6 March 2024
EMADOC-1700519818-1200791
Committee for Medicinal Products for Human Use (CHMP)

DRAFT Qualification opinion for Centiloid measure of Amyloid PET to quantify brain amyloid deposition

Draft agreed by Scientific Advice Working Party (SAWP)	26 October 2023
Adopted by CHMP for release for consultation	24 January 2024[1]
Start of public consultation	7 March 2024[2]
End of consultation (deadline for comments)	18 April 2024

Comments should be provided using this template. The completed comments form should be sent to ScientificAdvice@ema.europa.eu

Keywords	Qualification of Novel Methodology, Biomarkers, Brain Amyloid Burden

[1] Last day of relevant committee meeting

[2] Date of publication on the EMA public website

Official address Domenico Scarlattilaan 6 ● 1083 HS Amsterdam ● The Netherlands
Address for visits and deliveries Refer to www.ema.europa.eu/how-to-find-us
Send us a question Go to www.ema.europa.eu/contact **Telephone** +31 (0)88 781 6000 An agency of the European Union

1. CHMP Qualification Opinion

Based on the evidence and data presented in the provided qualification opinion request and additional material provided for a discussion meeting with the applicant, CHMP considers that the Centiloid Unit for the measurement of brain amyloid level can be considered a validated measure of global amyloid load in the brain for enrichment in clinical trials, if properly used with quality control procedures. The advantage would be potential use of different PET tracers and scanning and analysis procedures (scanning pipelines) in cross sectional settings, as non-normalised raw data for different available tracers are not comparable. However, estimated variability of the measure needs to be considered for application and the technical recommendations from the AMYPAD consortium should be followed.

Context of Use (CoU)

The proposed method Centiloid Unit is intended for the measurement of brain amyloid burden in subjects with early or established Alzheimer's disease pathology to be used in clinical trials. The Centiloid Unit can be used as adjunct to visual reads (negative with white matter retention only, positive with cortical tracer retention) with different tracers and PET scanning and analysis procedures. Based on this, the Centiloid Unit can be used for enrichment in clinical trials considering existing qualification opinions on the use of amyloid PET-imaging as biomarkers for enrichment in regulatory clinical trials in mild to moderate Alzheimer's disease (EMA/CHMP/SAWP/893622/2011) and in pre-dementia AD (EMA/CHMP/SAWP/892998/2011). Quality control procedures including visual inspection should be in place to detect issues with PET scans before use in clinical trials (e.g., focal uptake, atrophy). If Centiloid Units are used in cross-sectional studies or longitudinal settings, variability between PET scanning and analysis procedures (between-pipeline variability) is low in subjects with negative visual reads but is noticeably higher in subjects with positive visual reads and rises with increasing amyloid load. Use of the Centiloid Unit as prognostic or predictive measure is currently not in scope of the Context of Use.

Technical recommendations

A standard procedure defined by the work of Klunk et al. (Klunk WE et al., Alzheimer's & Dementia 2015) is still valid as basis for technical implementation and calculation of Centiloid Units. Centres that want to establish the Centiloid Unit should follow the steps outlined in the publication. Combined PET scanning and analysis procedures are termed pipeline. Concordance and agreement between the centre-specific in-house pipeline and the GAAIN reference data pipeline quantifications should be checked, and it is necessary to meet the acceptance criteria mentioned by Klunk et al. As the calibration method may lead to some bias at the upper end of the CL scale due to use of a limited number of subjects, the maximum potential bias that was assessed with simulation by the applicant should be considered for the application. Using the same PET tracer and pipeline would be preferable when variability is intended to be minimised.

The applicant provides a set of technical recommendations for utilizing the CL method:

- Use consistent acquisition and reconstruction settings over time for comparative use.
- Longitudinal evaluation of amyloid levels should be performed with the same tracer and scanning pipeline
- 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 to your local regulation (e.g., CE-marked).
- CL quantification should be quality controlled (i.e. via visual inspection of each image) to assess factors that may bias CL values (e.g., wrong positioning of the ROIs, atrophy, etc.).
- Use consistent reference region, preferably whole cerebellum.
- Corrections for partial volume effects are not normally applied.

Additional recommendations pertain to issues to be avoided:

- Do not apply published CL calibration equations developed for one analytic pipeline to another pipeline.
- If comparing data collected from other sites, do not use different pipelines (check for consistency of pipeline used for CL measure).

Check of correct brain anatomy mapping should be part of the quality control process. While use of a pipeline involving MRI scanning would not be mandatory from a variability perspective and pipelines using only PET may be appropriate, it should be checked if brain atrophy or cerebrovascular disease has an impact on quantitative reads (and visual reads). Impact of ventricular expansion and different volumes of interest is expected to be limited. As for all static PET acquisition methods, SUVr and CL values are marginally impacted by cerebral blood flow, as opposed to dynamic methods as e.g., non-displaceable Binding Potential (BP_{ND}).

Robustness of the Centiloid Unit

The applicant provided a re-analysis of the data used in work package A1 for the discussion meeting to quantify the sensitivity to design of the scanning and evaluation pipeline. This work used 32 different combinations of 'pipelines' with the 3 currently authorised PET tracers. The results are included as part of the briefing documentation for the discussion meeting and are documented there. The between pipeline variability was evaluated using two different pipelines that represent two extremes of pipeline design. Results suggest a between-pipeline difference of 2 CL for subjects with negative visual read. For subjects with positive visual reads, differences are larger for flutemetamol (5.42 CL) and flobetaben (8.77 CL). Reconstruction method had an impact in the range of 2.5 CL and atrophy could impact measurements in subjects with dementia (CL values in the range of 85) with an observed mean difference of 3.99 CL.

For the discussion meeting the applicant also presented an updated analysis of simulations presented in the work package A2 to include florbetapir using the open access GAAIN dataset (https://www.gaain.org/centiloid-project). The simulations of potential systematic bias due to the propagation of errors for florbetapir showed similar results as florbetaben and flutemetamol. Maximum potential bias could be larger (~10 %) in subjects with high amyloid load (> 75 CL). Results also show that biases in the range of 0<CL<50 are ±5 CL (80% confidence interval) in the region of transition between visual read negative and visual read positive subjects.

Quantitative values for Centiloid Units (CL)

The SUVR, the Centiloid scale and the reference-based z-scores are the most commonly used quantitative measurements of amyloid burden on amyloid PET scan. The Z-score represent standard deviations from the mean of the control group and z=2 is the threshold for an abnormal result of the PET scan. Both z-scores and CL are based on SUVR and thus inherit some of its benefits and drawbacks. Z-scores and CL-values were found to be equally useful in the AMPYPAD DPMS study, but

the CL scale might be considered as a more illustrative scale and easier to comprehend. Several studies support the use of quantitative CL data as adjunct to visual reads. These used a retrospective design and show high concordance between quantitative assessments and visual reads. Data suggest value of quantitative assessments as adjunct to visual read in settings with borderline cases and less experienced readers. The AMYPAD consortium has performed the first prospective read study to investigate the value of adjunct reads for challenging cases for visual reads. This study used assessment before and after disclosure of quantitative assessments and supports the use of quantitative assessments. A large number of studies are available that aim to define CL thresholds for different purposes. These should be considered when quantitative values are intended to be used, e.g., for population selection or enrichment. When using quantitative Centiloid (CL) values, a cut-off of <10 CL may be used to rule out amyloid load. The over limit >30 CL may be used to indicate pathologic amyloid levels. Values between 10 and 30 CL (grey zone) need to be interpreted with caution. Larger variability and potential bias at the upper end of the scale should be considered. Longitudinal evaluation of amyloid levels should only be performed with the same tracer and scanning pipeline. The lower threshold to rule out amyloid load can be considered established, while for an upper threshold it should be considered if sensitivity or specificity for detecting a defined amyloid load level would be of importance. Use of quantitative CL data may be considered more sensitive than using visual reads to detect changes in amyloid load and could help detect accumulation of amyloid. Regional amyloid load quantification in specific brain regions is currently not proposed for any Context of Use relating to Centiloids as the proposed methodology uses a cortical composite mask as volume of interest.

Relevance of longitudinal changes

A clinically relevant rate of accumulation has not been established yet. An assessment of amyloid accumulation is not included in the Context of Use.

AMYPAD work packages

For a detailed discussion of the CHMP assessment of the AMYPAD work packages provided with the background document, please refer to the answers to the questions 1 to 4 below.

See also sections of the briefing package.

2. Applicant Executive Summary

Biomarker Qualification Opinion (BQO) for the Centiloid Measure as a universal metric for the assessment of brain amyloid burden:

A robust standardised tracer-independent methodology for measuring global amyloid load in subjects with early or established pathology.

Applicant: IMI funded Amyloid Imaging to Prevent Alzheimer's Disease (AMYPAD) consortium

Biomarker Qualification Opinion (BQO) Process:

Developed by the European Medicines Agency (EMA) to facilitate the acceptability of specific use for a novel method or imaging modality to enable progress in the development of novel treatment and management regimes. The opinion process involves the assessment of submitted data and additionally a further public consultation with the scientific community. The process starts with the applicant submitting a detailed proposal and is projected to take approximately 9 months to 1 year.

AMYPAD Program:

AMYPAD is a public-private partnership of 15 European partners who have two active clinical programs in the field of brain amyloid positron emission tomography (PET) imaging, with the ultimate goal to improve knowledge of dementia pathology and clinical progression. One major objective is the development and validation of robust standardised methodology for the measurement of amyloid in the brain. The project is now in its 6th year of funding and was granted a no-cost extension till the 30th of September 2022. The F-18 tracers Vizamyl ([^{18}F]flutemetamol) and Neuraceq ([^{18}F]florbetaben) are approved by EMA and broadly available in Europe, both are being studied in the AMYPAD program. Additional data from the US based IDEAS study which has a large proportion of Amyvid ([^{18}F]florbetapir) scans has also been included in the results section.

What currently exists:

Fluorine-18 labelled brain amyloid PET tracers have been available for routine use in Europe since 2013 and have been validated against Consortium to Establish a Registry for Alzheimer's disease (CERAD) pathology as the standard of truth. Clinical routine use of brain amyloid PET tracers involves categorisation of static scans by visual read as either negative or positive. All three amyloid PET tracers approved in the EU have quantification included in their SmPC as an adjunct to a visual read to assist in the assessment of an amyloid PET scan. Additionally in the research space, quantitative measures are being employed with many of the standard software packages able to calculate both regional and composite levels of amyloid burden, enabling a continuous measure of amyloid load in addition to the dichotomous read that the visual inspection allows. Methods such as the standardised uptake value ratio (SUVr) yield tracer uptake values, which vary depending upon the chosen reference region and the analytical implementation. In turn, non-displaceable Binding Potential (BP_{ND}) reflects specific tracer uptake, as it takes several technical and physiological factors into account and is therefore considered a more accurate and precise measure. This measure, however, faces a similar dependency on radiotracer and analytical approaches, and requires a longer "dynamic" acquisition protocol, which may limit routine clinical use.

What is the knowledge gap:

Recently, the field of Alzheimer's disease (AD) research has focused on the value of both the topographical distribution and burden of amyloid pathology present, rather than a binary classification of the amyloid status. Studies so far have illustrated the added value of this information for both disease-modifying therapies and clinical use. There is a need to reliably quantify the presence of early amyloid pathology as secondary prevention trials move to treat subjects with low but detectable levels of amyloid. Additionally, there is value to improve the prognostic value of amyloid imaging in clinical routine, by considering the overall pathological load, which could improve subject placement along the AD trajectory. Although controversial regarding the clinical benefit demonstrated so far, the recent Aduhelm approval by the Food and Drug Administration (FDA) also highlights the potential value of a universal metric to assess the amyloid burden by PET as the label was updated in April 2022 to include the following 'confirm the presence of amyloid beta pathology prior to initiating treatment'. This could include both a baseline measure of amyloid to initiate treatment and potentially further scans for the purposes of managing the therapy regime. In addition to Aduhelm, other promising anti-amyloid therapies are in the final stages of closing out Phase III studies and submitting NDA/MAAs in both the USA and Europe. Managing both the inclusion into therapy as well as therapy monitoring across both global territories and with multiple tracers will require a consistent and robust approach.

One method increasingly gaining traction in the dementia neuroimaging space is the Centiloid measure, a tracer independent metric that can be easily grasped beyond Nuclear Medicine as well as

providing thresholds to answer different questions. Thus, while visual binary read of global amyloid provides useful information for clinical routine and research purposes, it does not consider the wealth of information that brain PET scans provide, both from a regional and continuous quantitative measure perspective.

Premise of BQO:

To facilitate the wider utility of standardized, tracer independent, and sensitive methods for 1) measuring cross-sectional levels (and potentially longitudinal changes) of brain amyloid pathology across PET tracers and 2) support amyloid PET biomarker use in both clinical routine and research by providing information on the extent of pathology for differing scenarios. These could include the evaluation of both early and established amyloid pathology as well as the possibility to predict disease trajectory (i.e., prognosis). Currently, the Centiloid measure could be considered the most developed quantitative methodology within the field of amyloid PET and has been reliably implemented in multiple studies, including AMYPAD and clinical trials of anti-amyloid drugs. Other quantitative methods to optimally measure amyloid burden or accumulation have been proposed, such as Aβ load and Aβ index. However, these approaches are currently less mature, having only been assessed in limited data sets.

Sources of data:

The primary sources of data presented in this BQO is amyloid measures from the two AMYPAD studies (i.e., the Diagnostic and Patient Management study, DPMS; and the Prognostic Natural History Study, PNHS). Additionally, work has been performed by members of the AMYPAD consortium on other cohorts (e.g. ALFA+, ABIDE, IDEAS etc) and will be appropriately referenced. There has also been a large body of data published in the literature and/or presented at recent conferences and this too is considered in this application.

Analysis proposed in this BQO:

A wealth of data and analysis primarily from both the AMYPAD DPMS and PNHS studies are presented in this BQO dossier. The analysis described in Chapter 6 is broadly divided into three sections which cover a) analytical robustness of the quantitation of cortical amyloid, b) cross sectional results of image analysis in the clinical subgroups of DPMS and other studies and c) the longitudinal analysis of amyloid PET in both DPMS and PNHS.

Value to the field of AD:

To provide a framework for the validation of quantitative assessment of amyloid burden, which is suitable for use/implementation by the general dementia field. The Centiloid method is the example for this BQO. The approach by AMYPAD has been endorsed by the European Association of Nuclear Medicine (EANM) (see letter of endorsement in Appendix A).

The application could also provide a template for further methodologies to be introduced as well, as future uses of amyloid PET are expected, e.g., more widespread applicability of longitudinal scanning to monitor therapeutic efficacy of cases with developing pathology.

The use of the Centiloid method allows the dementia field to use a central, universal metric, which is valid across all three approved brain amyloid PET tracers. This method aligns the use of target and reference regions and harmonizes the outcome measures.

Additionally, the BQO will demonstrate best practice PET acquisition parameters for the acquisition and reconstruction of amyloid PET images gained via collection and analysis of over 2000 images acquired

in the AMYPAD program (either newly acquired for prospective AMYPAD studies or in collaboration with other consortia).

3. Applicant questions and CHMP answers

Based on the Coordinators' reports the CHMP gave the following answers:

Question 1

Does EMA agree it valuable to have a single universal metric that is tracer independent to measure amyloid burden in the brain?

CHMP answer

CHMP agrees that standardisation in measuring amyloid burden by PET would be valuable, e.g. in terms of regions of interest, reference region, tracers used, and cut-offs for guiding interpretation.

Based on the material provided with the briefing document and material provided for the discussion meeting, the use of the Centiloid (CL) method can be considered validated across the three approved brain amyloid PET tracers for the agreed Context of Use. The method aligns the use of target and reference regions and harmonizes the outcome measures. However, even if the method has undergone validation work using the currently approved tracers, it may not be sufficient to call it "tracer independent" regarding all possible future tracer developments, as the available local PET data quality and quantitative properties may differ.

Three fluorine-18 amyloid PET tracers ([^{18}F]florbetapir (Amyvid®), [^{18}F]flutematmol (Vizamyl®) and [^{18}F]florbetapen (Neuracq®) are currently approved for routine clinical use, i.e. for measuring plaque density of adult patients with cognitive impairment who are being evaluated for Alzheimer's disease (AD) and other causes of cognitive impairment. In addition, although formally not registered in this indication, [^{11}C]PiB is one of the most widely used tracers and has been used as a benchmark for imaging Aβ in vivo.

Positive qualification opinions to use amyloid PET-imaging (positive/negative) as biomarkers for enrichment in regulatory clinical trials in mild to moderate Alzheimer's disease (EMA/CHMP/SAWP/893622/2011) and in pre-dementia AD (EMA/CHMP/SAWP/892998/2011) are already available. Of note, amyloid related positive/negative PET has not been qualified as diagnostic tool or outcome or longitudinal measure. At that time scans were only interpreted dichotomously as "negative" (white matter retention only) or "positive" (cortical tracer retention) and it was clearly stated that "neither the actual value of PET (+) or (-) to accurately predict rate of progression have been reported", so quantification of amyloid was not available.

Since then, substantial progress in use of semi-quantitation methods, analysing tracer uptake mainly based on SUVR has been made. Semi-quantitative measurements, which were validated in tracer specific settings have been tested against post-mortem confirmation of brain amyloid status as a universal standard of truth (SOT) and are recommended for use as an adjunct to visual read in the SmPCs of all three amyloid detecting PET tracers approved in the EU. However, SUVrs are tracer-specific and may be impacted variety of factors. Therefore, a more universal and standardised method for semi/quantitative measurement of amyloid is welcome for a Context of Use for which sufficient and robust data is provided.

CHMP agrees that the use of amyloid PET offers a high negative predictive value vs. pathology seen in autopsy. For many therapeutic interventions which have been aimed at treating AD patients, a positive amyloid scan has been used as (one of the) selection criteria for treatment inclusion. The approved radiotracers have been validated to detailed extent with autopsy data as standard of truth, indicating that they give specific, direct, and in vivo information on cerebral amyloid load.

The centiloid (CL) methodology is one method of standardizing Aβ quantification, that is increasingly applied in clinical trials across the AD continuum (e.g. in trial on Aducanumab, Lecanemab, Donanemab). Most recently the centiloid-based quantification of amyloid load was applied in a study testing use of quantitative methods in the PET imaging compared to visual reads with Neuraceq. Diagnostic performance tested against post-mortem histology showed high level of sensitivity and specificity (>90%) when the centiloid method was applied. Quantitative measurement methods based on SUVRs, including those analysed with centiloid, were included in the Neuraceq SmPC (see EMEA/H/C/002553/II/0038). For the two other commercially available tracers' data on quantitative methods have been presented based on SUVr thresholds derived for florbetapir (Pontocorvo MJ et al., Eur J Nucl Med Mol Imaging 2017) and flutemetamol (Thurfjell L et al., J Nucl Med 2014), including comparisons against post-mortem histology. Adjunctive use of quantitative image information is included in SmPCs of all three available tracers with different recommendations and description of use. Data on Centiloids for quantitative have yet been only presented by the MAH for Neuraceq.

However, there are limitations related to PET use. Amyloid PET alone, even with clinical assessment, cannot exclude other (concomitant) causes of dementia or amnestic MCI (Mild Cognitive Impairment). A positive amyloid PET scan is not synonymous with AD since patients with for example Lewy body dementia might exhibit elevated beta-amyloid levels as well. The fact that non-symptomatic patients with positive amyloid PET scans may never develop amnestic symptoms is also problematic. In asymptomatic patients the benefit of early determination of PET positivity might even have negative effects with respect to unjustified patient burden e.g., over-management and decreased Quality of Life. Finally, as demonstrated in previously failed trials with antibodies against amyloid, a decreased amyloid burden according to PET is not synonymous with clinically relevant treatment effects.

In principle, it can be agreed that a single universal metric that can be used with different tracers to measure amyloid burden is valuable. However, Centiloid Units itself are based on SUVr measurements and the factors influencing SUVr measurements in general (Krishnadas N et al., Semin Nucl Med 2021, Adams MC et al., Am J Roentgenol 2010) need to be considered. The Applicant presented updated data for the discussion meeting, supporting interchangeability of methods of quantification and a discussion of the factors affecting SUVr measurements in a population that includes patients with various types of amyloid burden. Based on this, the context of use statement for the Centiloid Unit quantitative methodology was refined and narrowed

Reference is made to the following questions with detailed comments on the data sources used for this biomarker qualification (see the answers to questions 2 and3).

Question 2

Does EMA agree that the Centiloid metric is suitable for measuring amyloid burden in the brain?

CHMP answer

Centilod as an adjunct to visual read

The proportion of pre-dementia patients assessed in memory clinics has significantly increased over the past few years reaching up to ~25% of patients presenting with SCD (Subjective Cognitive Decline). In these subjects, amyloid deposition may be developing, or may be more focal compared to global, which could make visual assessment more challenging, especially by less experienced readers and the dichotomous approach may be more prone to subjectivity. Adjunct quantitative measures of amyloid measurement and potentially more sensitive thresholds may be beneficial. (Pemberton HG et al., EJNMM 2022).

As discussed in the briefing document, the Centiloid scale has become an increasingly used approach for the harmonization of amyloid PET data. The framework described by Klunk et al. (2015) includes validation of local processing against the original Centiloid method, and conversion of tracer-specific metrics such as the standardized uptake value ratio (SUVr) to a common scale referred to as Centiloid (CL). The scale is anchored on [11C]PiB SUVr data and constructed such that CL = 0 represents the mean level of amyloid PET tracer uptake in young controls, while CL = 100 reflects the average signal observed in typical mild-to-moderate AD dementia patients. This method has also been validated against neuropathological data by two independent studies (La Joie R et al, Neurology 2018 and Amadoru S et al., Alzheimer's Res Ther 2020). As mentioned in context of question 1, an adequately designed study (albeit small for centiloid subgroup) was conducted with Neuraceq, applying centiloid as semi-quantitative measure. The validity of the CL method as adjunct to visual read in detection of beta-amyloid, is therefore accepted currently (see also question 1). For other tracers the Applicant presents two studies which are stated to have validated centilod against neuropathology (Amadoru et al., Alzheimer's Res Ther 2020; La Joie et al., Alzheimers Dement. 2019) where CL<10 correlates with absence of neuritic plaques, CL>20 specified at least moderate plaque density, and >50 CL best confirmed both neuropathological and clinico-pathological evidence of AD.

Centilod as quantitation tool for amyloid burden

As the Centiloid method is based on SUVr and shares properties with derivation of uptake ratios between target and reference regions for the individual tracers, it can be agreed that the CL method is a suitable method for measuring amyloid burden in the brain. However, as is noted in the briefing document, it is not the only suitable method. Other normalised measures as z-scores may be used, and Aβ load and Aβ index are also under development (p. 11, Briefing Document). Both, Z-scores and CL, are based on SUVr and thus inherit some of its benefits and drawbacks. Z-scores and CL-scores were found to be equally useful in the AMPYPAD DPMS study, but the CL scale might be considered as a more illustrative scale and easier to comprehend.

For use in clinical trial settings with longitudinal measurements, the robustness of data would be paramount. The Applicant provided additional data to estimate "between-pipeline" variability. The procedures that define scanning and analysis procedures for the different PET tracers are summarised as "scanning pipelines" (pipeline). The Applicant refined in the discussion meeting recommendations for use of Centiloid Units in longitudinal clinical trial settings based on assessment of between-pipeline variability and this is reflected in the Qualification Opinion. Longitudinal evaluation of amyloid levels should only be performed with the same tracer and scanning pipeline.

The same fundamental challenges to PET amyloid tracer methods and the "pipeline" regarding e.g. scanner design, radiotracer accumulation, time factors, image reconstruction methods, partial volume effects and motion artefacts apply to all quantitative methods and are not limited to the factors listed.

In addition, the influence of brain atrophy and ventricular expansion or regional cerebral blood flow and corrections needed for different VOIs need to be considered. Technical recommendations as provided by the Applicant for the discussion meeting and reflected in the Qualification Opinion should be followed.

Treatment effects of amyloid targeted therapies should be evaluated in the light of these potential sources of error including the natural course of AD where atrophy of the brain inevitably increases over time. Quantitative measures of amyloid PET scans should likely always be assessed in adjunct to visual read and not least in adjunct to the cognitive symptoms of the patient.

A large number of studies are available that aim to define CL thresholds for different purposes. These should be considered when quantitative values are intended to be used, e.g., for population selection or enrichment. When using quantitative Centiloid (CL) values, a cut-off of <10 CL may be used to rule out amyloid load. The over limit >30 CL may be used to indicate pathologic amyloid levels. Values between 10 and 30 CL (grey zone) need to be interpreted with caution. Longitudinal evaluation of amyloid levels should only be performed with the same tracer and scanning pipeline. The lower threshold to rule out amyloid load can be considered established, while for an upper threshold it should be considered if sensitivity or specificity for detecting a defined amyloid load level would be of importance. Use of quantitative CL data may be considered more sensitive than using visual reads to detect changes in amyloid load and could help detecting accumulation of amyloid.

Question 3

Does EMA concur that the Centiloid measure has been sufficiently characterised for use in both research and clinical applications?

CHMP answer

As suggested in previous Scientific Advice (EMEA/H/SA/4003/1/FU/1/20 19/SME/II), the Applicant provided further data analyses from the AMYPAD program comprising the DPMS (Diagnostic and Patient Management) and PNHS (Prognostic and natural History) studies. Three sections of work were undertaken: Section A concerns the technical robustness of the Centiloid metric, section B provides cross sectional results in clinical subgroups from DMPS, PNHS and two independent studies (IDEAS and ABIDE), and section C contains longitudinal analyses of amyloid PET in DPMS and PNHS.

CHMP has critical comments in the following sections on work packages provided and acknowledges that additional material was provided for the discussion meeting. Based on all information, CHMP is of the opinion that the Centiloid measure has been sufficiently characterised for use according to the Context of Use statement.

Comments on Work section A:

Most of the Applicant's validation work (work section A) to confirm robustness was done with florbetaben und flutemetamol. It is stated that due to the Covid pandemic it was not possible to add in head-to-head comparisons between tracers. This is considered a limitation of the presented evidence. The Applicant states that in the absence of head-to-head scans acquired with both tracers, data from work packages A1 and A4 are considered the best alternatives to demonstrate the accuracy and precision of Centiloid Units. This is acknowledged.

Regarding the different analyses presented by the Applicant on robustness of the CL metric, it is acknowledged that a standard procedure is defined by the work of Klunk et al. (Klunk WE et al.,

Alheimers&Dementia 2015). It is also noted that a standard quantification "pipeline" is proposed. Nevertheless, other pipeline options could be used. Modifications of the pipeline may e.g. concern scanner design, dosing and acquisition timing, reconstruction method and brain anatomy mapping. For the discussion meeting, the Applicant presented an additional analysis of variability based on data from largely differing pipeline design (GAAIN standard vs. subject based). These data allow estimation of difference between extremes for visual read negative and visual read positive subjects for different tracers and influence of different reconstruction methods, as well as impact of atrophy. All data lead to technical recommendations that are reflected in the Qualification Opinion. The overall conclusion that the approach proposed by Klunk et al. is still appropriate to derive CL calibration equations can be supported.

A1. Evaluating the sensitivity of Centiloid quantification to pipeline design

For the analysis provided with the initial Briefing Document, the Applicant chose 4 design factors to assess robustness to pipeline design options (4 reference regions, 2 target VOIs, 2 reference region types, 2 analysis spaces). The analysis was performed using a GEE model including also tracer, MMSE and visual read results as variables. Differences in marginal means of the factors were used to assess the impact of a factor. Comparing to a difference with relevance, e.g. 2.5 CL units proposed for test-retest variability of the CL method can be endorsed, while interpretation of p-values for the factors is considered of limited importance.

PET scans from 330 participants of the DPMS with available MRI data were quantified with 32 calibrated CL pipelines. The subjects were not selected based on specific criteria, and clinical status data indicate a range of subjects from SCD+, MCI and Dementia. Additionally, analysis in subjects with positive and negative visual reads were provided.

The initially provided results show impact of reference region as a factor with relevantly lower values when using Pons as reference region. Reference region delineation also had relevant impact. It is not fully clear why the Applicant interprets marginal differences between factor levels in the range of 6 CL (e.g. between Cerebellum grey matter and Whole cerebellum + Brainstem, Table 8, p. 40, Briefing Document) between reference regions as "similar".

Overall, the analysis can be regarded informative for exploration of impact of the factors included in the analysis. Further discussion on influence of pipeline design factors and the distribution of amyloid load in the population used as new data for Level-2 calibration was provided for the meeting. In addition, a discussion of general challenges to amyloid PET acquisition was presented. The new analysis and discussion are reflected in the Qualification Opinion statement and is acknowledged.

A2. Impact of error propagation in the development of the Centiloid conversion equation

This analysis addresses the impact of errors in the CL calculation approach of different tracers that use linear regression fitting to SUVr values of a tracer and PiB (e.g. Battle MR, EJNMMI Research 2018). The analysis focuses on florbetaben and flutemetamol. The approach with bootstrap simulations using the GAAIN data to simulate 10000 datasets and adding heteroscedastic Gaussian noise together with additional use of the Jacknife method as alternative is considered appropriate. Expected differences between tracers when pooling data was estimated. For the discussion meeting, the Applicant presented an updated analysis that also included information on florbetapir. The simulations have been updated to include florbetapir using the open access GAAIN dataset (https://www.gaain.org/centiloid-project). The simulations of potential systematic bias due to the propagation of errors for florbetapir showed similar results as for florbetaben and flutemetamol. Results show that even maximum potential bias could be large in subjects with established amyloid-beta deposition (CL > 75), the bias 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). Additional (graphical) information is available in the material provided for the discussion meeting.

Results for error propagation when comparing to theoretical values show an expected influence of sample size of the equation development dataset. When using 95% confidence interval limits as metric for the assessment of maximum expected differences between tracers, the impact at 0 CL +/- 3.5 CL and at 100 CL with +/- 10.5 CL is considerable. In a range around proposed cut-off values for classification of subjects (15 to 30 CL) the impact is comparable to that at the low end of the scale.

This result illustrates the maximum potential bias of one factor in an image processing pipeline, namely the tracer calibration to the PiB reference, on potential variability. The method proposed by Klunk et al. may lead to a larger bias at the upper end of scale in the CL calibration equations because these are generated from a limited number of cases. It is acknowledged that even though significant bias may occur, likely smaller systematic bias will be typical. The systematic bias will not affect estimates of amyloid load changes over time when using a single tracer or estimates of amyloid load in cross-sectional clinical trials using only one single tracer, and will likely have limited impact on classification of amyloid status. When applying the calibration equations, users should be aware of their limitations and potential bias to avoid over-interpretation of small CL differences across tracers. Overall, these results support the recommendation in the Qualification Opinion to use only one tracer in longitudinal clinical trial settings.

A3: Cross comparison of Centiloid values from analysis pipelines used in AMYPAD

The comparison of results from different pipelines is considered very relevant to assess potential differences to quantitative analysis with different approaches in the "pipelines". The analysis includes 82 selected subjects from the DPMS and PNHS studies scanned after flutemetamol dosing. However, from Bland-Altman plot it can be deduced that the distribution of CL values is clustering in the low CL range and very limited data in the higher CL range may not allow conclusions on comparability of results. Most relevant are mean absolute differences and results of the Bland-Altman analysis. It can only be assumed that +/- 1.96 SD lines are shown in the Bland-Altman plot.

While the initially provided results indicate some degree of agreement between pipelines, the observed differences may be impacted by clustering of CL data and cannot be generalised to the complete CL range. Apparently, the scarce data in relevant CL ranges around 20 to 60 CL even exceed the 1.96 SD range in the Bland-Altman analysis. The Applicant provided additional discussion for the meeting and results from 283 subjects from the DPMS study with a more representative spread across the AD pathology continuum. In this analysis, 96% of the subjects fell with the 95% CI (12 outliers). Results overall support the importance of the recommended visual quality control.

A4. CL stability using CSF measures as anchors: Identify factors affecting CL and their impact.

This analysis compares CL data from scans with flutemetamol (N=125) and florbetaben (N=28) with pTau/Aβ42 as proxy for amyloid load. As predictions of positive visual reads with Aβ42 and Aβ42/pTau show increased variability in the low and high range of the CL scale, respectively, CL and CSF data were log-transformed before applying a linear model. This is acceptable.

Results for marginal means suggest small differences between flutemetamol and florbetaben when accounting for pTau/Aβ42 (1.09 (0.84, 1.42) CL). The indirect comparison to the proxy may limit robustness of conclusions. The Applicant clarified for the meeting that additional factors were not included by design, to obtain an upper bound of the estimate of between-tracer differences.

A5. Validation of centiloids as an adjunct to visual assessment of florbetaben PET -a multi-software analysis

This retrospective analysis focuses on florbetaben PET images acquired in subjects with at least one PET scan in previous clinical trials (N=589). Florbetaben scans were quantified with five analytical methods reporting centiloids, including the standard centiloid pipeline proposed by Klunk and co-authors. Method operator influence was minimised, and operators were blinded to clinical data and visual read results. The Applicant clarified that majority reads were used for comparisons.

Results suggest high sensitivity, specificity and accuracy presumably comparing to visual reads. Mean percentage of agreement to visual majority read was 93.2% (presumably 95% CI 0.4).

These results support the notion that quantitative reads can be used in addition to visual assessment as indicated in the SmPC of Florbetaben (Neuraceq®). The Applicant presented a new prospective read study for the discussion meeting. Visual read was performed by 5 trained readers, before and after disclosure of AMYPYPE quantitation. Besides the assessments according to reader guidelines, assessors also documented their confidence and whether quantification was supportive on a 5-point Likert scale. Data suggest value of quantitative assessments as adjunct to visual reads in settings with borderline cases and less experienced readers. Results are presented with supplementary material of the Qualification opinion.

Overall conclusion on work section A

Regarding the overall conclusion of section A, it can be agreed that the presented results on robustness of the Centiloid metric allow qualification of the method in the proposed Context of Use

Comments on work section B:

The work provided on cross-sectional data in a population that would be expected in clinical trials is obviously of high importance to assess the utility of CL values for estimating amyloid burden, classifying patients and selecting populations for clinical trial settings.

B1. The Centiloid scale applied to florbetaben and flutemetamol PET renders comparable estimates of amyloid burden in both memory clinic patients and those in the natural history study.

The Applicant used data from DPMS and PNHS for assessing CL distributions across the amyloid load spectrum. PHNS data are in the lower CL range and DPMS data cover a broad range of Cl values. Gaussian Mixture Modeling was used as data-driven approach to describe the CL data distribution and a bimodal distribution was expected due to the distribution of patients scanned in the databases, with an expected clustering of negative classifications in the PHNS data. A non-Gaussian distribution was added to better describe the cases in the intermediate CL zone, named 'gray zone' by the Applicant to improve model fit. It may be considered that this 'grey zone' is of major importance as it covers the CL range that contains the threshold values that are proposed to differentiate between positive and negative scans. The purpose of the exercise as stated by the Applicant was to compare results from florbetaben and flutemetamol as tracers. For this analysis, the models were stratified by tracer. The usefulness of the approach for comparison of tracers can be questioned. Still, the CL distribution data with fitted extended GMM curves illustrate the distribution of SCD+, MCI and dementia patients to negative, gray zone and positive as model curves in the GMM mixture model. Bootstrapping was used to calculate confidence intervals for the model parameters. This is an acceptable approach.

Results show that the curves provide comparable estimates of amyloid burden across the two tracers. The 95% confidence interval included 0 for the estimated negative Gaussian as would be expected considering the calibration of the CL scale to 0 and 100. The mean of the positive Gaussian was lower

than 100, which illustrates that selection of patients used for Level-2 calibration is of relevance. Fit data split by tracer show some difference in distributions between tracers.

Overall, the analysis is considered of exploratory value. Comparisons between tracers are not interpretable due to potential differences between the populations scanned with the two tracers.

B2. Quantitative Analysis of 6150 Real world amyloid PET scans from IDEAS.

This analysis was performed with a large independent cohort (IDEAS, N=6150 scans) with 3 approved tracers (florbetapir, florbetaben and flutemetamol). Centiloids were generated at one centre using a single pipeline (rPOP) without MRI data involved. Comparisons to local visual reads using a "pathology-based" CL threshold of 24.4CL units to define positivity independent of the visual read were made.

Results show high agreement of 86.5% (53.3% +/+, 33.2% -/-) with 13.5% discordant results (approximately equally distributed to +/- and -/+).

This analysis shows some utility of CL values for classification of patients as negative and positive. The proposed threshold of 24.4 was derived from ROC curve analysis and picked as optimal threshold based on Youden's index in an independent data set of 179 subjects (La Joie R et al. Alzheimer's Dementia 2019) For the discussion meeting, the Applicant provided additional analysis by tracer. Across all radiotracers, agreement between majority expert read and local readers of a random sample of 500 cases was excellent (kappa 0.76; 95% CI 0.73 – 0.80, p<.0001) with 86.6% (791/913) agreement for positive scans and 90.9% (532/585) agreement for negative scans. Agreement by individual radiopharmaceuticals was good to excellent: kappa 0.78 (95% CI 0.72 – 0.83, p<.001) for florbetaben, kappa 0.72 (95% CI 0.66 – 0.78, p<.001) for florbetapir, and kappa 0.78 (95% CI 0.73 – 0.84) for flutemetamol. Results indicate performance consistency for the three tracers used in IDEAs and support observations relating to the specific pathology derived Centiloid cutoff (24.4 CL)

B3. Centiloid Quantification from a second independent Clinical Cohort (ABIDE).

For this analysis scans with florbetaben were used to derive CL data in a cohort of patients with memory problems associated with different aetiologies across the spectrum of subjects with MCI (N=63), SCD (N=130) and AD/non-AD dementia (44.5%). Comparison between visual reads and CL quantified classification was performed. The CL threshold for quantitative classification was derived as optimal CL cut-off as indicated by Youden's index with CL=21 from the same data set apparently.

Results show high concordance between visual reads and quantitative reads as positive or negative (93.1% agreement). Association with etiological diagnosis was observed and results plotted by aetiology illustrate the distribution of CL data by aetiology.

This analysis has exploratory value and shows CL distributions and amyloid load for different aetiologies in a mixed cohort. Results for agreement between visual reads and quantitative reads have to be interpreted with caution, as the threshold of 21 CL proposed is derived from the same dataset. Of note, the current publication by Collij et al. (Collij LE et al., EJNMM 2021) on an analysis with flutemetamol images based on a different dataset proposes a different optimal threshold of 17 CL when comparing to visual reads (see section B4).

B4. Visual assessment of flutemetamol PET images can detect early amyloid pathology and grade its extent.

This analysis of Collij and co-authors (Collij LE et al., EJNMM 2021) used flutemetamol scans in a pooled cohort of patients from two data sources with 28.4% of the scans read as amyloid positive. They compared to visual reads from 3 expert readers as per product SmPC. Quantification of CL was

performed with the standard method and for regional CL in five anatomic regions of interest. An optimal threshold for global quantitative classification of CL=17 was derived from the same dataset using Youden's index.

Results show high agreement between visual reads and quantitative reads with a sensitivity of 97.9 % and specificity of 97.8% when using the optimal global threshold. Regional results would allow derivation of separate regional thresholds.

The results on agreement have to be interpreted with caution, as the thresholds are not derived in an independent data set. At the meeting, the Applicant clarified that use of regional data is not generally proposed for application of Centiloid Units, as the method is validated only as a global, composite measure. This is acknowledged.

B5. Visual and Quantitative amyloid-PET measures in the AMYPAD DPMS Study

This analysis uses two quantification methods for amyloid PET scans, CL and z-scores investigated in a mixed cohort of patients with MCI (N=293), SCD+ (N=220) and dementia (N=216) from the AMYPAD DPMS study. 49.9% of patients were classified as visual read positive. PET scans with flutemetamol and florbetaben were processed with GE AMYPYPE pipeline to calculate global CL and z-scores as well as regional z-scores.

Results show associations of positive visual reads with CL and z-scores and clinical stage. In patients with primary etiological diagnosis of AD higher overall amyloid burden was observed in CL and z-scores. Regional analysis for z-scores show higher amyloid burden in the pre-frontal cortex. Global CL and z-scores were highly correlated.

This analysis suggests that using both, CL or z-scores, in patients may support visual reads. Some agreement between CL and z-scores for global distributions of amyloid load can be observed from the exploratory analysis.

Overall conclusion on work section B

For Section B, the cross-sectional data support use of quantitative data in clinical trials. A large number of studies are available that aim to define CL thresholds for different purposes. These should be considered when quantitative values are intended to be used, e.g., for population selection or enrichment. No definitive thresholds can currently be singled out for a broad application.

Comments on work section C:

Two pieces of work are presented to understand the behaviour of PET tracers over time. C1 concerns the estimation of longitudinal within subject variability and C2 deals with the ability to establish a centiloid window for the prediction of amyloid accumulation.

C1. Estimation of longitudinal within-subject variability

Within subject variability was assessed either with FFM or FBB in the context of two longitudinal studies:

DPMS: time interval between scans 1.3 years, n=22 patients

PNHS: two follow-ups after 2.1 ± 0.3 (follow-up 1) and 4.8 ± 1years (follow-up two), n=46 patients

It is noted that in order to capture the variability over time a subset of individuals expected to be stable over the measurement time was selected (DPMS inclusion criteria: SCD+ at both time points, baseline CL < 10; PNHS inclusion criteria: CL <10 and VR negative at both time points, CSF aβ42/40

or aβ42 and CSF ptau negative, ApoE ε4 non carrier; MMSE at baseline for DPMS 28.6±1.3 and for PNHS 29.3 ±0.9). Although the variability was low in both cohorts (~3CL/year) with an ICC of 0.82 (DMPD) and 0.86 (PNHS), this was shown only for a small subset of patients that met criteria for this analysis and it is unclear how the variability would be in less stable patients. The Applicant clarified at the meeting that data with ~4-year follow-up have been presented for this analysis, as the rates of change were computed using generalized estimating equations with all the time points available per subject

C2. A Centiloid window to help predict true amyloid accumulation.

The Applicant concludes in this work package that baseline CL can help identify subjects more likely to accumulate pathology and could therefore assist subject selection and therapy response monitoring in clinical trials. Three different approaches were used to model longitudinal change in CL in 686 cognitively unimpaired individuals, 1. based on PET visual read, 2. based on baseline CL load, 3. based on rates of amyloid accumulation (≥3.3 CL/year). The Applicant states that individuals with a baseline CL in the Grey-zone (12≤CL≤50) show a similar pattern than the VR groups ($\beta_{Stable\ VR-}$=0.3; $\beta_{Converters}$=4.9, p<.005). It is unclear what the advantage versus VR exactly is. The Applicant clarified on an analysis among individuals with baseline VR- and CL in the Grey –zone, from whom 30 % will convert to amyloid positivity in the cohort of patients analysed in this work package.

Overall Conclusion on work section C

It can be agreed that the longitudinal analysis of DPMS and PNHS data explores factors that are relevant for understanding trajectories of measurement with PET tracers over time.

While it is agreed that there is value in measuring pharmacodynamic response on amyloid load and that the establishment of an optimal window for the selection of patients for anti-amyloid treatments could be theoretically helpful, the link to the clinical response (e.g. cognitive and/or functional outcome parameters) is missing. Hence, the provided analyses do not allow to estimate the risk of preclinical or MCI progression to AD or to determine whether the use of PET and/or CL metric is associated with an improved clinical outcome (see also Rabinovici 2019 IDEAS). The prognostic or predictive value cannot be estimated with the current data. For the discussion meeting, the Applicant proposed to exclude assessment of prognostic risk of future cognitive decline from the Context of Use. This is agreed.

Question 4

Does EMA concur that the body of evidence provided by AMYPAD supports the diverse utility of the Centiloid metric as a means for example to (i) support the current visual inspection of tracers as an adjunctive tool, (ii) for the consistent inclusion of patients for AD targeted therapies and (iii) to provide a potential baseline measure for future therapy monitoring/follow up scanning as indicated in the context of use summary?

CHMP answer

The Applicant initially proposed the Centiloid Unit as a universal metric for the assessment of brain amyloid burden. The Context of Use statement was refined after discussion with the Applicant.

Regarding the initial questions from the Applicant the following statements can be provided:

i. Support the current visual inspection of tracers as an adjunctive tool.

It is agreed that the use of quantification software tools could be beneficial when images are assessed by more inexperienced readers, or when amyloid levels of patients are close to pathology thresholds. Importantly, quantitative information should always be used as an adjunct to visual read, not least since atrophy of the brain may lead to lower tracer uptake than expected with regards to the patient's clinical disease stage. If there is a discrepancy between the reader´s visual assessment of the PET scan images a number of steps should be taken to assess the differences between visual read and quantitative information as described by the manufacturer of the tracer. However, visual read will have primacy if a discrepancy still exists. As outlined in the answer to question 1, PET tracers are already qualified for enrichment in clinical trials across the AD spectrum from pre-dementia to mild to moderate AD (EMA/CHMP/SAWP/893622/2011 and EMA/CHMP/SAWP/892998/2011). It is agreed that the use of the Centiloid metric as adjunctive tool to visual reads in research trials could add granularity to the information when defining thresholds for amyloid.

ii. For the consistent inclusion of patients for AD targeted therapies.

In Europe until now no amyloid targeted therapies are approved, however PET imaging data including the CL metric have already been widely used for trial enrichment and assessment of pharmacodynamic response in early AD (Mintun et al. 2021; Budd Haeberlin et al. 2021, Swanson et al 2021, Karran and De Strooper et al. 2022). Since anti-amyloid treatments will only be successful on patients with established amyloid positivity, the use of amyloid biomarkers including the CL metric as diagnostic tool for amyloid targeting therapies is therefore plausible. An optimal universal threshold is currently not available. Even though only demonstrating amyloid positivity on amyloid PET scan without quantification might suffice to diagnose and qualify the patient for disease modifying therapies targeting beta amyloid, it is agreed that a quantitative measurement such as CL might provide valuable additional information for the consistent inclusion of patients for AD targeted therapies and possibly also to identify the optimal window for therapeutic intervention.

iii. To provide a potential baseline measure for future therapy monitoring/follow up scanning.

Use of the CL scale can provide a potential baseline measure for future therapy monitoring/follow up scanning. However, the clinical utility would depend on the clinical data that need to be generated for the specific future therapy. Amyloid PET is only one of the tools used to monitor patients receiving therapies against AD, since it is still not uncontroversial if lowering beta amyloid actually translates into clinically relevant treatment effects and if yes, to which magnitude. Therefore, any use, e.g. for surrogacy of efficacy or monitoring treatment response, is currently premature. Furthermore, the complex interactions between Amyloid beta, tau, neurodegeneration, and neuroinflammation and their relationship to the AD clinical syndrome are still being untangled and it is unclear how this interplay influences any prognosis with respect to rate of decline or predictions to future treatment response. Also, the influence of negative health and lifestyle factors would need to be taken into account (Bischof & Jacobs, 2019).