FINAL PUBLISHABLE REPORT

Grant Agreement number	18HLT09
Project short name	NeuroMET2
Project full title	Metrology and innovation for early diagnosis and accurate stratification of patients with neurodegenerative diseases

Project start date and duration:		July 2019, 42 months
Coordinator: Will Webster, LGC Tel: +44 20 8943 7523 E-mail: william.webster@lgcgroup.com Project website address: https://www.lgcgroup.com/our-programmes/empir-neuromet/		
Internal Funded Partners: 1 LGC, United Kingdom 2 LNE, France 3 PTB, Germany 4 RISE, Sweden	External Funded Partners: 5 Charité, Germany 6 CHU Mpt, France 7 Modus, France 8 Uni-Greif, Germany 9 VUmc, Netherlands	Unfunded Partners: 10 HKR, Sweden
RMG: -		

TABLE OF CONTENTS

1 Overview

Neurodegenerative diseases (NDDs), such as Alzheimer's disease (AD), are a group of chronic, progressive disorders, which lead to deficits in specific brain functions including cognition and movement. NDDs can affect people of all ages and are one of the most pressing medical issues of the modern time. The NeuroMET2 project addressed the challenges associated with diagnosis of AD by building upon the results of the preceding EMPIR 15HLT04 NeuroMET project and its unique patient cohort to apply metrological principles of ADdiagnosis. The most promising NeuroMET minimally invasive methods for early diagnosis of AD were advanced through longitudinal studies and transferred to clinical settings. The project also developed novel approaches and reference measurement procedures (RMP) to address the current measurement challenges of early NDD diagnostics and therapies.

2 Need

There are over 9.9 million new cases of dementia each year worldwide, with one new case every 3.2 seconds and a desperate need for treatments. NDDs are one of the leading medical and societal challenges faced by European society with costs for care currently around **€130 billion per annum.**

Research suggests that the brain changes associated with AD (as with many other NDDs) begin fifteen or more years before symptoms appear, and that treatment of NDDs is typically most effective when started at the early stages. Translational research is however needed to fine-tune the methods developed and define their measurement uncertainty and accuracy before they can become fit for clinical use.

The 15HLT04 NeuroMET project was the first metrological project to combine the diverse expertise of NMI/DIs together with that of clinicians and academics to build the infrastructure required to translate research into clinical or pharmaceutical settings and overcome specific metrological barriers in NDD diagnosis and treatment. In the 15HLT04 NeuroMET project, person-centred outcome measures (PCOMs) were, for the first time, metrologically validated using clinical laboratory data, RMP for protein biomarkers and ultra-high field magnetic resonance imaging (MRI) and spectroscopy (MRS) protocols. All these have high potential for use in early diagnosis or to facilitate the uptake into clinics and industry of novel assays. However longitudinal studies were required to determine their prognostic value. The fully characterised 15HLT04 NeuroMET patient cohort (90 individuals) and data associated to it constituted an invaluable European and international resource to carry out longitudinal studies and validate methods and biomarkers developed elsewhere. It was therefore important that the cohort was maintained in order to address the most up to date NDD measurement issues and validate new assays.

This project addressed needs for: (i) screening programs for early NDD diagnosis; (ii) RMPs and protocols to facilitate implementation of new assays into clinics and improve differentiation by reducing measurement uncertainty; and (iii) improved specificity of drugs by developing new methods for monitoring NDD protein aggregation.

3 Objectives

This project aimed to consolidate and further develop the 15HLT04 NeuroMET metrological infrastructure and validate biomarkers and procedures for early NDD diagnosis and accurate patient stratification, leading to new patient screening programs and increased rate of success in clinical trials. The specific objectives were:

1. To maintain the already established and stratified 15HLT04 NeuroMET patient cohort and enrol new patients to cover the defined stages of neurodegeneration and account for patient drop-out. Patients will be clinically assessed, and blood as well as cerebrospinal fluid (CSF) samples will be collected and distributed to partners for longitudinal studies in Objective 2, 3 and 4.
2. To advance the 15HLT04 NeuroMET PCOMs for cognition and early diagnosis. This will lead to the validation of a 'NeuroMET Memory Score' and the development of an app (software application downloadable onto mobile devices) for clinicians and patients to deliver validated cognitive tests.
3. To refine ultra-high field MRI and MRS protocols from 15HLT04 NeuroMET through longitudinal studies for application into clinics. Additionally, new *in vivo* approaches will be developed to monitor supplemental biomarkers in the project cohort.

4. To advance biomarker measurements for early and accurate diagnosis through the validation and implementation into clinics of the 15HLT04 NeuroMET methods and other methods. Biomarkers such as Aβ1-42, Aβ 1-40, neurofilament light chain (NfL), total-tau (t-tau) and α-synuclein will be monitored in the NeuroMET cohort and new RMP for NfL and *p*-tau will be developed. Methods for monitoring aggregation of NDD proteins will be also developed and validated to improve specificity of therapeutic targets and as potential diagnostic tools.
5. To enhance Causal Rasch mathematical models to define prototype metrological references for cognition expressed as "construct specification equations (CSE)". This will provide an extensive explanation of how able a human can act as an "instrument" when measuring the difficulty of a task such as a cognitive test. Those models will be applied to the PCOMs, MRI and MRS, biomarker data to define and improve the prognostic values of the methods developed.
6. To transfer the project's results to the measurement supply chain, standards developing organisations (ISO/TC212, the International Federation of Clinical Chemistry (IFCC), and the Joint Committee for Traceability in Laboratory Medicine (JCTLM)), instrument manufactures and end users (e.g., clinical laboratories and pharma) and promote the 15HLT04 NeuroMET multidisciplinary infrastructure to become the ideal space for NDD translational research.

4 Results

This project, NeuroMET2 built upon the success of the preceding EMPIR 15HLT04 NeuroMET project to further apply metrological concepts and procedures to NDD diagnosis.

The structure of the NeuroMET2 project is shown below and the key outputs of this project have included:

- Maintenance and further development of the metrological infrastructure and network of multidisciplinary laboratories established under 15HLT04 NeuroMET
- Development of enhanced tools to enable early NDD diagnosis and accurate patient stratification in cognitive assessments (Objectives 1, 2 & 5), MRI and MRS (Objective 2), and fluid biomarkers (Objective 3)
- Development and validation of mathematical RMP to provide a reference for cognitive ability (Objective 5)

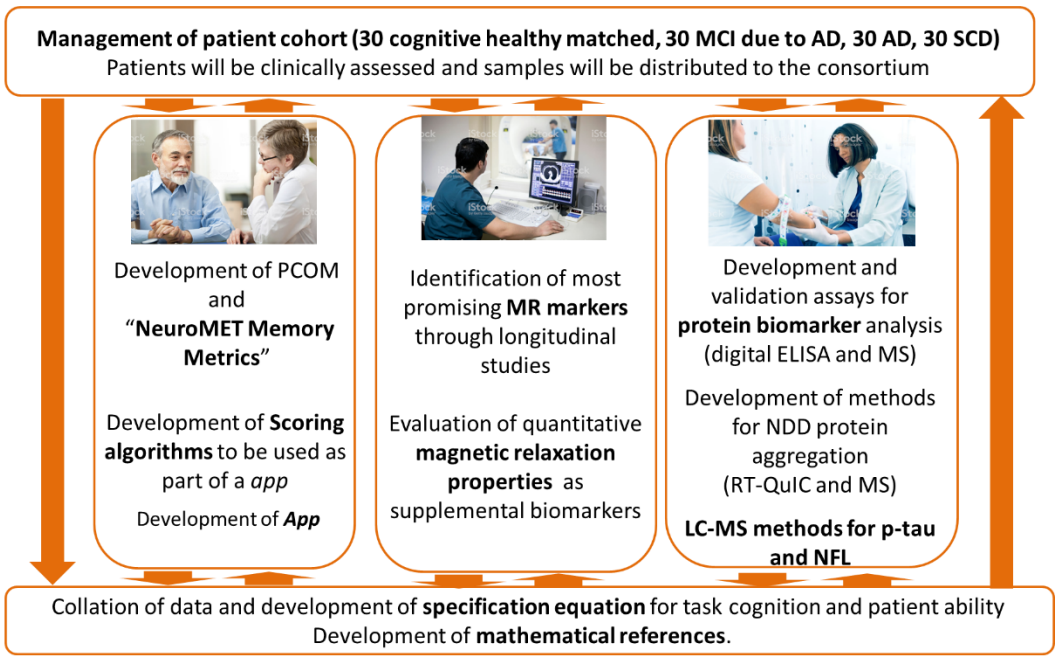

Figure 1. Schematic representation of the project

4.1 Objective 1: To maintain the already established and stratified 15HLT04 NeuroMET patient cohort and enrol new patients to cover the defined stages of neurodegeneration and account for patient drop-out. Patients will be clinically assessed, and blood as well as CSF samples will be collected and distributed to partners for longitudinal studies.

Patient cohort recruitment

The NeuroMET2 project built upon the unique patient cohort which was recruited by Charité for the 15HLT04 NeuroMET project. At the conclusion of 15HLT04 NeuroMET in 2019, a total of 39 Healthy Controls (HC), 23 Mild Cognitive Impairment (MCI) patients and 26 Alzheimer's Disease (AD) patient had been recruited.

NeuroMET2 aimed to develop the patient cohort to cover at least 30 individuals in the HC, MCI, and AD groups; and maintain these numbers accounting for attrition (e.g recruitment of new patients to offset patient dropouts due to immobility in advanced stages of the disease, or death). An additional aim was to recruit a group of 30 patients with Subjective Cognitive Decline (SCD) from a pre-existing SCD cohort at Charité.

Participants aged between 55 and 90 years old were recruited by (i) neurologists of the Charité memory clinic, (ii) external neurologists, (iii) advertisements on websites and distributed flyers, or (iv) from a preceding study at Charité named SmartAge.[1]

Participants were stratified into one of the following groups: SCD, MCI, AD, or HC.

- SCD participants reported persistent self-perceived cognitive decline for over 6 months and associated worries that would motivate the individual to seek medical help, analogue to SCD plus criteria proposed by (14).
- Participants of the MCI and AD group were required to have received a diagnosis by a neurologist and scored around 1 (MCI) or 2 (AD) standard deviations less in standardised memory-related tests.
- AD participants additionally showed impairment in daily life due to cognitive deficits. Besides severe or untreated medical, neurological or psychiatric diseases which could potentially interfere with cognition, exclusion criteria were history of drug or alcohol abuse, and eating disorder.

All participants were native German speakers.

During the first participant visit blood, saliva, and CSF samples (when applicable) were collected. Where possible participants came in for the second visit within the same week and underwent a neurological and neuropsychological assessment using the most promising tests from 15HLT04 NeuroMET; including the Mini Mental State Examination (MMSE), Digit Span, Corsi Block tapping etc.

Where possible (see note effects of COVID-19, below) participants were scanned at PTB using the 7T scanner. Participants were also given a routine medical examination at Charité, that included standard blood tests to rule out a number of different diseases such as infections, anaemia, leukaemia. When applicable standard CSF measurements related to AD (Aβ-40, Aβ-42, t-tau and phosphorylated-tau (p-tau)) were also carried out.

Longitudinal Studies

In total, we recruited 129 participants (Table 1). Of these 129, two participants did not meet the inclusion criteria (too young n = 1, psychiatric disease n = 1) and were therefore not stratified into the groups and not followed-up and were excluded from analyses.

Additionally, we performed 186 follow-up visits after 1 year (T2), 3 years (T3), 4 years (T4) and 5 years (T5).

At follow-up, participants underwent the same of battery tests as at baseline visit (with the exception of T2, which did not include blood collection or MRI).

Visit	Total	HC	SCD	MCI	AD
T1	129	35	35	30	27
T2	102	28	33	22	19
T3	44	17	14	6	7
T4	31	15	10	3	3
T5	9	6	2	0	1

Table 1 - Number of visits in the NeuroMET cohort for whole cohort (Total) and group-wise (HC, SCD, MCI, AD). Participants included in visit T1 who did not meet the inclusion criteria (n = 2) were excluded from the group-wise counting.

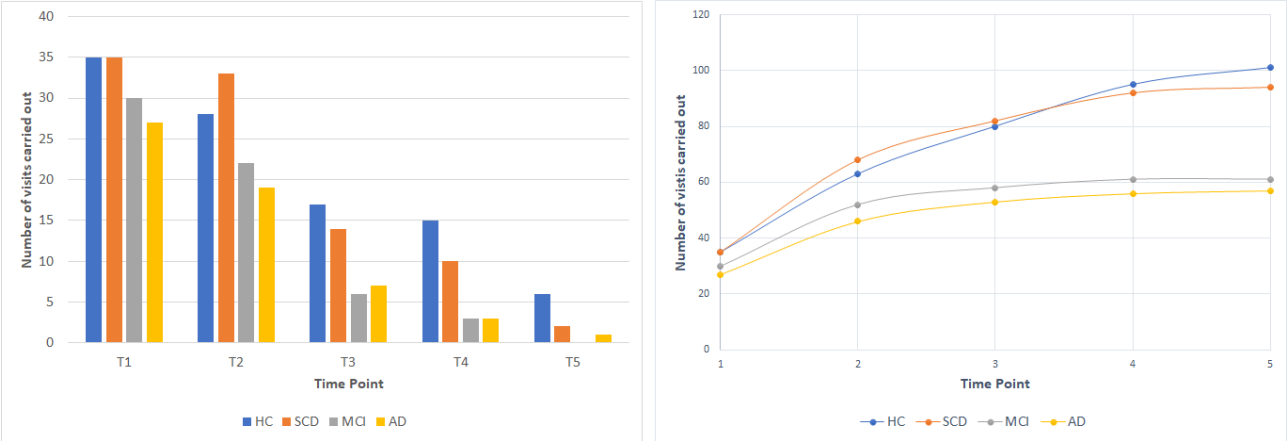

Figure 2 – cohort visits carried out, by time-point (left) & cumulative (right)

	Total N = 127	HC N = 35	SCD N = 35	MCI N = 30	AD N = 27	p-value
Age [years]	72 (7)	71 (8)	69 (7)	71 (6)	75 (6)	0.005 [1]
Female	63 (50%)	18 (51%)	22 (63%)	8 (27%)	15 (56%)	0.027 [2]
Education [years]	15 (3)	15 (3)	16 (2)	15 (3)	14 (3)	0.200 [1]
APOE e4 carrier	52 (41%)	9 (26%)	11 (31%)	17 (57%)	15 (56%)	0.018 [2]

[1] Kruskal-Wallis rank sum test
[2] Pearson's Chi-squared test

Table 2 – Patients' characteristics at baseline. Values are reported in mean (SD) or N (%).

Effect of the COVID-19 Pandemic

The COVID-19 pandemic had a significant impact upon the recruitment and characterisation of the cohort at Charité. Effects included:

- All patient visits to Charité for neuropsychological testing and blood and saliva sample collection were suspended between February and June 2020, and again between December 2020 and March 2021.
- Patient visits to PTB for 7T MRI were similarly suspended and could only be resumed at a lower throughput level due to the need to maintain a strict hygiene regimen within the scanning facility.

This is reflected in the trends observed in Figure 2.

Distribution of fluid biomarker samples

The collected blood, CSF, and saliva samples were sent to partners CHU Mpt, LGC, LNE, and VuMC for use in RMP and clinical method development (Objective 3) and to generate biomarker data to feed into mathematical models being developed by RISE and Modus (Objective 4).

Conclusions

The project met Objective 1, to maintain the already established and stratified 15HLT04 NeuroMET patient cohort and enrol new patients to cover the defined stages of neurodegeneration and account for patient drop-out. Patients were clinically assessed, and blood and CSF samples were collected and distributed to partners for longitudinal studies.

From the 90 participants recruited in 15HLT04 NeuroMET, 42 participants were followed up in this project. Combining the NeuroMET and NeuroMET2 cohorts, the final dataset comprises of 129 participants with a total of 315 visits. The NeuroMET patient cohort represents a unique resource, as it is the first ever metrologically characterised cohort in the field of AD research.

The collected blood, CSF, and saliva samples (from participants in the NeuroMET cohort) were distributed to partners for use in RMP and clinical method development (Objective 3) and to generate biomarker data to feed into mathematical models (Objective 4).

4.2 To advance the 15HLT04 NeuroMET PCOMs for cognition and early diagnosis. This will lead to the validation of a 'NeuroMET Memory Score' and the development of an app (software application downloadable onto mobile devices) for clinicians and patients to deliver validated cognitive tests.

Development of the NeuroMET Memory Metric (NMM)

Under the preceding EMPIR 15HLT04 NeuroMET project, a prototype NMM based on legacy cognitive PCOMs was developed. This project aimed to further develop and validate this metric, and then develop an app which could be used to roll the new metric out to the clinical community.

Commonly used legacy tests have been shown to lack discrimination power between levels of cognitive performance, failing to detect early cognitive decline.[2] By bringing together different types of memory ability tests under one common measurement system, the NMM enables a more precise and valid estimation of a persons' locations on the cognitive spectrum.

The project's NMM is comprised of a careful selection of items from six legacy memory tests and five cognitive performance tests: the Corsi Block Test (CBT) forward items, the Digit Span Test (DST) forward items, the CBT, the Rey Auditory Verbal Learning Test (RAVLT) first trial A list, a subset of the Mini-Mental State Exam (MMSE), and the CERAD battery Word Learning List in the CERAD test battery (WLL CERAD) first trial as well as the MMSE memory items. The RAVLT C and D lists, which are composed of different words than the A list, were originally part of the selection, but were not included in the NMM (due to reasons explained below).[5] [28] [29]

Composition of the NMM by careful selection was guided by prediction of how each item can "fill the gaps" on the task difficulty axis and the "equivalence" of items from different tests based on our best understanding of each item as explained with entropy-based CSE (see Objective 5).

Rasch Model

The preceding EMPIR 15HLT04 NeuroMET project demonstrated the advantages of applying Rasch Measurement Theory (RMT) to neuropsychological assessments. RMT separates estimates of person and item attributed values and their scaling on the same interval logit scale. Modus, in collaboration with RISE, conducted a Rasch analysis of the items that were selected for the NMM for early and accurate diagnosis of NDD.

The measurement properties and calibration of the NMM was performed using the Rasch model. The below model (Equation 1) is a probability formula for the outcome of a person, *i*, answering a dichotomous item, *j*:

$$P(X_{in} = x) = \frac{e^{x(\beta_n - \delta_i)}}{1 + e^{\beta_n - \delta_i}}, \qquad x = 0, 1.$$

Equation 1

The Rasch model, also known as the 1-parameter logistic model because it has one parameter (*δ_jδ_i*) for each item, has sufficient statistics for all of its parameters. This important property allows item estimation that is independent of the sample distribution through an estimation method called conditional maximum likelihood (CML).[3]

Derived item locations are therefore independent of the sample, up to its size, in that standard errors for item estimates are inversely proportional to the number of people in the sample. Although CML is the only method to have person-free item estimation, studies on well-targeted samples have shown that results from other estimation methods (JML, CML) are comparable, as long as the underlying population is not mis-specified.[4] Work at RISE has also demonstrated the links between the Rasch model and measurand restitution.[5] [6] [7]

Properties and potential benefits of the NMM

The overall item hierarchy revealed in Rasch model analysis was concurrent with clinical expectations. A benefit of the NMM is that it unifies individual instruments into one common metric. With all items calibrated with each other, crosswalks and adaptive testing are also possible. Crosswalks enable direct comparison between scores obtained on different legacy tests included in the NMM.[8] The NMM is a valuable tool as it brings together existing legacy tests into one common measurement system for more reliable, traceable, and comparable assessments of memory ability.

NMM as a diagnostic tool

The NMM shows promising results as a potential diagnostic tool, although it must be noted that this is early exploratory research and analyses will require confirmation in larger studies. Patients who were administered items from the NMM were classified as either healthy control (HC), subjective cognitive decline (SCD), MCI, or AD. Results from this first round of analyses show a clear group distinction in patient locations on the NMM metric. This can be seen in Figure 3.

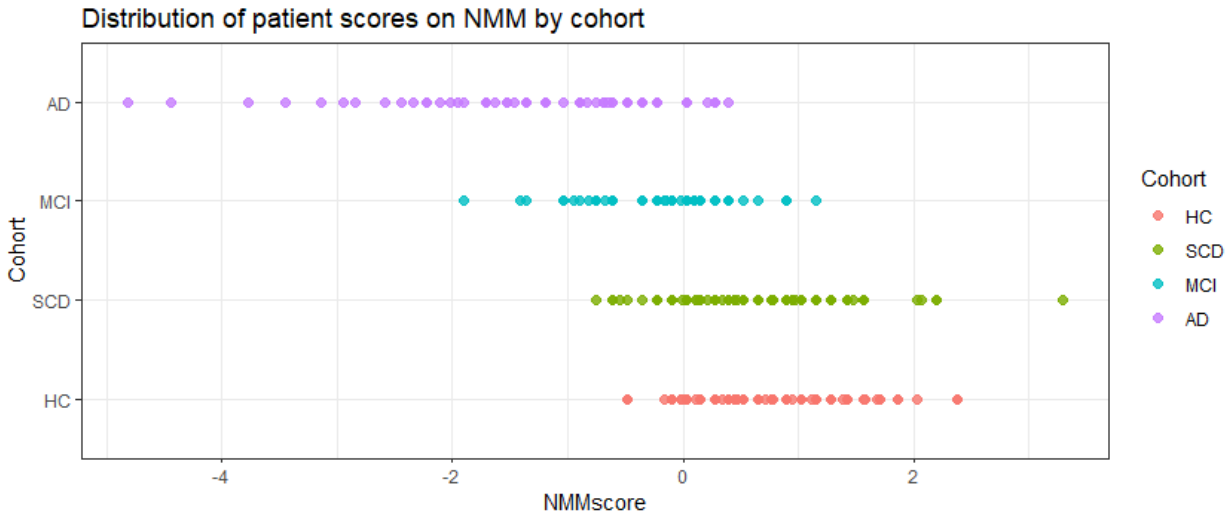

Figure 3 - Distribution of patient scores on NMM by cohort

Development of an app to deliver cognitive tests for the NMM

The development of an electronic delivery system for the items and scoring conventions of the NMM is key output of the NeuroMET2 project. The app was formally known as NMM electronic Clinical Outcome Assessment (NMM eCOA) and provides a modality of data collection (i.e., responses to items from patients, clinicians, and observers) using a handheld, tablet, computer, or associated devices.

The NMM eCOA project team was assembled of individuals from THREAD Research (a collaborator linked to Modus), Modus and Charité. THREAD was chosen to develop the NMM eCOA in order to harness the features of their pre-existing platform that provides configurable data collection methods for eCOA. Modus was responsible for managing this overarching task and liaising with study partners. The THREAD team worked on the NMM eCOA development, and members from Modus and Charité teams oversaw the development, provided input, feedback, and assistance with training. RISE, who led on the NMM development, also provided high-level feedback throughout the duration of the project. The development process is outlined in figure 5, below.

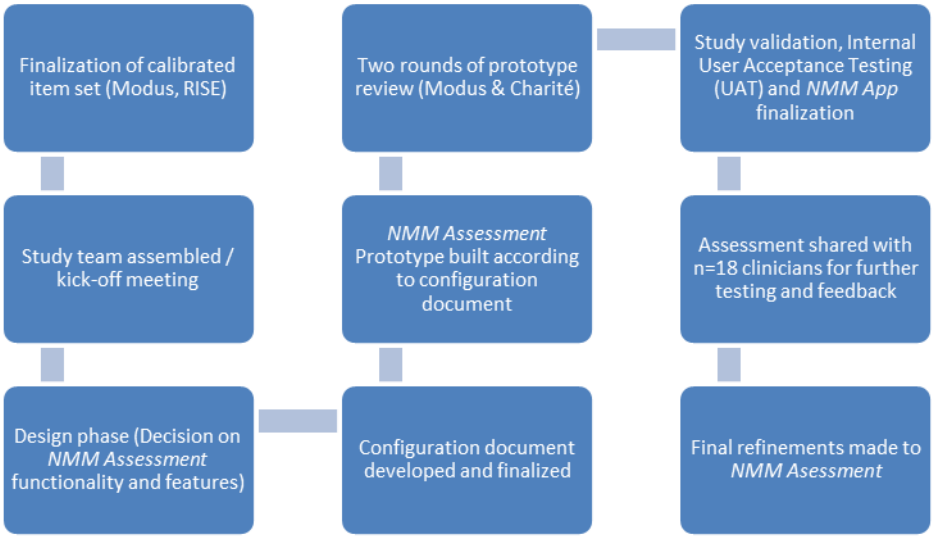

Figure 5 - NMM eCOA development flowchart

Design

The purpose of the design phase was to align on features and functionality from the THREAD platform to be implemented in the study. It was decided that the NMM eCOA would be developed using the THREAD platform and based on the requirements of the NMM, that it would use the clinician reported features only. Further decisions were made about how the NMM eCOA would be seen by clinicians, the order of tests, how data would be entered into the platform, and the scoring capabilities of the NMM eCOA.

The resulting NMM eCOA was built to be compatible with a Google Chrome browser and designed to be used either on a laptop or iPad. The individual memory tests are programmed all in one activity, in the following order: CBT, DST, CERAD MMSE, CERAD WLL CERAD, RAVLT Version A.

The steps for a clinician to use the assessment were designed to be as follows: (i) Log into the THREAD platform; (ii) Create a new participant; (iii) Open the NMM form and fill in the test results as the participant answers; (iv) Export the data. The portal is simple and user-friendly and has the potential to be an eventual alternative to filling out a variety of different batteries on paper. The NMM eCOA has capabilities to produce a total score (sum of all responses), with conversion to the person's memory ability measured with the NMM conducted subsequently. Therefore, the decision was made that scoring would be performed following data export in a CSV file.

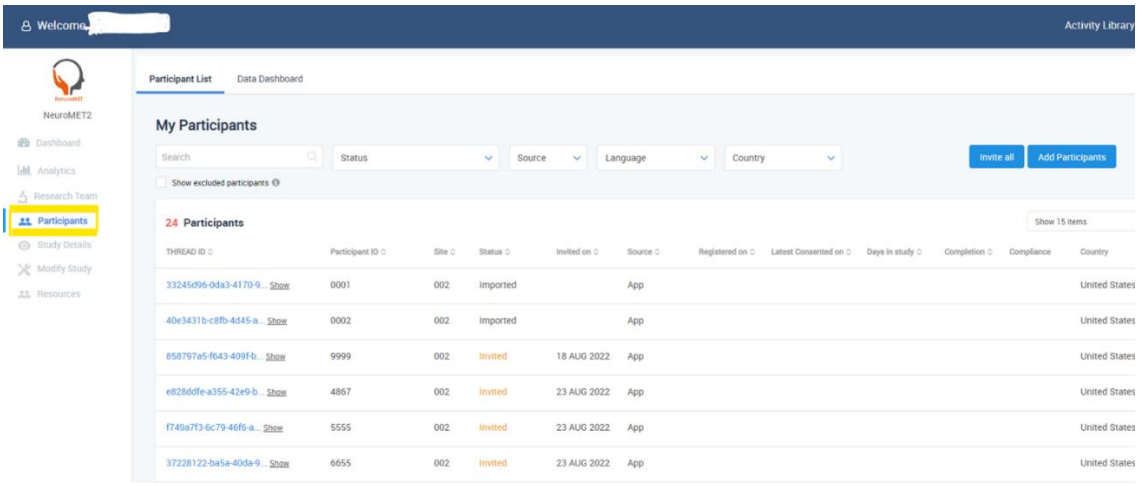

Figure 6 – Screenshot from the NeuroMET app: sample participant list

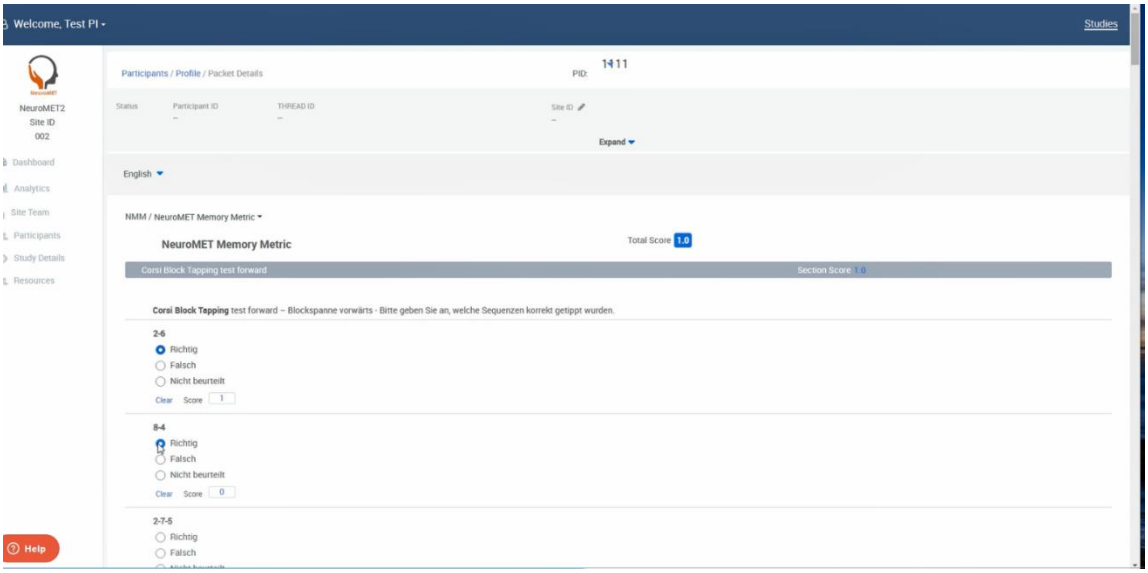

Figure 7 – Screenshot from the NeuroMET app: example interface: entering data

Testing of the NMM eCOA in clinical settings

Following the development cycle and user acceptance testing, the NMM eCOA was transferred to Charité for external training, testing and feedback. An online training course was delivered to expert clinicians to gain their feedback on the NMM eCOA.

Access to the NMM eCOA was granted to eighteen healthcare professionals from the NeuroMET2 partners, including those from Charité (n=11) and Uni-Greif (n=6), but also from external collaborators Fachklinik Briese (n=1), and Immanuel Klinik Rüdersdorf (n=1). The participants were from a variety of backgrounds, including neuroscientists (n=11), neurologists (n=4), psychologists (n=2), project managers (n=1) and study assistants (n=1).

The training was distributed by e-mail and performed digitally. Participants received an e-mail that included all necessary information about the study, including background information on the NMM and THREAD, a training video and training tasks, and a request to fill in a feedback form. The results demonstrated that the project was successful in developing a user-friendly eCOA that clinicians can easily use to deliver the newly developed NMM.

Conclusions

The project met Objective 2, to advance the 15HLT04 NeuroMET PCOMs for cognition and early diagnosis through the validation of a 'NMM' and the development of an app for clinicians and patients to deliver validated cognitive tests.

Under the preceding EMPIR 15HLT04 NeuroMET project, a prototype metric based on legacy cognitive PCOMs was developed. The 18HLT09 NeuroMET2 project further developed and validated this metric using data obtained in Objective 1. The project also developed an app which could be used to roll the new metric out to the clinical community.

The NMM and its associated app represent both a novel and more accurate measurement system for cognitive testing. They are also an important and new method for cognitive testing through an app for use by clinicians and patients.

4.3 Objective 3: To refine ultra-high field MRI and MRS protocols from 15HLT04 NeuroMET through longitudinal studies for application into clinic. Development of new in vivo approaches to monitor supplemental biomarkers in the project cohort.

Neuroimaging techniques are very promising non-invasive approaches for early detection of NDD and can be used to improve understanding of underlying principles of neurodegeneration. Neuroimaging techniques are also currently used in routine practice for diagnosis of AD.

The 15HLT04 NeuroMET project identified several promising MR-derived biomarkers for AD. These included the Grey Matter (GM) volume adjusted for the total intracranial volume, the average cortical thickness of the brain, the volumes of the left and right hippocampus, and the left and right posterior cingulate gyrus. The 15HLT04 NeuroMET project also highlighted a number of neurometabolite markers which may have significance in monitoring AD progression, including myo-inositol, N- acetyl-aspartate, γ-amino butyric acid (GABA), and glutamate. The NeuroMET2 project aimed to build on this work, applying fundamental metrological concepts to MR measurements and applying the enhanced techniques developed to both longitudinal studies to evaluate biomarker significance, and to measurements made in clinical settings.

Evaluation of Measurement Uncertainty

The main goal of this work was to generate a statistical analysis framework to investigate the impact of different MRS parameters on the reproducibility and repeatability of measurements *in vivo*. The impact of different adiabatic pulses on the variance of *in vivo* measurements was selected for the study.

The conventionally used hyperbolic secant adiabatic inversion pulse (HS) has the disadvantage of a large chemical shift displacement (CSD) and high necessary peak voltage. These drawbacks can be mitigated by using gradient-modulated pulses such as the gradient offset independent adiabaticity (GOIA) and the wideband, uniform range, smooth truncation (WURST) pulse. This a HS, GOIA, and WURST pulse was designed[9] [10] [11]in MATLAB (The MathWorks, Natick, MA, USA) to achieve identical pulse duration, inversion slice thickness, and pulse energy.

Nine healthy volunteers/participants (aged 39 ± 13, 1:7:1 male:female:nonbinary) were scanned using a 7T whole body MR system (Magnetom 7T, Siemens Healthineers, Erlangen, Germany) equipped with a 1-channel transmit/32-channel receive head coil (Nova Medical Inc., Wilmington, MA, USA). All participants gave written informed consent according to local ethics regulation requirements. Each volunteer was scanned in two sessions on two different days approximately one week apart (six to eight days). Both sessions consisted of two measurements (M1-M4) of SPECIAL acquisitions with the HS, GOIA, and WURST adiabatic inversion pulses, each.

During session one, the volunteer was repositioned between M1 and M2, while in session two M3 and M4 were acquired without repositioning in between. A schematic overview of the different measurements, scan blocks, and sessions is shown in Figures 8c-d.

The order of the SPECIAL versions within the different scan blocks was cyclically permuted between the different volunteers to ensure that the performance of the pulses was not biased due to the acquisition time point within the protocol, e.g., due to increased likelihood of volunteer movement towards the end of each scan block. With this design, it was possible to distinguish between three scenarios:

i. the repeatability11 (R_0), which refers to two consecutive measurements without repositioning the subject
ii. the reproducibility11 between two measurements performed on the same day including repositioning and new calibration (R_(1,M) for minutes in-between)
iii. the reproducibility between two measurements approximately one week apart (R_(1,W) for week in-between).

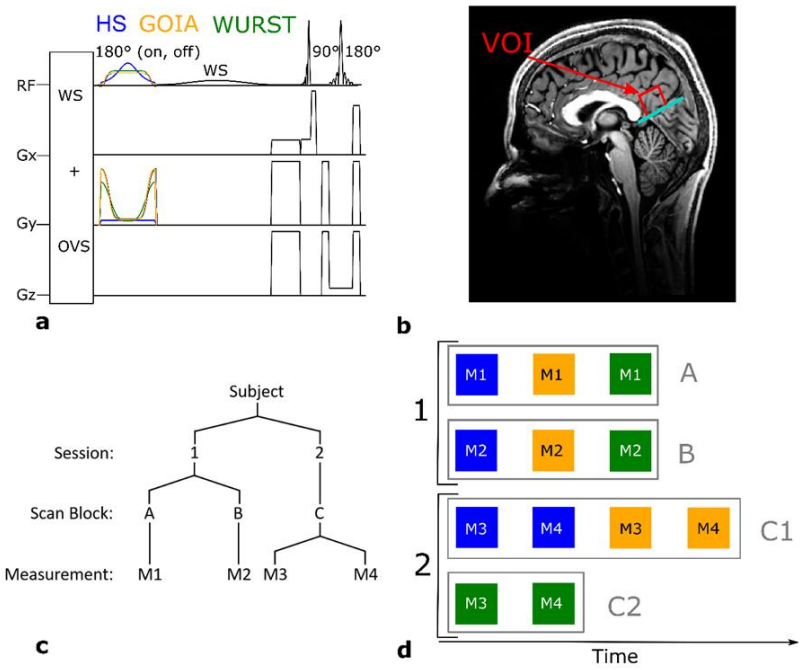

Figure 8- a) Pulse sequence diagram of SPECIAL with different adiabatic inversion pulses: HS (blue), GOIA (orange), and WURST (green). b) Exemplary voxel position in the posterior cingulate cortex (PCC). The turquoise line indicates the connection between the lower edge of the corpus callosum and the outer edge of the parieto-occipital fissure. c) Unbalanced nested study design performed for every pulse sequence variant. The subject-wise between-session reproducibility (R1,W); M1 and M3), the between-positioning reproducibility (R1,M; M1 and M2), and the repeatability (R0; M3 and M4) were assessed. d) Scan scheme (exemplary for the first three volunteers): On the first day in the first session (M1), SPECIAL with HS, GOIA, and WURST was measured. After repositioning the volunteer (M2), the sequences were measured in the same order as in M1. On the second day, i.e., one week later, in the first session, HS- and GOIA-SPECIAL were measured twice without repositioning (M3 and M4). Then, the SPECIAL sequence using the WURST pulse was measured twice without repositioning. Note that the repeatability measurements are split into two scan blocks due to time restrictions in our ethical regulations. Figure was published by Riemann et al.[12]

The acquired MRS data from the 9 healthy volunteers were analysed using a restricted maximum likelihood estimation (REML) and the analysis used to generate results for the calculation of the standard deviation of the metabolite concentrations such as GABA, glutamine, glutamate, glutathione, myo-inositol, lactate, N- acetylaspartate, N-acetylaspartylglutamate, myo-inositol, choline containing compounds, creatine, taurine, and phosphorylethanolamine.

The results indicated that Cramér-Rao Lower Bounds (CRLBs), which are commonly used to assess the reliability of metabolite concentrations from MRS, appear to correlate with the standard deviation calculated from the REML method. However, the CRLBs appear to underestimate measurement uncertainty. The results present a methodology to estimate the measurement precision of in vivo metabolite concentrations obtained by MRS, and consequently the MDCs for 13 metabolite concentrations in vivo for the used setup.

Metabolite	MDC / µmol g^{-1}		Metabolite	MDC / µmol g^{-1}		Metabolite	MDC / µmol g^{-1}
Asp	1.87		Ins	1.66		tCho	1.92
GABA	1.20		Lac	0.59		tCr	1.46
Gln	1.22		NAA	2.16		Tau	0.85
Glu	2.23		NAAG	0.92			
GSH	0.66		PE	0.40			

Table 4: The minimal detectable change (MDC) for each metabolite, derived by the restricted maximum likelihood estimation (REML) analysis, is shown. Table was published by Riemann et al.[12]

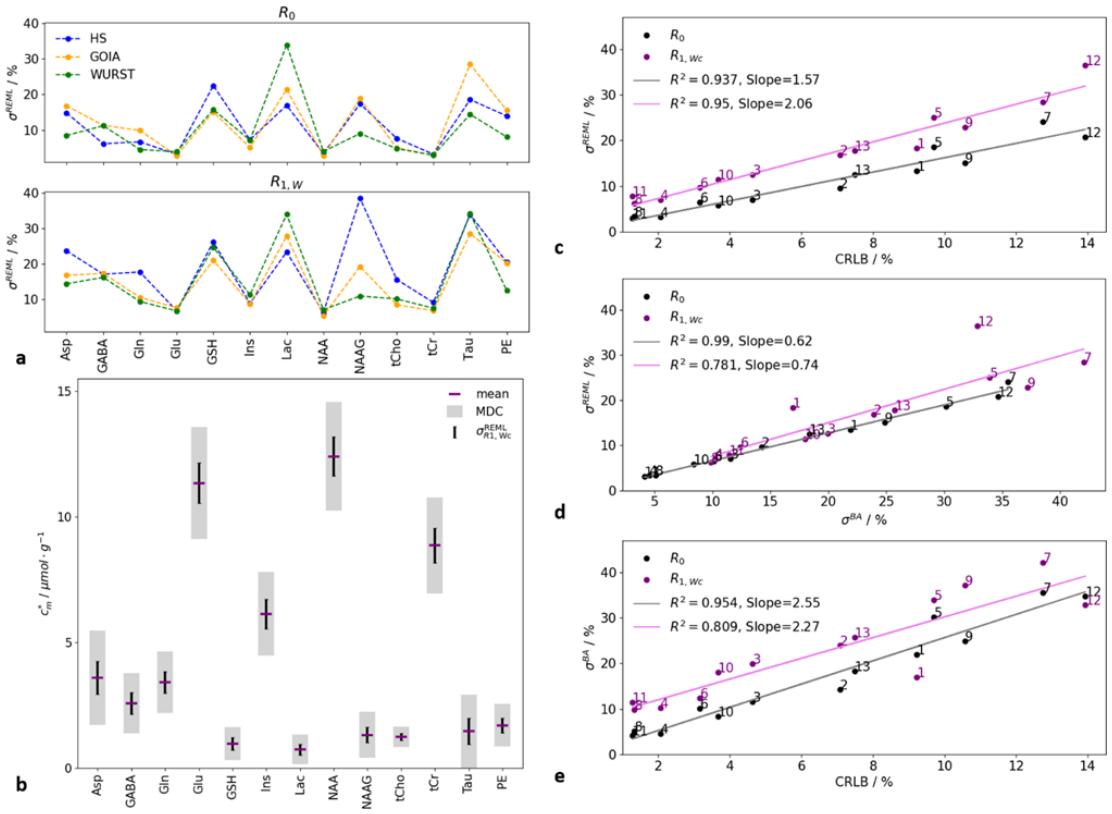

Figure 9 - Pulse-wise relative SDs (= coefficients of variances) averaged over the subjects, obtained by the restricted maximum likelihood estimation (REML) analysis, for repeatability (R_0, upper plot) and the combined reproducibility scenario (R_(1,Wc), lower plot) of all quantified metabolites. No pulse substantially outperforms another one, and hence, data of all three pulses were subsequently pooled to strengthen the statistical analysis regarding σ^REML and MDC. b) Mean concentrations (purple horizontal bars), ± 〖σ〗_(R1,Wc)^REML (black vertical bars), and ± minimal detectable changes (MDC, indicated with the gray box) of metabolites. Correlation plots between c) relative CRLBs and σ^REML, d) σ^BA and σ^REML, and e) relative CRLBs and σ^BA averaged over all three pulses and all volunteers. The R_0 scenario is denoted in black, while the R_(1,Wc) scenario is indicated in purple. Each point represents one metabolite: 1: Asp, 2: GABA, 3: Gln, 4: Glu, 5: GSH, 6: Ins, 7: Lac, 8: NAA, 9: NAAG, 10: tCho, 11: tCr, 12: Tau, 13: PE. Figure was published by Riemann et al.[12]

Enhanced MR techniques

One important component affecting MRS measurement uncertainty is the model that is used to generate quantitative data. Whilst it is known that the macromolecules in the investigated tissue can influence the results obtained for metabolite concentrations, the correct inclusion of macromolecules in analysis models has not been sufficiently investigated. NeuroMET2 addressed this issue by developing a model to better describe and include macromolecular signal contributions in data analysis.

A quantitative T1 mapping sequence was implemented and tested at the 7T scanner in phantoms and *in vivo*, and parameters were optimised. In total, 100 quantitative T1 maps were obtained from participants from the NeuroMET cohort. To exploit synergies, meetings with partners from the 18HLT05 QUIERO project took place, and a reconstruction pipeline from both project's data was developed in collaboration between partners from both projects.

A candidate improved macromolecular model for metabolite quantification was developed. Experiments on a 3T scanner showed very satisfactory results, and measurements at 7T showed both differences and similarities of the macro-molecular signals at 3T and 7T. Further work is needed beyond the lifetime of NeuroMET2, to validate the model by applying it to the reproducibility data set described above, and to evaluate its impact upon measurement uncertainty. After finalising the model, MRS data from the NeuroMET patient cohort needs to be re-analysed with the new model, to obtain more reliable values for the metabolite concentrations.

Longitudinal studies

The NeuroMET 2 project continued the work of 15HLT04 NeuroMET to characterise the patient cohort recruited. All *in vivo* MR measurements were performed at PTB using the aforementioned 7T MR system. Each *in vivo* MR measurement on the participants from NeuroMET cohort consisted of three major sections:

- structural or "morphometric" measurements, in order to assess structural information on each individual participant's brain structure and anatomy;
- MRS measurements, to measure concentrations of neurometabolites within the brain tissue of participants; and
- resting-state functional MRI (rs-fMRI), to extract information on functional connectivity within the participant's brains.

After analysis, the obtained results were sent to RISE and Modus for inclusion in PCOMs (including the NMM – see Objective 2) and in mathematical models of cognition (see Objective 4).

Visit	Total	HC	SCD	MCI	AD
T1	118	33	31	28	26
T2	24	1	17	4	2
T3	38	15	14	3	6
T4	26	14	8	1	3
T5	9	6	2	NA	1

Table 5 - MR spectroscopy measurements. In total, the following number of subject measurements were carried out at five different timepoints T1 – T5. T1 and T2 were within the 15HLT04 NeuroMET project, while T3-T5 were within NeuroMET2. Analysis of the obtained data will be ongoing beyond the end of the project.

Effect of the COVID-19 Pandemic

As noted in Objective 1, the COVID-19 pandemic was a limiting factor in the partners' ability to acquire MR data due to multiple suspensions in patient visits, and reduced throughput at other times due to hygienically motivated restrictions. However, mitigation strategies including scanning patients at weekends allowed the acquisition of fit-for-purpose datasets for this objective

Transfer to clinical setting

An agreement between MR vendor Siemens and two of the partners (PTB and Uni-Grief) was obtained as a prerequisite for transferring the MRS sequence files between the groups. The relevant MRS sequences were transferred from PTB to Uni-Grief and a 3T protocol was implemented at Uni-Grief. In a common effort of both groups, the sequence parameters were then optimised for the lower field strength.

Ethical Approval to perform additional MR measurements on a sub-cohort of the NeuroMET cohort in Uni-Grief was obtained. To ensure comparability, three subjects were examined with the 3T protocol at PTB's own 3T scanner and at the 3T scanner at Uni-Greif. After finalisation of the 3T measurement protocol, 22 subjects from the NeuroMET cohort underwent an additional MR exam at Uni-Greif.

Four one-on-one hands-on training courses on MRS data acquisition and MRS data analysis were given by PTB to MR physicists at Uni-Greif. Additionally, in May 2021 a NeuroMET virtual MR spectroscopy workshop was held, to transfer critical background knowledge on MRS protocols as well as the 3T NeuroMET protocol in particular to clinical users of MRS.

The first 7T system received FDA and CE approval in 2019, and 7T scanners are now starting to enter clinical routine environments, however there are still many more 3T systems being used for clinical examinations and research. A 3T system is not able to obtain the same detail and reduced measurement uncertainty as a 7T system, therefore the NeuroMET 3T protocol was designed to obtain as similar information as possible within a timeframe feasible for clinical research. Implementing this metrologically validated 3T protocol into a clinical setting is an important step for disseminating the project's results.

Conclusions

The project met Objective 3, to refine ultra-high field MRI and MRS protocols from 15HLT04 NeuroMET through longitudinal studies for application into clinics and develop new *in vivo* approaches.

The NeuroMET 2 project continued the work of 15HLT04 NeuroMET by characterising patients using the high accuracy of a 7T MR system. 7T scanners are now starting to enter clinical routine environments, however there are still many more 3T systems (than 7T systems) being used for clinical examinations and research. Currently, 3T systems are not able to obtain the same detail and reduced measurement uncertainty as a 7T system. Thus, implementing into a clinical setting 3T protocols which are mapped to metrologically validated 7T protocols is a very important step.

Significant advances were also made in fundamental metrological aspects of MRS, including data analysis and measurement uncertainty. A statistical framework was also developed to estimate the measurement uncertainty of *in vivo* MRS. The data set associated with this work including the raw data were published open access to allow their uptake in the examination of newly developed MRS modelling, postprocessing, and quantification pipelines and of their influence on the measurement uncertainty.

4.4 Objective 4: To advance biomarker measurements for early and accurate diagnosis through the validation and implementation into clinics of the 15HLT04 NeuroMET methods and other methods. Biomarkers such as Aβ1-42, Aβ 1-40, NfL, t-tau and α- synuclein will be monitored in the NeuroMET cohort and new RMP for NfL and p-tau will be developed. Methods for monitoring aggregation of NDD proteins will be also developed and validated to improve specificity of therapeutic targets and as potential diagnostic tools.

In terms of the analysis of fluid biomarkers for NDD, the NeuroMET2 project built upon the achievements of the 15HLT04 NeuroMET project, through longitudinal monitoring of key biomarkers, further development and validation of SI-traceable RMP, and development of assays for use in clinical settings.

Longitudinal monitoring of key biomarkers

The NeuroMET 2 project continued the work of 15HLT04 NeuroMET on characterising the NeuroMET cohort (Objective 1) for key NDD biomarkers. Where feasible, the obtained results were sent to RISE and Modus for inclusion in PCOMs (including the NMM – see Objective 2) and in mathematical models of cognition (see Objective 4).

Key AD biomarkers were measured in plasma by VUmc (Aβ 1-40, Aβ 1-42, p-tau, NfL, Glial fibrillary acidic protein (GFAP) and in CSF by Charité (Aβ 1-40, Aβ 1-42, t-tau, p-tau, NfL). The analysis included patients recruited under both 15HLT04 NeuroMET and NeuroMET2, and across multiple time points in order to provide longitudinal data in relation to disease progression. See Table 6

Visit	Total	HC	SCD	MCI	AD
t1	127	35	35	30	27
t2	32	4	20	3	5
t3	44	20	11	5	8
t4	31	16	8	3	4
t5	9	5	3	0	1

Table 6 – biomarker measurements carried out at VUmc. Each entry represents a suite of measurements (Aβ 1-40, Aβ 1-42, p-tau, NfL, GFAP) in plasma.

Development of RMPs for NDD protein biomarkers

The development of reference methods (RMPs) in in vitro diagnostics (IVD) is essential to enable harmonisation or standardisation of the results across locations and over time. It is widely recognised that standardisation of clinical measurement results is achievable by ensuring SI-traceability. While procedures are established to develop a traceability chain for small molecules, the development of SI-traceable RMPs for proteins is still in its infancy. This is due to difficulties with the definition of the measurand and the complexity of protein biomarkers. To ensure traceability to the SI of a RMP, the availability of a higher order primary calibrator is essential.

Candidate RMP for NfL

NfL is part of a complex forming a heteropolymer called Neurofilament (Nf). Nfs are found both in the central and peripheral nervous system and are composed of four subunits: Neurofilament-heavy (NfH), Neurofilament-medium (NfM), Neurofilament-light chains and alpha-internexin or peripherin. Nfs are abundant in neurons where they are essential for maintaining the axon structure. Following neuronal damages or neuronal degeneration, Nfs are released resulting in an increase of subunits concentration like NfL into CSF and blood. Thus, NfL is considered as a non-specific biomarker for neurodegeneration.

Selection and characterisation of primary calibrators

Three potential sources of NfL were chosen for the development of a primary calibrator: (i) Human Neurofilament light polypeptide unlabelled (Promise Proteomics, Grenoble), (ii) 68kDa Neurofilament Ag Bovine (MyBioSource Inc., San Diego, CA) and (iii) Neurofilament NF-L Full Length Recombinant Protein (Encor Biotechnology Inc., Gainesville, FA).

Preparations of these 3 sources of NfL were characterised using (a) Intact analysis of the protein using high resolution mass spectrometry (HRMS); (b) LC-MS/MS of resulting peptides following proteolytic digestion to confirm the sequence of and relative quantity of NfL in each preparation; and (c) Quantification to the SI via amino acid hydrolysis and Isotope Dilution Mass Spectrometry (IDMS).

Following characterisation, the Promise Proteomics material was found to be the most suitable primary calibrator as (i) no major chemical impurities were detected (see figure 10, below); (ii) the highest NfL content was observed based on analysis of the constituent peptides; and (iii) quantification via amino acid analysis (AAA) and IDMS demonstrated sufficient purity of the materials relative to the other source preparations.

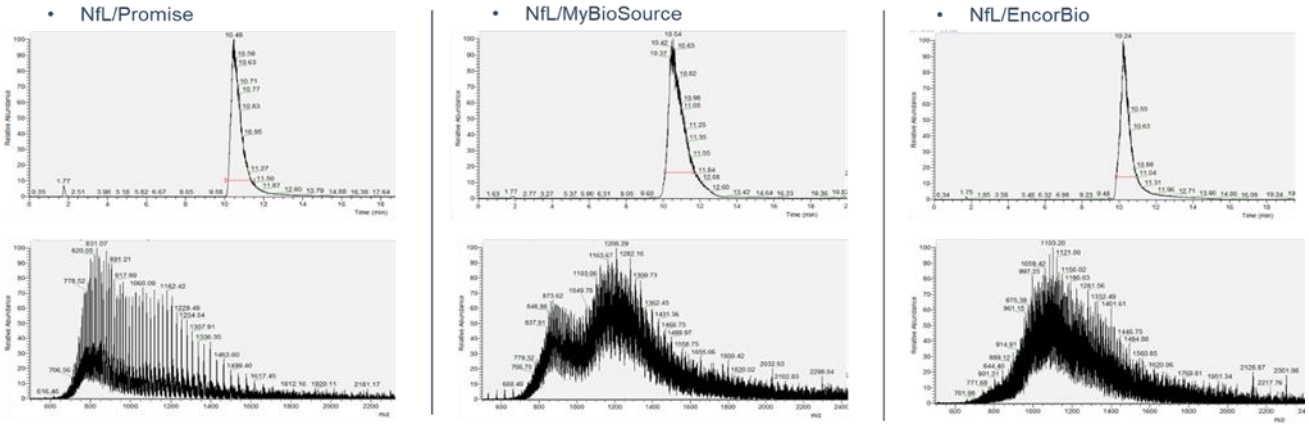

Figure 10 – Chromatogram and mass spectrum of three respective NfL preparations.

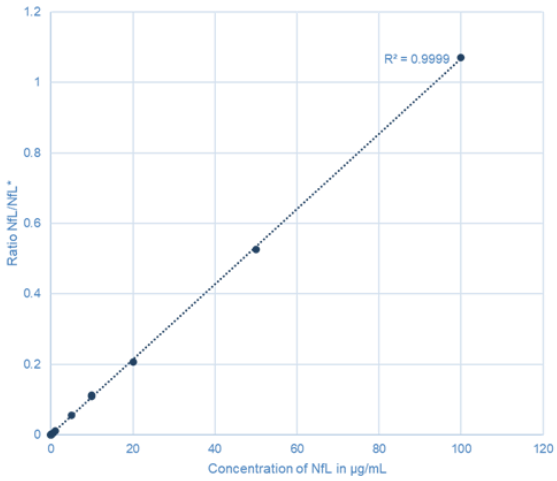

Figure 11 – Dilutional linearity of NfL/Promise.

Linearity and assessment of sensitivity

As quantification and characterisation of the primary calibrator was carried out at higher concentrations than the clinically relevant amounts, the potential loss of protein due to dilution steps was investigated. Several dilutions from 100 µg/mL to 1 ng/mL were prepared, digested in the presence of the labelled NfL, then injected on a C18 column and analysed on a Waters triple quadruple (QQQ) mass spectrometer coupled to a M-class liquid chromatography system using a Multiple Reaction Monitoring (MRM) method. The method included fifteen peptides located on the head, core, and tail domain of NfL. Natural/labelled ratios were calculated and good linearity was found in the indicated concentration ranges as shown in Figure 11.

Method development

Two MRM methods were set up to monitor NfL tryptic peptides (i) using QQQ, Xevo TQ-XS (Waters) and (ii) TQ8060nx (Shimadzu), mass spectrometers coupled to a capillary liquid chromatography (LC). Following this, two sample preparation workflows were developed and optimised: (i) including protein precipitation with methanol (see Figure 12) and peptide fractionation on an offline LC (Waters) after digestion and (ii) another immunocapture at protein level prior digestion.

The methods have not yet been fully validated to date due to the technical complexities arising during the method development coupled with the low concentration of NfL in CSF. However, antibody-free and immuno precipitation-based strategies have been developed that can meet the required sensitivity for the detection of NfL in clinical CSF samples.

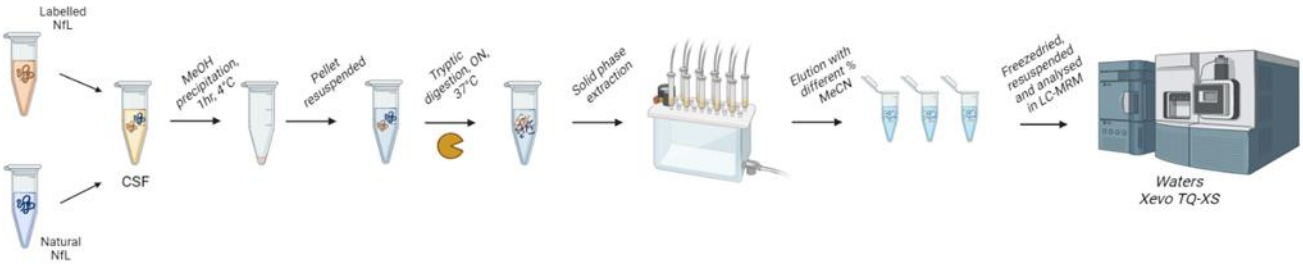

Figure 12 – Candidate assay workflow

Candidate RMP for p-tau

Tau hyper-phosphorylation is a significant indicator of AD. The detection and quantification of this modification occurring on serine and threonine amino acids provides higher accuracy to discriminate AD from other forms of dementia and to stratify patients at an early stage. Commercially available immunoassays are able to target specific phosphorylated epitopes: CSF p-tau181 is a well-established biomarker, routinely measured in clinical practice. Therefore, to improve the traceability of these measurements, a feasibility study on the development of a candidate reference method for p-tau181 in CSF was carried out as part of the NeuroMET2 project.

Characterisation of the candidate primary calibrator

A candidate primary calibrator consisting of a synthetic peptide was sourced and its purity fully characterised.

Impurity sequence	Delta Mass	RT	m/z	CS	MH+	Area	% to main component
TPPAPKpTPPSSGEPPPK	79.9684,97.0497	18,24	588,957	3	1763,8481	9052543,58	0,371555088
TPPAPKTPPSSGEPPK	21.9590 (sodium)	18	537,27	3	1608,7891	8439776,35	0,346404501
TPPAPPPKpTPPSSGEPPK	79.9684,97.0494	18,76	588,957	3	1763,8479	7465098,89	0,306399571
TPPAAPKpTPPSSGEPPK	79.9684,71.035ccccc3	18,56	580,285	3	1737,8337	5829091,35	0,239250828
TPPAPKpTPPSSSGEPPK	166.9941(79,9684+87,03)	17,62	585,615	3	1753,8242	5733545,65	0,235329224
TPPAPKTPPSSGEPPK -7	-7.0668	18,06	527,595	3	1579,7633	3837780,48	0,157518917
PPAPKpTPPSSGEPPK	Loss Threonine (101)	16,41	522,924	3	1565,7488	3220490,06	0,13218268

Table 7 - list of the impurities identified in the primary calibrator solution and used to correct the mass fraction obtained by AAA. Four impurities correspond to the phosphorylate 175-190 sequence with the insertion of an amino acid, one impurity corresponds to the unmodified 175-190 peptide, one impurity is the 175-190 with the deletion of the N-terminal threonine. The triply charged impurity at RT 18.06 having an m/z=527.595 presenting a delta mass of -7.07 Da compared to the unphosphorylated 175-190 sequence was not identified.

Method development

Calibration solutions were prepared and used to calibrate the candidate method for p-tau. Important work was done to optimise sample preparation, chromatographic conditions, and mass spectrometry conditions in order to reach an appropriate limit of quantification with sufficiently low measurement uncertainty. Method development and validation were conducted using 3 pools of frozen CSF prepared by CHU Mpt covering low, medium, and high tau concentrations.

A multiplex method was developed for measuring t-tau concentration via the peptide 156-163 (GAA peptide) and p-tau(181) via the peptide 175-190 simultaneously using an IDMS protocol. Calibration blends were prepared by using the t-tau protein primary calibrator and its relative labelled internal standard as well as the p-tau(181) peptide candidate primary calibrator and its corresponding labelled internal standard. This calibration approach was then evaluated for its feasibility for the simultaneous quantification of t-tau and p-tau(181) in three CSF pool samples.

Application of the method for value assignment of pooled samples

The developed method was applied to 3 pools of frozen CSF prepared by CHU Mpt covering low, medium, and high tau concentrations.

For p-tau quantification in the samples (QCs and CSF pools), peak area ratio R_{sample} was calculated and amount of substance ratio Q_{sample} was derived from a linear regression model. The concentration of p-tau in the samples $C_{p\text{-}tau}$ was then determined by taking into account the amount of p-tau solution added to the sample m_{spike} and the mass of sample m_s.

The standard measurement uncertainty associated with the concentration of p-tau in the samples (u_{sample}) was determined using the GUM approach[13] by taking into account uncertainties associated with the mass of sample $u(m_s)$, the amount of p-tau* solution added to the sample $u(m_{spike})$ and the uncertainty associated with calibration u(Q), which combines u(Qlin) and u(Qcal) according to the following equation:

$$u(Q) = \sqrt{\left(\frac{u(Q_{lin})}{Q_{lin}}\right)^2 + u(Q_{cal})^2}$$

Equation 2, where, u(Qlin) = standard uncertainty associated with the amount of substance ratio calculated through the calibration regression model (linearity of the calibration curve); u(Qcal) = standard uncertainty associated with the gravimetric preparation and value assignment of the calibration blends, which includes i) weighing of p-tau and p-tau and ii) determination of the p-tau concentration by AAA.*

Standard uncertainty of p-tau mass fraction $u(C_{p\text{-}tau})$ includes a precision component (u_{rep}), which corresponds to the standard deviation of the mass fraction values divided by the square root of the number of independent replicates.

$$u(C_{p-tau}) = \sqrt{{u_{sample}}^2 + {u_{rep}}^2}$$

Equation 3

Concentrations were determined in pmol/g and converted to pg/g by considering the molecular weight of 45717 Da (corresponding to Tau-441). This conversion allowed the comparison of the results by LC-IDMS obtained at LNE with the results by immunoassay, obtained at CHU Mpt using the Lumipulse® G1200 instrument from Fujirebio (Tokyo, Japan). The results are summarised in Table 8.

Pool	[p-tau] ± U (k=2) (pmol/g) LC-IDMS	[p-tau] ± U (k=2) (pg/g) LC-IDMS	U (%)	pTau (pg/mL) IA
Low	0.00598±0.00046	273±20.8	7.62%	26.3
Medium	0.00829±0.00073	379±33.5	8.85%	55.3
High	0.00985±0.00101	450±46.4	10,3%	86.2

Table 8 - concentration of p-tau(181) obtained for the CSF pools by using the LC-IDMS method and expressed in pmol/g and pg/g. The uncertainty is also reported. In the fourth column the values obtained by Immunoassay

The ID-LC-MS/MS method measures a concentration 10.4-fold higher for the pool low, 6.9-fold higher for the pool medium and 5.2-fold higher for the pool high. This is consisent with results from work on the development of the t-tau RMP undertaken during 15HLT04 NeuroMET. It is explained by a calibration bias and/or poor antibody affinity of antibodies used in IA. However, the use of SI-traceable material can compensate for this.

Clinical assays for α-synuclein

MS Clinical assay

Building upon work to develop a MS reference method for the quantification of α-synuclein in CSF carried at LGC in the 15HLT04 NeuroMET project, CHU Mpt developed and optimised an MS clinical assay for the quantification of this key NDD biomarker in both CSF and plasma.

A method using a Shimadzu LC (mikros) and a triple quadrupole - Shimadzu 8060 mass spectrometer (Duisburg, Germany) in positive ionisation mode was developed and validated using CSF QC materials (see Table 9), below.

Protein names	Sequence	Intra-days precision (%)	Inter-days precision (%)	LOQ (ng/ml)	Carry over (%)	Stability in the autosampler (at 4°C) during 0h /24h / 36h / 48h / 72h	Stability when thaw on ice during 0h / 2h / 4h	Stability when thaw at room temperature during 0h / 2h / 4h
α/β-syn	EGVVAAAEK	8,4	18,6	<0,034	0,034	100/103/111/112/106	100/127/124	100/97/98
	EGVLYVGSK	9,9	13,5	0,168	0,000	100/109/106/110/109	100/108/120	100/84/103
β/γ-syn	QGVTEAAEK	8,7	12,9	0,168	0,067	100/121/108/119/100	100/104/101	100/125/92
α-syn	QGVAEAAGK	7,9	15,9	0,056	0,338	100/67/124/124/115	100/80/94	100/87/128
	EGVVHGVATVAEK	9,2	18,8	0,084	0,031	100/75/61/77/96	100/108/182	100/135/185
	EQVTNVGGAVVTGVTAVAQK	9,8	12,1	0,167	0,260	100/105/100/90/120	100/454/124	100/92/96
	TVEGAGSIAAATGFVK	8,8	23,3	NA	NA	NA	100/128/145	100/94/108
β-syn	EGVVQGVASVAEK	5,6	8,0	0,167	0,004	100/92/107/78/67	100/80/80	100/120/153
	EQASHLGGAVFSGAGNIAAATGLVK	NA	NA	NA	NA	NA	NA	NA
γ-syn	EGVVGAVEK	12,0	13,4	0,056	0,028	100/94/92/86/81	100/93/90	100/111/83
	ENVVQSVTSVAEK	10,3	22,3	0,111	0	100/106/131/98/111	100/147/130	100/95/74
	EQANAVSEAVVSSVNTVATK	NA	NA	NA	0	NA	NA	NA
	TVEEAENIAVTSGVVR	8,1	13,1	0,084	0,075	100/110/127/104/112	100/105/112	100/101/73

Table 9: Performance of the method in CSF QCs, such as intra- and inter-day precision, , LOQ: limit of quantification, carry over, and sample stability parameters such as: in order from left to right, accuracy of storage in the autosampler at 4°C for 0h, 24h, 36h, 48h, and 72h, thawing of samples on ice and at room temperature for 0h 2h, 4h (n=3).

Protein name	Peptide sequence	Intra-assays precision (%)			Inter-assays precision (%)			Mean Accuracy at 12h / 24h / 36h / 48h in the autosampler at 4°C, respectively			LOQ (ng/ml)	Linearity range (ng/ml)
		Low QC	Medium QC	High QC	Low QC	Medium QC	High QC	Low QC	Medium QC	High QC		
α/ß-syn common peptides	EGVVAAAEK	7	4	6	15	16	14	99 / 98 / 102 / 103	104 / 102 / 105 / 101	102 / 102 / 101 / 104	12,90	0 - 5981
	EGVLYVGSK	8	5	2	17	13	12	94 / 93 / 100 / 103	102 / 97 / 98 / 96	96 / 104 / 108 / 104	19,35	0 - 5981
α-syn	QGVAEAAGK	5	4	3	19	15	13	101 / 102 / 95 / 100	87 / 93 / 82 /80	113 / 113 / 113 / 104	6,45	0 - 5571
	EQVTNVGGAVVTGVTAVAQK	5	3	4	12	13	12	121 / 104 / 116 / 113	108 / 105 / 107 / 111	97 / 115 / 104 / 99	19,35	0 - 5571
	TVEGAGSIAAATGFVK	13	8	5	22	15	10	73 / 89 / 66 / 56	70 / 80 / 108 / 90	110 / 109 / 97 / 106	19,35	0 - 5571

Table 10: Method performance for plasma samples described as intra and inter assay precision for each level (high, medium and low), accuracy of the storage in the autosampler at 4°C during 12h, 24h, 36h and 48h, LOQ: limit of quantification and linearity range

Application to clinical cohort

The developed method was applied to 143 biobank samples, (n=82) for Parkinson Disease (PD) group, (n=8) for Multiple System Atrophy (MSA) group, (n=32) for Lewy Body Dementia (LBD) group and (n=21) for control group to assess the clinical performance results for alpha syn peptides (Table 11, below).

Peptide sequence	Position on α-syn sequence	Clinical performance: p value; sensitivity and specificity (AUC - area under ROC curve)		
		Distinguish PD vs LBD	Distinguish PD vs MSA	Distinguish PD vs control
α/β-syn EGVVAAAEK	13-21	0.3442	0.9887	0.0277; 0.83 and 0.67 (0.656)
α-syn QGVAEAAGK	34-32	0.0331; 0.49 and 0.78 (0.629)	0.9887	0.0014; 0.83 and 0.76 (0.726)
α/β-syn EGVLYVGSK	35-43	0.0014; 0.7 and 0.69 (0.694)	0.5706	0.0001; 0.84 and 0.76 (0.787)
α-syn EQVTNVGGAVVTGVTAVAQK	61-80	0.7098	0.3717	0.0205; 0.87 and 0.67 (0.664)
α-syn TVEGAGSIAAATGFVK	81-96	0.341	0.4355	0.2134

Table 11: Summary of clinical results for each peptide from α-syn and α/β-syn peptides

The results of this first clinical validation experiment on patient plasma samples showed that synuclein peptide concentrations were significantly different among synucleinopathies. Further validation needs to be performed (after the end of this project) to confirm this result and evaluate its clinical utility.

Orthogonal methods

Results were obtained using an LC-MS method for the quantification of α-synuclein in CSF that was developed by LGC under the 15HLT04 NeuroMET project. The method was validated as part of this project, NeuroMET2. Peptides monitored by LGC in the method consist of T6, T8, T12 and T13 (Figure 13).

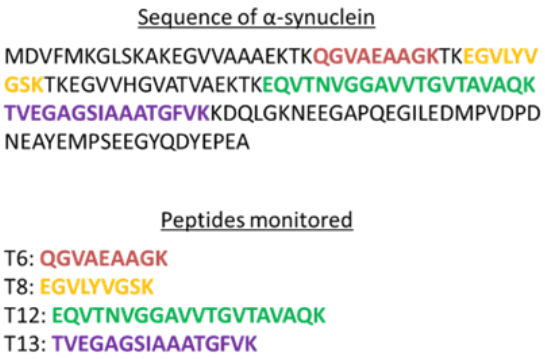

Figure 13. The sequence of alpha-synuclein and its peptides monitored for quantification

The method was verified and showed a linearity range of 0.1 ng/g to 10 ng/g with an r^2 value of 0.999 for three peptides (T8, T12 and T13). The intra-assay precision was ~ 2 %. The limit of quantification (LOQ) ranged from 10 – 100 pg/g an the intra-assay precision range was 2 – 17 %.

15 CSF samples from the NeuroMET cohort (Objective 1) were analysed using this targeted LC-MS method. The concentrations of peptides T12 and T13 were found to be lower compared with immunoassay results from partner CHU Mpt. LC-MS values compared with the immunoassay data, estimated higher levels for peptide T8 which is present both in α and β-synuclein.

VUmc also developed an RT-QuIC assay for α-synuclein measurement in order to compare the results of the LC-MS method. Two recombinant monomers of the full length wild-type α-synuclein, one untagged and one with a N-terminal His-tag were tested. The RT-QuIC assay protocol, using the his-tagged monomer, was able to provide a good signal after on average 75 hours in positive control samples. To verify the RT-QuIC assay performance for α-synuclein, samples that were previously shown to be positive on the RT-QuIC assay in a different laboratory were tested on the VUmc assay and gave a positive signal.

In addition, VUmc worked on the purification of recombinant α-synuclein from *E.coli* cells and different purification protocols (from other NeuroMET2 partners) were tested. However, the testing of clinical samples could not be fully completed within the project's lifetime.

Conclusions

The project met Objective 4, to advance biomarker measurements for early and accurate diagnosis through validation and implementation into clinics of the 15HLT04 NeuroMET methods and other methods.

The The NeuroMET 2 project continued the work of 15HLT04 NeuroMET by characterising the patient NeuroMET cohort (Objective 1) for key NDD biomarkers and to input this data into PCOMs (including the NMM, Objective 2) and in mathematical models of cognition (Objective 4).

Key AD biomarkers were measured in plasma (Aβ 1-40, Aβ 1-42, p-tau, NfL, GFAP) and in CSF (Aβ 1-40, Aβ 1-42, t-tau, p-tau, NfL). The analysis included patients recruited under both 15HLT04 NeuroMET and NeuroMET2, and across multiple time points in order to provide longitudinal data in relation to disease progression.

In parallel, a number of SI-traceable RMPs were developed and validated for key NDD protein biomarkers. This was achieved by characterisation and quantification of primary calibrators, followed by the development of mass spectrometry method(s) for quantification of the protein of interest.

- A LC-MS method for p-tau in CSF was developed and validated using pooled samples and was compared to immunoassay measurements in order to understand differences between the RMP and methods commonly used in clinical settings.
- A candidate LC-MS method for monitoring NfL in CSF was developed, encompassing novel antibody-free and immuno precipitation-based strategies which are required in order to reach the required sensitivity for the detection of NfL in clinical CSF samples.
- A reference method for α-synuclein in CSF developed in the 15HLT04 NeuroMET project was validated during and used for comparison with orthogonal techniques including a clinical MS method developed during this project and an RT-QuIC assay.

4.5 Objective 5: To enhance Causal Rasch mathematical models to define prototype metrological references for cognition expressed as CSE. This will provide an extensive explanation of how able a human can act as an "instrument" when measuring the difficulty of a task such as a cognitive test. Those models will be applied to the PCOMs, MRI and MRS, biomarker data to define and improve the prognostic values of the methods developed.

It is critical to have reliable and valid measures of both symptoms (e.g. memory ability) and signs (e.g. biomarkers and brain volumes), in order to determine diagnoses, monitor treatments, evaluate drugs, and better understand disease progression. This project NeuroMET2 and the preceding 15HLT04 NeuroMET investigated how far the conventional aspects of metrology for organic and physical measurements can be extended into the area of cognition.

The 15HLT04 NeuroMET project demonstrated how the application of RMT and a measurement system where the human acts as measurement instrument to cognitive test data could address the above issue and led to the development of mathematical models of cognition which explain and improve memory tests.[14] [15]

Applying metrological concepts to memory measurements

Memory measurements are, as any other human-based measurements, characterised by two coupling attributes to be considered: (i) task difficulty, and (ii) person ability. Clinicians are typically interested in the patient's memory ability when making or monitoring diagnoses, monitoring treatments, evaluating drugs and better understanding disease progression and treatments. However, for researchers, metrologists, psychometricians etc., both coupling attributes are of importance for providing clinicians with reliable and valid measures of person memory ability.

To date, the most commonly used legacy cognitive tests (e.g., MMSE and Alzheimer's Disease Assessment Scale-Cognitive Behaviour) do not have sufficient accuracy to distinguish between patients (especially in early stage disease) and lack metrological legitimation.[2] [5] Specifically, person-to-item targeting is often poor, owing to skewed distributions of both task difficulty across the test items and person ability across the cohort, which leads to large measurement uncertainties, particularly for those with early memory decline.[16] [17]

Traditional psychometric methods (i.e., based on classical test theory) do not account for the ordinal nature of data generated by human responses and lack a separation between person and item attributes.[2] [5] [18] [19] However, using RMT, the person's response, $P_{success}$, can be restituted into separate measures for memory task difficulty (δ) and person memory ability (θ) on a conjoint interval scale.[7] [20] [21] This addresses both ordinality and separability between item and person attributes. In turn, this takes the first steps towards metrological legitimation where the memory task difficulty, δ, can serve as metrological references to enable comparability of person memory ability, θ.

Construct Specification Equations (CSE)

A CSE describes a mathematical relationship for the 'something' that causes variation in the attribute of interest.

$$Z = \sum_{k} \beta_k \cdot X_k$$

Equation 4

RISE and Modus have developed (i) CSEs to describe task difficulty in cognitive tests (both verbal and non- verbal), (ii) CSEs which provide a composite metric of memory task difficulty, δ_j, (see the NMM in Objective 2, above), and (iii) person memory ability, θ_i.

RISE and Modus have also developed CSEs to describe both task difficulty, δ_j, and person ability, θ_i, in cognitive tests (both verbal and non-verbal). The entropy-based CSEs for task difficulty, δ_j, provide a specific and metrological unique guidance in the composition of the NMM (Objective 2).

Entropy based CSEs for memory task difficulty, δ_j

These NeuroMET studies focused on memory tests for the recall of forward sequences of blocks or digits[5 7 22] and freely recalling words.[5 23 24] Such recall tests are typically regarded as assessments of short-term memory.

Using Information theory[25 26], CSEs were developed to successfully explain the difficulty, δ_j, (see Equation 1, Objective 2, above) of recalling non-verbal sequences in memory tests (i.e., CBT and DST).[5] CSEs with entropy-based variables dominating, were found experimentally to explain and predict the construct of task difficulty for short-term memory tests:

$$zR_{j,CBT} = -6(3) + 1.2(6) \times Entropy_j - 0.3(1.1) \times Reversals_j - 0.02(0.25) \times AveDistance_j$$

$$zR_{j,DST} = -6(3) + 1.0(2) \times Entropy_j + 0.01(36) \times Reversals_j - 0.2(1.4) \times AveDistance_j$$

Equation 5

The information theory approach was also successfully applied to explanations of task difficulty δ_j in word list recall tests.[15 29] Specifically, for the 15 words in the RAVLT first trial (immediate recall), task difficulty, including serial position effects, is adequately explained by the CSE:

$$zR_{j,RAVLT\_A} = 5(3) + 0.7(5) \times Primacy_j + 0.8(5) \times Recency_j + 0.25(20) \times Frequency_j$$

Equation 6

For a shorter word list, such as for the 10 words in the word list test in the CERAD test battery, less influence from SPE is evident [29]:

$$zR_{j,WLL\_CERAD} = 5(5) + 0.7(5) \times Primacy_j + 1(1) \times Recency_j - 0.13(8) \times Frequency_j$$

Equation 7

This work proposes an entropy-based equivalence criterion, whereby different tasks (in the form of items) from different tests can be combined, enabling new memory tests to be formed by choosing a bespoke selection of items, thus leading to more efficient testing, improved reliability (reduced uncertainties for person ability) and validity (in the attributes themselves, here memory task difficulty).[5]

The CSE developed play a number of significant roles, particularly providing support for validity in the attributes themselves (memory task difficulty), and in guiding the composition of novel memory metrics (the NMM, Objective 2). Using CSEs for memory task difficulty for individual legacy tests, RISE and Modus have developed explanatory models with equal entropy that indicate the equivalence of different items to be combined into the NMM.[27 29]

The NMM, was obtained by making a bespoke selection of memory tasks from a battery of legacy tests, and provides a superior, metrologically quality assured scale for task difficulty to be used when measuring person memory ability.[5 28 29] In turn, the new metric is found to enable improved accuracy and the best degree of correlation with biomarkers seen so far.

CSEs for person memory ability and correlation of the NMM with biomarker studies

The NeuroMET2 project has demonstrated that person memory ability can be explained in terms of causal models by bringing together results from multidisciplinary measurements (see Objectives 3 & 4). Such causal models are so called CSEs in which multimodal statistical approaches are used in order to explain the construct person memory ability θ_i (see eq. 1, Objective 2, above) as a function of a set of explanatory (independent) variables, X_k.

The variables considered were (i) brain volumes measured by MRI, (ii) metabolites measured by MRS, and (iii) blood-based biomarkers. Changes in biomarkers can cause change in person memory ability and thus CSEs can provide a better understanding of what is causing variation in person memory ability.

Explaining person memory ability based on a combination of explanatory variables

When assessing how well a total set of 16 explanatory variables (age, MRI, MRS and blood-based biomarkers) could predict the empirical measured person ability, a multivariate regression of person ability and the 16 explanatory variables was made and a Pearson correlation coefficient of 0.77 was found.

As shown in Table 12 (below), several of the explanatory variables had β-coefficients suffering from large measurement uncertainties. Thus, we explored how an optimal combination of explanatory variables could be conducted. This resulted in seven explanatory variables having the same Pearson correlation coefficient for all 16 explanatory variables but a significant improvement in measurement uncertainties in several of the β-coefficients, and no deterioration in the other biomarker coefficient uncertainties.

Figure 14 (below) illustrates the contribution of the explanatory variables to person ability, θ_i, (eq. 1) from the CSE equation (Equation 8, below).

$$zR_{i,7} = -7(3) + 2(1) \times Thickness_i + 1.7(5) \times Hippocampus_{i,left} - 0.0045(8) \times GFAP_i \\ -0.03(10) \times pTau181_i - 0.127(2) \times Ins_i - 0.8(4) \times Amygdala_{i,right} + 0.035(3) \times Age_i$$

Equation 8

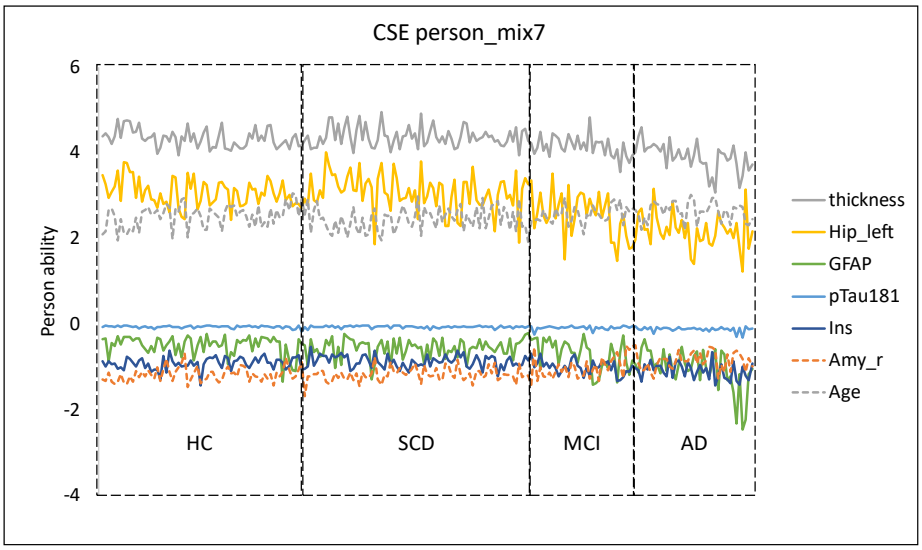

Figure 14 - Individual contributions to person ability (y-axis) from an optimal selection of 7 NeuroMET biomarkers (MRI, MRS and plasma-based). Measurements are ordered by decreasing person ability across the NeuroMET cohort (left to right along the x-axis) and approximate stratification into the sample groups (HC, SCD, MCI, AD) is shown in the figure. Abbreviations: AD = Alzheimer's disease, Amy_l = left amygdala GFAP = glial fibrillary acidic protein, HC = healthy control, Hip_l = left hippocampus, MCI = mild cognitive impairment, NfL = Neurofilament light, SCD = subjective cognitive decline.

There is a clear added value in using a multivariate approach compared to assessing univariate correlations. Table 12 presents contributions to person ability ($\beta k \cdot Xk$) and associated measurement uncertainties for the selection of the seven explanatory variables for univariate, grouped according to type of explanatory variable when performing multivariate Principal Component Regression with 16 explanatory variables and Principal Component Regression with 7 explanatory variables.

The same methodology is equally applicable when explaining person memory ability, especially in terms of ensuring the validity of the relevant attribute itself (person memory ability) and in enabling explanations of person memory ability as a function of the set of independent variables on which the ability depends.

Independent validation of RISE's approaches to explain memory ability as a function of the set of independent variables was carried out by LGC and Modus.

	Univariate	Grouped with same kind of explanatory variables	Combination of 16 explanatory variables	Combination of 7 explanatory variables (Eq. 8)
Pearson correlation coefficient				
MRI		0.65		
Blood-based biomarkers		0.58		
Combinations			0.77	0.77
Cortical Thickness	4.5 (8)	3 (1)	2 (1)	2 (1)
Hippocampus left	2.1 (4)	0.8 (5)	1.3 (1.7)	1.7 (5)
GFAP	-0.008 (2)	-0.0059 (8)	-0.005 (1)	-0.0045 (8)
pTau181	-0.4 (1)	-0.2 (1)	-0.1 (1)	-0.03 (10)
Myoinositol	-0.34 (10)		-0.14(44)	-0.127 (2)
Amygdala right	1.7 (5)		-1.0 (2)	-0.8 (4)
Age	-0.04 (2)		0.03(1)	0.035 (3)

Table 12 - Summary of Pearson correlation coefficients when correlating person memory ability with different explanatory variables, Xk, as well as the contributions to person ability ($\beta k \cdot Xk$)and associated measurement uncertainties (K = 2).

Measurement Uncertainties

The NMM (Objective 2) comprises several legacy tests that provide more information about the subjects being tested than the individual legacy tests and has been shown to reduce measurement uncertainties for person memory ability by up to a five-fold reduction.[8] [29]

However, the CSEs developed by RISE are based on legacy tests (e.g. CBT etc alone) which give larger measurement uncertainties than the measurements of person memory ability that when the NMM is being used. Consequently, as measurement uncertainties propagate through the CSE, (i) the predictive power reduces when explaining person memory ability; (ii) the Pearson correlation coefficient decreases from 0.77 to 0.46 and (iii) the measurement uncertainties in the β-coefficients become larger[5] (see Table 13, below).

Furthermore, the sample size also affects the measurement uncertainties in the CSE. As shown in Table 13, two random subsamples (subsample 1 n=105 and subsample 2 n=108, respectively) resulted in two CSEs which, within the measurement uncertainties, corroborated the CSE in '**CSEs for person memory ability'**, above'.

Importantly, the measurement uncertainties in the β-coefficients are up to twice as large for the subsamples compared to the full sample, due to the smaller sample sizes. At the same time, the Pearson correlation coefficients were 0.78 and 0.77, respectively.

θ_i	Combination of 7 explanatory variables (Eq. 8)	CBT	Subsample 1	Subsample 2
Pearson correlation coefficient	0.77	0.46	0.78	0.77
Thickness	2 (1)	-0.4 (2.2)	2.2 (1.5)	1.4 (1.6)
Hippocampus left	1.7 (5)	1.5 (2.1)	1.6 (1.0)	1.7 (0.6)
GFAP	-0.0045 (8)	-0.008 (2)	-0.005 (2)	-0.003 (1)
pTau181	-0.03 (10)	0.1(2)	0.03 (11)	-0.16 (17)
Myoinositol	-0.127 (2)	-0.3 (2)	-0.13 (8)	-0.14 (5)
Amygdala right	-0.8 (4)	-0.9 (5)	-0.8 (8)	-0.5 (4)
Age		0.007 (20)	0.04 (4)	0.024 (7)

Table 13. Summary of Pearson correlation coefficients from different subsets and derived from CBT when correlating person memory ability, θ_i, with the same set of explanatory variables, X_k, as well as the contributions to person ability ($\beta_k \cdot X_k$) and associated measurement uncertainties (coverage factor K = 2).

Given the relatively large measurement uncertainties of not only the NMM (Objective 2) but also each biomarker (Objective 4), it is important to give a critical figure of merit metrologically as an indication of the responsiveness of the various explanatory variables in the CSE. In turn, this will provide end-users with estimates of the current responsiveness of the tools being used for measuring person memory ability (and memory task difficulty). Hence allowing the determination of values for Smallest Detectable Change (SCD) in person memory ability.

Based on the CSE for person memory ability, the SCD for person abilities (see Table 14, below) are found in several cases to be greater than the value of their associated explanatory variables. For instance, a decrease in hippocampus volume of 0.62 + 0.08 mm^3 or aging with 30 + 6 years are needed to produce a measurable change in person memory ability. Thus, the responsiveness of the various explanatory variables can barely be detected, i.e., as the instrument is not sufficiently responsive.

θ_i	SDC (U)	Mean	Min	Max
Thickness	5 (2)	2.03	1.47	2.37
Hippocampus left	0.62(8)	1.65	0.72	2.35
GFAP	237(24)	145.46	45.35	545.59
pTau181	30 (60)	2.42	0.89	9.74
Myoinositol	8.0 (1.6)	1.45	0.66	2.22
Amygdala right	1.4 (4)	7.45	4.15	11.29
Age	30 (6)	72.06	55.00	87.00

Table 14. SDC in person memory ability due to changes in explanatory variables from the CSE for person ability (section 3.3) as well as empirical mean, min and max for explanatory variables in this work.

CSEs as RMPs for memory measurements

The work by RISE in the NeuroMET2 project proposes that CSEs for task difficulty can constitute metrological references analogous to 'recipes' for a RMP in chemical and materials metrology which enable instrument calibrations e.g., for reliable and traceable measurements of person memory ability.[27] When the human is placed at the heart of the system as the instrument,[30] the measurement object has no input but only produces an output (which acts as stimulus input to the instrument). Measurement objects can be robust and simple and as such are more suitable as metrological references and RMPs compared with the relatively sensitive and complex instrument, which is also more prone to the environment, context, and choice of method.[31] [32] Thus, the measurement object (e.g., the task) is a natural first choice of metrological standard.

A further advancement when using CSEs for task difficulty as metrological references is the ability to link items when transferring traceability from one test to another.[29] Transferability of traceability (within quoted uncertainties) is ensured by the entropy-based theory models above and the CSE specifies individual item difficulties so that new items can be designed ad hoc to have a certain task difficulty in terms of the structure of each test sequence.

The entropy-based models for task difficulty can also be used as a means of crosswalking between different cultural realisations. In the NeuroMET project, CSEs for the RAVLT have been compared and contrasted across different languages (German and Swedish), which has demonstrated satisfactory comparability between the cohorts.[23] However, while inter-laboratory studies are well-established as a means of evaluating measurement accuracy and ensuring metrological traceability, in chemical and materials metrology, in the human and social sciences such routines are less developed and need further work.[23] [29]

Traceability for memory measurements

Together with the arguments for metrological references and traceability, comes the consideration of a traceability pyramid for memory measurements. Figure 15, below, presents a proposed traceability pyramid for memory measurements based on ISO 17511:2003.[29]

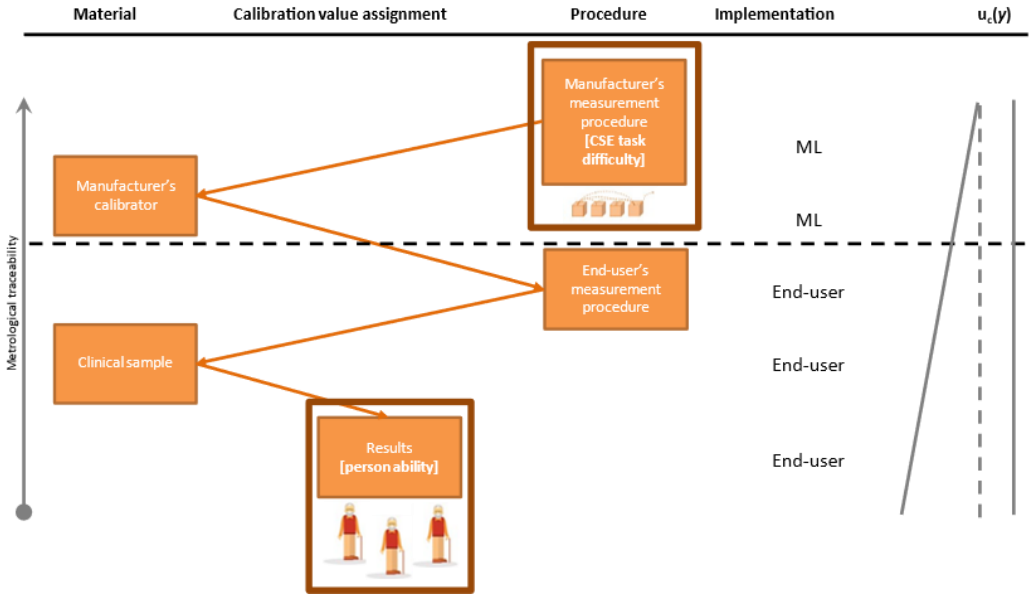

Figure 15. A proposed traceability pyramid for memory measurements based on ISO 17511:2003. To the left there is an arrow pointing upwards, this implies that the metrological traceability is increasing higher up the pyramid, while on the left you see how measurement uncertainties decrease higher up the pyramid. A dotted line has been included where measurement uncertainties are stable, which might be the case for memory measurements as the procedures may not vary between our study cohort and clinical practice, but this needs further evaluation.

As yet, there is no SI reference standard at the top of the pyramid; however, at this stage we have tentatively placed the reference procedure at the top of the pyramid, to include both persons and task.

The NeuroMET CSEs, developed for a specific cohort and set of legacy tests, are currently the most route towards development of a calibrator. However, it should be noted that these are initial results, based on the NeuroMET cohort alone, and will requires extensions to include other test persons in order to make the reference procedure as representative as possible. Potentially, in the future, data from cohorts across the world might be combined into a reference sample and the production of a global CSE in much the same way as traditional references are tested in inter-laboratory comparisons. A first step in this direction was taken by this project when the reproducibility of the task difficulty CSE between the NeuroMET cohort and the cohort from the Gothenburg Memory Clinic was assessed by RISE and HKR. [23]

Conclusions

The project met Objective 5, to enhance Causal Rasch mathematical models to define prototype metrological references for cognition expressed as “CSE”s. This provided an extensive explanation of how a human can act as an “instrument” when measuring the difficulty of a task such as a cognitive test. Those models were applied to the PCOMs, MRI and MRS, and biomarker data to define and improve the prognostic values of the methods developed.

Building upon the results of 15HLT04 NeuroMET, work was undertaken to enhance the understanding of the role of entropy when formulating CSEs for memory task difficulty, leading in turn to the development of the NMM (Objective 2). While existing individual legacy neuropsychological tests have lacked both sufficient accuracies to distinguish disease stages and have not been quality-assured, the new NMM is based on modern measurement theory, metrological quality assurance, causal multivariate analyses and cross-walking between items carefully chosen from different legacy tests when composing the metric.

The new NMM shows an up to five-fold reduction in uncertainties for measurements of memory ability along the AD continuum without jeopardising validity which is a significant improvement. The new NMM should also be more efficient and specifically suitable for studies of, for example, early detection of disease onset or the effects of drug intervention.

The project also demonstrated that person memory ability can be explained in terms of causal models by bringing together results multidisciplinary measurements. Such causal models are CSEs, in which multimodal

statistical approaches are used in order to explain the construct person memory ability as a function of a set of explanatory (independent) variables.

Finally the project developed a traceability pyramid for memory measurements based on ISO 17511:2003.30. This is a novel approach which applies established metrological principles which are familiar in physics and chemistry into psychometric testing.

4.6 Objective 6: To transfer the project's results to the measurement supply chain, standards developing organisations (ISO/TC212, IFCC, and JCTLM), manufacturers (MR manufacturers, immunoassay and MS manufacturers), and end users (e.g., clinical laboratories and pharma) and promote the NeuroMET multidisciplinary infrastructure to become the ideal space for NDD translational research.

The consortia delivering the NeuroMET2 and preceding 15HLT04 NeuroMET projects have always sought to maintain close links with standards organisations and end users in order to translate research outputs into tangible outcomes for the NDD research community. A number of initiatives were implemented during NeuroMET2 including interlaboratory value assignment and commutability studies in conjunction with the IFCC, submission of reference methods to the JCTLM database, and engagement with clinical end users (see the impact section).

Collaboration between NeuroMET2 and the IFCC

The IFCC, and in particular its Working Group 7 on CSF proteins was identified as a key collaborator for NeuroMET as it facilitates the development of certified reference materials (CRMs) and the uptake of methodologies and guidelines on a global scale, with the potential of technology transfer to instrument manufacturers.

Round Robin Study for t-tau measurement standardisation

The improvement of CSF biomarker measurement comparability needs the development of RMPs to calibrate clinical assays. During 15HLT04 NeuroMET, a candidate RMP for SI-traceable quantification of t-tau protein in CSF was developed[33], and during NeuroMET2 this RMP was used in an interlaboratory study in conjunction with the IFCC WG on CSF proteins.

The method comprises an IDMS protocol for the quantification of the peptide GAAPPGQK (GAA-peptide) of the t-tau protein by LC-MS/MS. Traceability to the SI is achieved by using a primary calibrator consisting of a recombinant Tau protein, the purity of which was extensively characterised by LNE, LGC and CHU Mpt. This primary calibrator was then made available to the participants of the round robin study, which were three laboratories/members of the IFCC-WG-CSF: (1) the University of Gothenburg (GOT), (2) the University of Pennsylvania (UPENN), and (3) LNE. These 3 laboratories have existing methods for the absolute quantification of t-tau in CSF. CSF samples of unknown concentration were distributed among the participants of the round robin study as well as samples of artificial CSF spiked with different amount of t-tau LNE's primary calibrator.

The aim of this round robin study (interlaboratory comparison) was to evaluate whether the 3 different laboratories could provide equivalent results by using LC-MS methods and a bottom-up approach, based on tryptic digestion of the endogenous protein to measure the obtained peptides, and in particular the peptide 156-163, which is a non-modified peptide. LNE and UPENN used recombinant protein calibration materials (LNE's primary calibrator was SI-traceably quantified by AAA), whereas GOT used a particular peptide material, the wing-peptide, having the 156-163 sequence surrounded by two wing sequences at the N- and C-term (this material was SI-traceable quantified by AAA).

Calibration standards were prepared gravimetrically at LNE by spiking a fixed amount of 15N-recombinant protein or peptide to variable amount of 14N-recombinant tau in an artificial CSF (aCSF) matrix, consisting of diluted human serum. The same amount of 15N-labelled protein was spiked into the CSF samples that were subjected to the same protocol as the calibration standards: i.e. perchloric acid protein precipitation and supernatant recovery, a hydrophilic-lipophilic balance solid phase extraction (SPE), and finally trypsin digestion.

LC-MS/MS experiments were carried out by using liquid chromatography coupled to orbitrap HRMS. MS detection was operated in parallel reaction monitoring (PRM) mode, following 12 peptides and their labelled counterparts to identify the presence of Tau protein.

Only the GAA peptide at m/z 363.2 and its labelled form at m/z 367.2 were used for protein quantification, the 3 most intense transitions per peptide were used as quantifiers. The ratio between the GAA peptide and its corresponding labelled form was obtained by considering the area ratios. Further to the CSF samples, different

solutions of t-tau nominal concentrations (based on AAA values) were gravimetrically prepared by spiking t- tau recombinant protein in aCSF and were measured by the RMP by spiking the labelled internal standard. The mean bias among the obtained concentrations and the theoretical concentrations were calculated. The results are shown in Table 15, below.

LNE Low			
Date	C(pg/mL)	C_{theo} (pg/mL)	**Deviation from theoretical value**
11/20/2020	609,557	940,425	**-35 %**
11/26/2020	1043,321	954,257	**9 %**
11/26/2020	925,932	917,952	**1 %**
LNE Medium			
Date	C(pg/mL)	C_{theo} (pg/mL)	**Deviation from theoretical value**
11/20/2020	1453,720	2201,451	**-34 %**
11/26/2020	1928,333	2222,827	**-13 %**
11/26/2020	1845,235	2178,064	**-15 %**
LNE High			
Date	C(pg/mL)	C_{theo} (pg/mL)	**Deviation from theoretical value**
11/20/2020	2735,469	3719,141	**-26 %**
11/26/2020	3669,326	3808,929	**-4 %**
11/26/2020	3593,359	3700,888	**-3 %**

Table 15: results obtained by LNE on the t-tau spiked solutions.

In Table 16, below, the mean results provided by each participant in the of the round robin study for the three CSF pools and the 3 above-mentioned spiked materials from LNE are represented with their relative CVs.

Data was obtained by performing three analytical replicates for each sample. In addition to this, LNE provided the relative uncertainty estimation for the CSF pools. The data for t-tau concentration can be compared to the nominal value for each CSF pool obtained by using the immunoassay Lumipulse

Sample	**C (nominal) [pg/ml]**	**nominal method**	**C (GOT) [pg/ml]**	**CV (GOT) [%]**	**C (UPENN) [pg/ml]**	**CV (UPENN) [%]**	**C (LNE) [pg/ml]**	**CV (LNE) [%]**	**U (LNE) [pg/ml]**
CSF pool A	299	Lumipulse	84.2	5.4	5228.3	1.5	3964.0	4.3	426 (11 %)
CSF pool B	752	Lumipulse	177.8	6.2	8043.3	4.9	7732.3	10.3	1315 (17 %)
CSF pool C	183	Lumipulse	52.1	1.2	3427.0	2.3	2433.0	4.5	271 (11 %)
rec Tau-High	1000	AAA	121.2	6.4	16103.3	4.5	3332.7	15.6	N/A
rec Tau-Mid	450	AAA	56.7	16.5	9315.7	9.0	1742.4	14.5	N/A
rec Tau-Low	150	AAA	26.7	18.6	4729.0	6.9	859.6	26.1	N/A

Table 16: results obtained by all the participants on the CSF pools and on the t-tau spiked samples. Concentrations and CV are reported for all the participants. LNE also provided uncertainty for the CSF pools.

It is clear that the values provided by the 3 participants in the round robin study are very far from the data obtained by immunoassay, implying that recalibrating immunoassays against the MS-based reference method would cause a major shift in results provided by immunoassays.

The ID-LC-MS/MS methods from LNE and UPENN gave higher results than the imunoassay for all CSF samples. This result is in agreement with other studies,[34] [35] showing a good correlation between imunoassay and LC-MS methods, but higher concentrations for the latter.

Results from GOT were lower than the nominal concentration obtained by Lumipulse. This could be due to the use of the wing-peptide material. Despite there being no agreement among the data from the individual laboratories, the results are well correlated, with a Pearson correlation coefficient R^2> 0,978 (Figure 16, below).Despite the good correlation, no agreement was found on the value to be attributed to t-tau in the different samples. Thus showing the need for the use of SI-traceable material to be used to calibrate LC-MS.

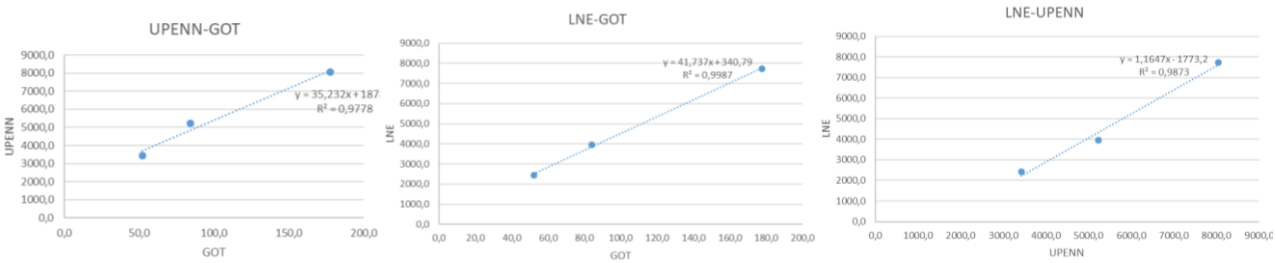

Figure 16 - correlation of the results for the different laboratories

t-tau commutability sdudy

LNE also organised a commutability study of 13 matrix-based CRMs involving 8 immunoassays from 5 IVD providers and following the latest IFCC recommendations. The involvement of the assay manufacturers paved the way toward establishing an external quality assurance (EQA) scheme to assess the accuracy and reproducibility of common methods before and after standardisation.

The aim of the commutability study was to i) evaluate comparability of immunoassays for t-tau, ii) evaluate commutability of different candidate reference materials and iii) establish correlation between the main t-tau immunoassays and the IDMS method.

This was achieved by distributing various samples including:

i. 40 single donor samples consisting of frozen CSF sourced and measured by CHU Mpt with Fujirebio's Innotest hTAU-Ag,

ii. 6 candidate CRMs consisting of pools of human frozen CSF samples prepared by CHU Mpt

iii. 7 spiked materials consisting of diluted serum spiked at 3 different concentrations with:
   - the NeuroMET Tau primary calibrator (see above),
   - a phosphorylated recombinant Tau protein, and
   - a pool of human frozen CSF containing endogenously high Tau concentrations.

In all samples, t-tau concentration was measured using:

- LNE's candidate RMP by LC-HRMS
- 5 immunoassays from 3 different IVD providers:
   - hTAU total ELISA from RJ Roboscreen,
   - Lumipulse G T-tau and INNOTEST hTau-Ag from Fujirebio,
   - T-tau ChLIA and T-tau ELISA from Euroimmun.

Samples were also distributed to Göthenburg University and to Roche Diagnostics to be measured respectively with Simoa Human T-tau Assay from Quanterix and the Elecsys T-tau CSF assay. However, results have not yet been received from these participants and will be analysed after the end of the NeuroMET2 project.

Based on data obtained on the 40 single donor samples consisting of frozen CSF, the mean CV between the 5 involved immunoassays was 19.0 % (n=40, SD=10.6 %), suggesting that comparability of immunoassays

for t-tau is largely perfectible and that standardisation of calibration material for all the laboratories might have allowed closer/equivalent results

The results of the commutability study (performed according IFCC recommendations) show that:

- commutability of the 6 candidate CRMs consisting of pools of human frozen CSF samples prepared by CHU Mpt is very good.
- commutability of the 6 materials obtained by spiking diluted serum with recombinant Tau protein is very low
- commutability of the material obtained by spiking diluted serum with a pool of human frozen CSF containing endogenously high Tau concentrations was acceptable.

These results suggest that any candidate CRMs should consist exclusively of pools of human CSF.

Finally, the correlation between the main t-tau immunoassays and the IDMS method was established. Results showed that correlation between the main t-tau immunoassays and the IDMS method is generally good, suggesting that standardising CSF t-tau measurements against IDMS is possible. An in-silico recalibration of results provided by immunoassays was performed using the 6 candidate CRMs value assigned with the candidate RMP. Based on the 40 single donor samples consisting of frozen CSF, the mean CV between the 5 involved immunoassays was decreased to from 19.0 % to 6.4 % (n=40, SD = 3.4 %), suggesting that standardisation of calibration greatly improves comparability of immunoassays for t-tau.

Overall, the results show that standardisation of t-tau immunoassays is both desirable and feasible, either through commutable CRMs or through comparison studies between immunoassays and the IDMS reference method.

JCTLM database

LNE's t-tau RMP has been submitted for inclusion in the JCTLM database. The JCTLM database lists higher-order reference materials, measurement methods and services to be used in calibration hierarchies for value assigning calibrators and trueness control materials for quantities measured by in vitro diagnostic medical devices. The JCTLM database is an important resource for ensuring traceability in laboratory medicine to reduce between method variability in the interests of improved clinical outcomes and patient safety. Therefore, the submission of the t-tau RMP on this database is an important outcome for NeuroMET2.

5 Impact

A website for the project is available at https://www.lgcgroup.com/our-programmes/empir-neuromet/ . It contains information about both this project and the preceding 15HLT04 NeuroMET project. A LinkedIn page for the project was also created and it is updated regularly https://www.linkedin.com/company/neuromet.

Newsletters were circulated to stakeholders interested in the standardisation of liquid biomarkers, MRI/MRS, and PCOMs. Two leaflets on the NeuroMET projects and their results were also prepared and distributed to cohort volunteers and their families. These leaflets were made available in both English and German through the project website and LinkedIn page. Further to this, information events for study participants and their caregivers were held at Charité in July 2019 and November 2022.

The project was the subject of 35 conference presentations. Conferences attended included the Joint Conference of the Society for European Magnetic Resonance and the International Society for Magnetic Resonance EUROISMAR 2019, the Annual Meeting of the International Society of Magnetic Resonance in Medicine (ISMRM 2020, 2021, 2022), the Annual Scientific Symposiums on Ultrahigh Field Magnetic Resonance (2019), Alzheimer's Association International Conference (AAIC 2019, 2021, and 2022); the International Society for Magnetic Resonance in Medicine (2021, 2022), the International Metrology Congress (CIM) 2021, IMEKO 2021, and the International Lewy Body Dementia Conference 2022.

The project generated 14 open access peer-reviewed publications, of which 10 had international co-authorship between consortium partners.

Impact on industrial and other user communities

One of this project's goals was to bring benefits to the pharmaceutical and in vitro diagnostic industry, clinicians, and ultimately patients, by providing a set of metrologically validated methods to: (i) improve targeted NDD recruitment in clinical trials and to accurately monitor the efficacy of new therapeutics; (ii) facilitate the regulatory approval of new assays and their uptake into clinics; and (iii) enable accurate diagnosis of NDD patients, facilitate clinical decisions, and hence improve clinical outcomes. This project has achieved this goal through the deployment of the NMM and its associated app (Objective 2), and the provision of RMPs for key NDD biomarkers in both biological fluid testing and MIR/MRS (Objective 4) and their translation to clinical settings.

The project's metrologically validated app (Objective 2), is the first of its kind in the field of cognitive assessment and based data from the project's unique cohort (Objective 1) and is a significant project output and represents more accurate a robust route to memory testing. The phase one roll out of the app to clinicians at Charité and other institutes was completed during the project and a wider roll out will continue after the end of the project. The app will provide mobile NDD health information to clinicians with the potential for dynamic engagement with patients and health care providers, and a new means of improving health outcomes. The NMM and the app were disseminated to a wider group of users beyond the original testing group via a webinar in November 2022.

The validated RMPs developed for key NDD biomarker provide an SI-traceable reference against which performance of in-vitro diagnostics can be assessed (Objective 4). Roll out of these to clinical communities has included the development of a candidate MS clinical assay for monitoring α-synuclein in CSF, and the delivery of a commutability study on t-tau under the IFCC WG-CSF which has provided IVD manufacturers with insights into the performance of their products relative to a reference method.

Within MRI and MRS, direct links to clinical end users were maintained though the involvement of clinicians and industry as partners (i.e. partners Charité, Uni-Greif, Modus), and via transfer of the project's reference 7T MRI/MRS sequences to widely used scanners in a clinical setting at Uni-Greif. An agreement was made with Siemens (one of the leading manufacturers of MRI machines in Europe) to enable transfer of this project's 3T protocol (Objective 3) to the clinical environment at Uni-Greif. The 3T protocol was derived from the 7T NeuroMET protocol in the 15HLT04 NeuroMET project. Furthermore, hands on training on MRS data acquisition was provided by PTB to MR physicists at Uni-Greif in order to support MRS data acquisition uptake in clinical settings, and a webinar was organised to disseminate best practice to the clinical community more widely (Objective 3).

The outputs of the project were also presented to stakeholders at key clinical conferences including AAIC 2019, 2021, and 2022 which is the largest AD conference worldwide, and at International Lewy Body Dementia Conference 2022.

Longer term, the projects outputs will support pharmaceutical development with the PCOMs and blood tests (Objectives 2, 3, 4 & 5) developed and validated within the project contributing to identification of NDD patients before clinical on-set. This will support the recruitment of groups of patients to clinical trials for new therapeutics and hence increase their rate of success. The project's RMPs for CSF biomarkers, and the high-resolution MRI and MRS protocols (Objectives 3 & 4) will also increase confidence in the recruitment of targeted NDD patients for new drug trials.

Impact on the metrology and scientific communities

This project has built on the work of 15HLT04 NeuroMET to enhance the application of metrological concepts, which are commonplace in disciplines such as physical and chemical measurements, into the area of cognition and mathematical RMPs. The standardisation of the results from cognitive assessments and their application to develop a memory metric (the project's NMM) is the first example of standardisation of PCOMs (Objectives 2 & 5). This is a unique advancement for the metrological and scientific community, together with the development of mathematical models to be used as primary RMPs for cognition.

The application of the project's metrological concepts such as the measurement uncertainty of *in vivo* MRI and MRS results (Objective 3) should enable significant progress, not only in the definition of NDD diagnostic thresholds, but also in establishing novel RMPs for *in vivo* MRI and MRS. The data associated with this work has been made open access and freely available via the project's Zenodo community which will allow other researchers to replicate and enhance this novel uncertainty determination framework. The project also achieved pan-European impact in this area by close alignment and collaboration with the 18HLT05 QUIERO project in relation to T1-mapping examinations.

The project's RMPs for protein biomarkers (Objective 4) will help the scientific and metrological community address the challenges faced in standardising measurements of larger and more complex biomolecules. SI-traceable RMPs for protein biomarkers is still an emerging field relative to similar methods for small molecules, and work under this project (for example on tau) has demonstrated the feasibility and utility of such RMPs and their associated primary calibrators. LNE's validated RMP for t-tau reference measurement procedure was accepted for inclusion to the JCTLM database, which will further increase its impact on the laboratory medicine and IVD communities.

The project pursued close alignment with key European Metrology Networks (EMNs) on 'Traceability in Laboratory Medicine' (TraceLabMed) and MATHMET, in order to support their integration into international initiatives such as EFLM (European federation of Laboratory Medicine) and EUFIND (European Ultrahigh-Field Imaging Network for Neurodegenerative Diseases).

Work from the project was also presented at key metrology conferences including CIM 2021, IMEKO 2021, and MSMM 2021; and key technical conferences including the International Society for Magnetic Resonance in Medicine (2021, 2022), Joint Conference of the Society for European Magnetic Resonance (EUROMAR) and the International Society for Magnetic Resonance (ISMAR), EUROISMAR 2019, and the 3rd International Hydrogen Deuterium Exchange Mass Spectrometry conference.

Impact on relevant standards

The project maintained close alignment with a number of standardisation committees including ISO TC 12 Quantities and Units and ISO TC 212 Clinical laboratory testing and in vitro diagnostic test systems, ISO TC 215 on Health informatics. Many of the scientists involved in the delivery of the project sit on key standardisation committees.

The project also provided input to the IFCC WG-CSF, the Society of CSF analysis and clinical Neurochemistry, EUFIND, JCTLM and BIPM's Consultative Committee for Amount of Substance: Metrology in Chemistry and Biology (CCQM) Working Groups. Standardisation activities to highlight include (i) a p-tau interlaboratory value assignment study and (ii) a t-tau commutability study, both organised in association with the IFCC WG-CSF, (iii) acceptance of the t-tau RMP to the JCTLM database, and (iv) presentation of the project's work at key standardisation conferences including JCTLM Accurate Results for Patient Care Workshop 2019.

Longer-term economic, social and environmental impacts

Many NDDs such as AD are irreversible and progressive. In addition to large socioeconomic costs, they severely affect the quality of life of patients and their caregivers. Early diagnosis through implementation of screening programs, the identification of people with risk factors, and the development of new therapeutics are vital for delaying the onset of symptoms and improving the quality of life of NDD patients. Thus, in the long term, this project will help to decrease the socioeconomic burden of NDD, reduce the resources spent by the pharmaceutical industry, and improve the quality of life of NDD patients and their caregivers.

6 List of publications

1. Melin, J., Cano., S.J. and Pendrill, L. The Role of Entropy in Construct Specification Equations (CSE) to Improve the Validity of Memory Tests. Entropy. Dec 2020. https://doi.org/10.3390/e23020212
2. J. Melin, S.J. Cano, L. Göschel, A. Fillmer, S. Lehmann, C. Hirtz, A. Flöel, L.R. Pendrill, Metrological references for person ability in memory tests, Measurement: Sensors, Volume 18, 2021, https://doi.org/10.1016/j.measen.2021.100289
3. J. Melin, S.J. Cano, A. Flöel, L. Göschel, L.R. Pendrill, Construct specification equations: 'Recipes' for certified reference materials in cognitive measurement, Measurement: Sensors, Volume 18, 2021, https://doi.org/10.1016/j.measen.2021.100290
4. Riemann, LT, Aigner, CS, Ellison, SLR, et al. Assessment of measurement precision in single-voxel spectroscopy at 7 T: Toward minimal detectable changes of metabolite concentrations in the human brain in vivo. Magn Reson Med. 2022; 87(3): 1119 – 1135, https://doi.org/10.1002/mrm.29034
5. Hui, SCN, Mikkelsen, M, … Fillmer, A, et al. Frequency drift in MR spectroscopy at 3T. NeuroImage 241: 118430 (2021), https://doi.org/10.1016/j.neuroimage.2021.118430
6. Melin, J., Cano, S., Flöel, A., Göschel, L., & Pendrill, L. (2022). The Role of Entropy in Construct Specification Equations (CSE) to Improve the Validity of Memory Tests: Extension to Word Lists. Entropy, 24(7), Art. 7. https://doi.org/10.3390/e24070934
7. Melin, J., Cano, S. J., Flöel, A., Göschel, L., & Pendrill, L. R. (2022). Metrological advancements in cognitive measurement: A worked example with the NeuroMET memory metric providing more reliability and efficiency. Measurement: Sensors, 100658. https://doi.org/10.1016/j.measen.2022.100658
8. Melin, J. et al, The Role of Construct Specification Equations and Entropy in the Measurement of Memory, in Person-Centered Outcome Metrology: Principles and Applications for High Stakes Decision Making in Springer International Publishing, William P. Fisher Jr. & S. J. Cano (ed.) (s. 269–309), https://doi.org/10.1007/978-3-031-07465-3_10
9. Pendrill, LR. et al, Assuring measurement quality in person-centred care, in Person-Centered Outcome Metrology: Principles and Applications for High Stakes Decision Making in Springer International Publishing, William P. Fisher Jr. & S. J. Cano (ed.) (s. 311-335), https://doi.org/10.1007/978-3-031-07465-3_11
10. Person-Centered Outcome Metrology: Principles and Applications for High Stakes Decision Making in Springer International Publishing, William P. Fisher Jr. & S. J. Cano (ed.), https://doi.org/10.1007/978-3-031-07465-3
11. Das, S.; Dewit, N.; Jacobs, D.; Pijnenburg, Y.A.L.; In 't Veld, S.G.J.G.; Coppens, S.; Quaglia, M.; Hirtz, C.; Teunissen, C.E.; Vanmechelen, E. A Novel Neurofilament Light Chain ELISA Validated in Patients with Alzheimer's Disease, Frontotemporal Dementia, and Subjective Cognitive Decline, and the Evaluation of Candidate Proteins for Immunoassay Calibration. Int. J. Mol. Sci. 2022, 23, 7221. https://doi.org/10.3390/ijms23137221
12. Riemann, LT, Aigner, CS, et al. Fourier-based decomposition for simultaneous 2-voxel MRS acquisition with 2SPECIAL, Magnetic Resonance in Medicine, 88(5):1978–1993, 2022, https://doi.org/10.1002/mrm.29369

13. Giangrande C, Vaneeckhoutte H, Boeuf A, Lalere B, Hirtz C, Lehmann S, Quaglia M and Delatour V. Development of a candidate reference measurement procedure by ID-LC-MS/MS for total tau protein measurement in cerebrospinal fluid (CSF), Clinical Chemistry and Laboratory Medicine (CCLM), 2023, http://dx.doi.org/10.1515/cclm-2022-1250

14. Melin, J., Cano, S.J., Gillman, A. et al. Traceability and comparability through crosswalks with the NeuroMET Memory Metric. Sci Rep 13, 5179 (2023). https://doi.org/10.1038/s41598-023-32208-0

This list is also available here: https://www.euramet.org/repository/research-publications-repository-link/

7 Contact details

Coordinator Will Webster, LGC

Tel: +44 20 8943 7523

E-mail: william.webster@lgcgroup.com

8 References

[1] Wirth M, Schwarz C, Benson G, Horn N, Buchert R, Lange C, et al. Effects of spermidine supplementation on cognition and biomarkers in older adults with subjective cognitive decline (SmartAge)—study protocol for a randomized controlled trial. Alzheimer's Research & Therapy. 2019;11(1).

[2] Hobart J, Cano S, Posner H, Selnes O, Stern Y, Thomas R, et al. Putting the Alzheimer's cognitive test to the test II: Rasch Measurement Theory. Alzheimer's & dementia. 2013;9(1S):S10-S20.

[3] Andrich D. Rasch Models for Measurement, 1988.

[4] Linacre J. Understanding Rasch measurement: estimation methods for Rasch measures. Journal of outcome measurement. 1999;3:382-405.

[5] Melin J, Cano S, Pendrill L. The Role of Entropy in Construct Specification Equations (CSE) to Improve the Validity of Memory Tests. Entropy. 2021;23(2):212.

[6] Pendrill LR. Assuring measurement quality in person-centred healthcare. Measurement Science and Technology. 2018;29(3):034003.

[7] Pendrill LR. Quality Assured Measurement: Unification across Social and Physical Sciences.: Springer International Publishing. ; 2019.

[8] Melin J, Cano S, Gillman A, Marquis S, Floel A, Goschel L, et al. NeuroMET Memory Metric: Traceability and comparability through crosswalks. 2022.

[9] Bernstein MA, King KF, Zhou XJ. Handbook of MRI Pulse Sequences. Elsevier; 2004. doi:10.1016/B978-0-12-092861-3.X5000-6

[10] Tannús A, Garwood M. Improved performance of frequency-swept pulses using offset-independent adiabaticity. J Magn Reson - Ser A. 1996;120(1):133-137. doi:10.1006/jmra.1996.0110

[11] Andronesi OC, Ramadan S, Ratai E-M, Jennings D, Mountford CE, Sorensen AG. Spectroscopic Imaging with Improved Gradient Modulated Constant Adiabadicity Pulses on High-Field Clinical Scanners. J Magn Reson. 2011;23(1):1-7. doi:10.1161/CIRCULATIONAHA.110.956839

[12] Riemann LT, Aigner CS, Ellison SLR, et al. Assessment of measurement precision in single-voxel spectroscopy at 7 T: Toward minimal detectable changes of metabolite concentrations in the human brain in vivo. Magn Reson Med. 2022;87(3):1119-1135. doi:10.1002/mrm.29034

[13] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. Evaluation of measurement data | Guide to the expression of uncertainty in measurement. Joint Committee for Guides in Metrology, JCGM 100:2008. URL: https://www.bipm.org/documents/20126/2071204/ JCGM_100_2008_E.pdf/cb0ef43f-baa5-11cf-3f85-4dcd86f77bd6.

[14] Pendrill L. R., (2019) Quality Assured Measurement - Unification across Social and Physical Sciences, Springer Series in Measurement Science and Technology, DOI: 10.1007/978-3-030-28695-8

[15] Leslie R Pendrill, Jeanette Melin, Stefan J Cano, the EMPIR NeuroMET 15HLT04 consortium, Metrological references for health care based on entropy, International Congress of Metrology 07001 (2019), DOI: 10.1051/metrology/201907001

[16] Mari L, Wilson M. An introduction to the Rasch measurement approach for metrologists. Measurement. 2014 May;51:315–27.

[17] Hobart JC, Cano SJ, Thompson AJ. Effect sizes can be misleading: is it time to change the way we measure change? Journal of Neurology, Neurosurgery & Psychiatry. 2010 Sep 1;81(9):1044–8.

[18] Hughes LF, Perkins K, Wright BD, Westrick H. Using a Rasch scale to characterize the clinical features of patients with a clinical diagnosis of uncertain, probable, or possible Alzheimer disease at intake. JAD. 2003 Nov 17;5(5):367–73.

[19] Hobart J, Cano S, Posner H, Selnes O, Stern Y, Thomas R, et al. Putting the Alzheimer's cognitive test to the test I: Traditional psychometric methods. Alzheimers Dement. 2013 Feb;9(1 Suppl):S4-9. doi: 10.1016/j.jalz.2012.08.005.

[20] Rasch G. Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests. Oxford, England: Nielsen & Lydiche; 1960.

[21] Cano SJ, Pendrill LR, Melin J, Fisher WP. Towards consensus measurement standards for patient-centered outcomes. Measurement. 2019 Jul;141:62–9.

[22] Melin J, Pendrill LR, Cano SJ, EMPIR NeuroMET 15HLT04 consortium. Towards patient-centred cognition metrics. J Phys: Conf Ser. 2019 Nov;1379:012029.

[23] Melin J, Kettunen P, Wallin A, Pendrill L. Entropy-based explanations of serial position and learning effects in ordinal responses to word list tests. 2022; IMEKO Conference, Porto.

[24] Melin J, Pendrill L. A Novel Metrological Approach to a More Consistent Way of Defining and Analyzing Memory Task Difficulty in Word Learning List Tests with Repeated Trials. In Marseille; 2022. Available from: http://www.lrec-conf.org/proceedings/lrec2022/workshops/RaPID4/2022.rapid4-1.0.pdf

[25] Shannon CE. A Mathematical Theory of Communication. The Bell System Technical Journal. 1948;55.

[26] Brillouin L. Science and Information Theory. Second Edition. New York; 1962. Available from: https://www.amazon.com/Science-Information-Theory-Second-Physics/dp/0486497550

[27] Melin J, Cano S, Flöel A, Göschel L, Pendrill L, EMPIR NeuroMET and NeuroMET2 consortiums. More than a memory test: A new metric linking blocks, numbers, and words. Alzheimer's & Dementia. 2021;17(S6):e050291.

[28] Melin J, Cano S, Flöel A, Göschel L, Pendrill L. The Role of Entropy in Construct Specification Equations (CSE) to Improve the Validity of Memory Tests: Extension to Word Lists. Entropy. 2022 Jul;24(7):934.

[29] Melin, J., Cano, S. J., Flöel, A., Göschel, L., & Pendrill, L. R. (2022). Metrological advancements in cognitive measurement: A worked example with the NeuroMET memory metric providing more reliability and efficiency. Measurement: Sensors, 100658. https://doi.org/10.1016/j.measen.2022.100658

[30] Pendrill L. Man as a Measurement Instrument. NCSLI Measure. 2014 Dec;9(4):24–35.

[31] Pendrill L. Quantities and units in quality assured measurement. Conference presentation presented at: Pacific RIM Objective Measurement Symposium 2021; 2021 Dec 6. Available from: https://proms.promsociety.org/2021

[32] Melin J. Neurogenerative disease metrology and innovation: The European Metrology Programme for Innovation & Research (EMPIR) and the NeuroMET projects. Conference presentation presented at: Pacific RIM Objective Measurement Symposium 2021; 2021 Dec 6. Available from: https://proms.promsociety.org/2021/

[33] Chiara Giangrande et al, Development of a candidate reference measurement procedure by ID-LC-MS/MS for total tau protein measurement in cerebrospinal fluid (CSF), CCLM, *Approved awaiting publication*

[34] Bros P, Vialaret J, Barthelemy N, Delatour V, Gabelle A, Lehmann S et al. Antibody free quantification of seven tau peptides in human CSF using targeted mass spectrometry. Front Neurosci 2015;9:302. doi: 10.3389/fnins.2015.00302. PMID: 26388715; PMCID: PMC4555028

[35] Barthélemy NR, Gabelle A, Hirtz C, Fenaille F, Sergeant N, Schraen-Maschke S et al. Differential mass spectrometry profiles of tau protein in the cerebrospinal fluid of patients with Alzheimer's disease, progressive supranuclear palsy, and dementia with Lewy bodies. J Alzheimers Dis 2016;51(4):1033-43. doi: 10.3233/JAD-150962. PMID: 26923020.