was performed as an alternative to direct head-to-head comparison between tracers. The difference in the estimated marginal means with FMM and FBB was of 1.09 CL (95% Confidence Interval: [0.84, 1.42]), after accounting for the amyloid load with the CSF p-Tau/Abeta42 ratio. All tests for heteroscedasticity were negative (p>0.05). Differences in CL values associated with the tracer are within the tolerance limits of the Centiloid transform (-2, 2) and within the range of the

test-retest variability (2.5 CL) of amyloid PET imaging. This result confirms that the CL transformation accurately render comparable quantitative estimates of amyloid load irrespective of the PET tracer used. These results were submitted and accepted for presentation at the CTAD 2023 conference.

7) Additional analyses on pipeline variability and influence of other parameters

To address this question, the between pipeline variability was evaluated using two different pipelines that represent two extremes of pipeline design with the assessments being performed on both visual read (VR) negative and positive images:

- GAAIN standard pipeline: Using GAAIN cortical VOI and GAAIN predefined RRs, and MNI space
- Subject-based pipeline: Delineating both cortical target and RRs in native space

The between pipeline variability was assessed on VR-negative and VR-positive scans. The between pipeline differences are below 2 CL for all tracers for VR-negative scans. The VR-positive scans showed between pipeline differences of 0.966 (95%CI: -0.33 to 2.25) for FBP. Higher differences were observed for FMM (-5.42 CL) and FBB (-8.77). As the mean CL in the VR-positive group is about 77 CL, this difference has no impact on categorizing the scans.

To assess the impact of differences in the scanner and reconstruction method, the Harmonization status was introduced to the model. Using WCB and WCB+BST as reference region resulted in a mean difference below 2.5 CL between harmonized and unharmonized images for both negative and positive VR group, which is in the range of test-retest variabilities.

Atrophy could affect the determination of CL values in subjects with dementia with a mean CL level of ~85. A mean difference of 3.99 ± 8.21 CL was observed while using the GAAIN cortical mask versus the subject-based cortical mask. This difference has no effective impact on considering scans as positive.

In conclusion, these additional analyses demonstrate in a variety of ways the robustness of the Centiloid metric as a tracer-independent measure of amyloid burden in the brain. AMYPAD has shown that pipeline design options do not markedly influence the CL measure nor did the propagation of error in development of the initial CL calibration equations (particularly in the low CL range). Additionally, when using CSF p-Tau/Abeta42 as a central anchor there was minimum differences in the CL levels between tracers too.