Letter of support for N170 ERP as a prognostic biomarker for adaptive social functioning and its potential to stratify study populations in people with Autism spectrum disorders (ASD) without intellectual disability

On 17 April 2020 the Applicant Roche Registration GmBH requested a follow-up qualification advice on behalf of the AIMS-2-Trials Consortium on the latency of the N170 event-related potential (ERP) as a prognostic biomarker for adaptive social functioning in people with Autism spectrum disorders (ASD) without intellectual disability pursuant to Article 57(1)(n) of Regulation (EC) 726/2004 of the European Parliament and of the Council. A more general use of use of the N170L as a stratification factor in randomised clinical trials was additionally proposed in the course of the assessment.

AIMS-2-TRIALS (Autism Innovative Medicine Studies-2-Trials) is a public-private research project at the forefront of autism research funded under the Innovative Medicines Initiative-2. It began in June 2018 and will run until May 2023. The research programme includes a range of studies carried out by different groups exploring how autism develops, from before birth to adulthood, and how this varies in different people. The Consortium will seek to identify prognostic biomarkers which indicate whether a person with autism has or may develop particular characteristics, and predictive biomarkers that could also help to identify who may ultimately benefit from particular treatments.

Background and proposed context of use

Autism Spectrum Disorder (ASD) is a heterogeneous neurodevelopmental disorder characterized by qualitative impairments in two core domains: social interaction and communication; and the presence of repetitive or restricted behaviours, interests, or activities and sensory abnormalities. To date, there are no approved pharmacological treatments that effectively improve the core symptoms of ASD and, given the significant impact on the quality of life of individuals with ASD and their caregivers, the unmet medical need is high.

It is likely that different pathophysiological pathways contribute to the core symptomatic domains and account for the heterogeneity of the autistic population in terms of symptom expression and severity. This heterogeneity presents a challenge to the development of pharmacological interventions aimed to alleviate the core symptoms of ASD. The inclusion of more homogenous groups of patients in treatment trials could accelerate the pace of intervention development in ASD (Loth et al. 2016).

Expertise with faces is central to social interaction and face expertise is often associated with increased speed of processing. A common approach to measuring the speed of particular neural processes is the

generation of 'event-related potentials' (ERPs) – brain responses that reliably occur time-locked to the onset of a particular stimulus.

The N170 is an ERP that has been shown to be sensitive to expertise effects, particularly social expertise, and has been linked to core regions of the "social brain" such as bilateral temporal cortex and posterior fusiform (Schurz et al. 2014)

A long history of research has shown that there are atypicalities in the N170 in ASD (McPartland et al. 2004). A recent meta-analysis confirmed that across 16 studies, autistic people show a longer latency (slower) N170 to faces stimuli than neurotypical controls (Kang et al. 2018).

The EU-AIMS Longitudinal European Autism Project (LEAP) is a large multicentre observational study that aims to identify factors that contribute to differences in brain development, difficulties in social behaviour and other core symptoms of ASD. Results of the LEAP study confirmed that the N170 latency to upright faces is on average altered in individuals with ASD relative to the neurotypical population, providing initial evidence of its putative suitability as a diagnostic biomarker for core symptom domains. The LEAP study data also offers:

1. Evidence that the variation in N170 latency within the ASD group is not confounded by the presence of associated psychiatric conditions.

2. Evidence of an association between N170 latency and the face-sensitive response in the right fusiform gyrus as measured by functional magnetic resonance imaging (fMRI).

3. Evidence that N170 latency predicts change in socialisation metrics (specifically, parent report of play and leisure time activities, a subdomain of the Vineland™-II Adaptive Behavior Scales).

Additional evidence from a US-based consortium (ABC-CT) has demonstrated the reliability of the N170 ERP and the replicability of results from the LEAP study. Further, additional evidence from a UK-based prospective longitudinal study of infants with older siblings with autism (BASIS) suggests that an early precursor of the N170 is altered in infants who are later diagnosed with autism. The early emergence of atypicalities raises the likelihood that a prognostic biomarker has value in longitudinal prediction.

The AIMS-2-TRIALS Consortium initially proposed use ERP N170 latency induced by upright faces as a biomarker for enriching clinical trials in a paediatric ASD population without intellectual disability for those who will show a poor prognosis of adaptive social functioning. During the course of the discussion meeting, the Consortium acknowledged that it may be premature to restrict the context of use to population enrichment and proposes to evaluate a more general use of the N170L as a stratification factor in randomised clinical trials where the Play and Leisure Time domain is the targeted primary outcome. This is acknowledged by the SAWP, but the prognostic ability of the marker and its utility as a stratification factor in future clinical trials (in order to increase the power, or likewise, the precision of effect estimates) is yet to be shown. Validation using an independent data set would be recommended to address this objective.

Establishing cut-offs

A normative modelling approach was used to generate z-scores that represent their position within the population adjusted for age, relative to a neurotypical control group (z-N170L). Even if the subpopulation of patients not expected to show natural improvement regarding the Vineland Socialisation Play and Leisure Time subscale could be considered a meaningful target population with expected better response to treatment, it is questionable whether the proposed z-N170L based on ERP N170 latency is a suitable tool to support enrichment.

The Consortium detailed during the discussion meeting that the normative modelling is based on a neurotypical sample, rather than an ASD population based sample, because the overall goal is to characterise patients with ASD by their degree of deviation from what would be expected in neurotypical persons matched on age. The neurotypical cohort from the LEAP sample (age 6 to 30 years) was used to build the initial model. The model considers age as a continuous variable and the goal is to measure individuals with ASD by their degree of deviation from a neurotypical trajectory. This is acknowledged and the robustness of the chosen modelling approach should be evaluated.

Based on an ROC curve (Area=0.67, standard error=0.04, 95%CI = 0.59-0.76), a cut-off of ≥ 0.5 for the z-N170L was selected. A cut-off of 0.5 yields a sensitivity of approximately 0.55 and specificity of approximately 0.76 for detecting non-improvers. The intra-individual variability of the z-N170L is unclear such that it is not obvious whether it can be reliably determined that a patient falls below a specific cut-off. It should be assessed how the reported ICC for ERP N170 could translate in intra-individual variability of the z-N170L, and in how far this impacts the prognostic value of the marker.

Validation of the cut-off in an independent data set would be required, where even lower sensitivity and/or specificity may well be expected. The justification for the specific cut-off (or the chosen trade-off between sensitivity and specificity) 'to maximise specificity over sensitivity' is not satisfactory, as maximizing specificity in the sample could be even improved by choosing a more conservative cut-off at the cost of the sensitivity. However, alternative choices would probably not be superior to the Consortium's proposal. It may also be questioned whether the dichotomisation of changes of the Vineland Socialisation Play and Leisure Time subscale into subjects who improved and those that did not improve is arbitrary and does not appropriately capture changes in development of the subscale over time. Alternatively, the prognostic model may just simply include the changes of the Vineland Socialisation Play and Leisure Time subscale in a regression model (e.g. subscale at specific time point as dependent variable and baseline value included as covariate). Cognitive ability is expected to affect this variable and may thus be accounted for in the model. Categorisation (and exclusion if IQ<70) may also be questioned for this variable, i.e. IQ may be considered as a continuous covariate and the functional relationship may be further explored. The dependency on age or other variables affecting the outcome may also be evaluated and eventually taken into account.

The Consortium provided also simulations to determine the gain in power by ERP N170 latency enrichment. However, the simulations are based on the same data set in which the association between ERP N170 latency and change in the subscale was observed, and that was also used to determine the cut-off for the z-N170L. Therefore, an increase in power by restricting to the subgroup of patients with slower ERP N170 latency that was already known to have a more homogeneous outcome than the overall study population is not surprising. Of note, the same treatment effect was assumed in all patients irrespective to which subpopulations they belong; therefore, increase in power is exclusively explained by the more homogeneous subpopulation. Thus, the simulations cannot answer the question if or to what magnitude sample size requirements can be reduced by enrichment. Even if the limitations of the simulations are ignored, the decrease from 78 patients to 48 patients to achieve a power of 80% seems not to be impressive.

The consortium notes that the underlying biology driving N170 latency may also predispose people to respond more or less strongly to any one intervention. This suggests that N170 latency may potentially be a marker that is (also) predictive for treatment response (rather than prognostic, as assumed for above considerations). This should be further investigated and may depend on the mechanism of action of the drug to be tested.

Overall, before a final conclusion on meaningfulness of enrichment by ERP N170 latency can be made, several assumptions made by the Consortium which are only based on hypotheses or based on exploratory findings in the LEAP study population need to be validated in independent populations. In addition, the performance of the ERP N170 latency regarding the potential for identifying patients with poor prognosis is even not convincing in the LEAP population, and potential for reduction in sample size seems to be limited. Therefore, it appears questionable that the validation of ERP N170 latency as a suitable biomarker for enrichment will finally be successful. Alternatives to enrichment such as including ERP N170 latency as a covariate in statistical analysis, or stratified randomisation may also increase efficiency and are intended to be investigated by the Consortium. This is welcomed.

Further studies to support qualification of the N170 latency biomarker

A longitudinal study in pre-schoolers, follow-up data of the LEAP study in wave 3 and a double-blind randomized placebo-controlled trial versus Abaclofen are proposed to further examine the FACE-ERP task and the Vineland scale.

a) The extension to pre-schoolers may generate further evidence on the prognostic value of N170 latency regarding trajectories of socialisation. However, it is noted that Vinleand^{TM}-3 will be used instead of Vineland^{TM}-II such that it needs to be justified in how far results are comparable. In addition, the prognostic value may be age-dependent.

b) The proposed replication by additional data collection in the LEAP cohort is an important first step to replicate the core findings. If the validation of the association between N170 latency and trajectories of Vineland Socialisation (Play and Leisure Time) scores is successful, the performance of the z-N170L regarding the identification of non-improvers with regard to this score should also be evaluated. However, as stated above and in the previous advice, to gain evidence for any prognostic or predictive value the consortium will need confirmation and replication in an independent data set. Additional data collection in the same data set is not sufficient in this respect as N170 latency as well as trajectories of Vineland Socialisation (Play and Leisure Time) scores can be expected to be correlated within patients (wave 2 and 3 differ only by the length of follow-up) such that these data cannot be considered independent.

c) The extension to a clinical trial context is supported, but the objective of assessing N170's utility for future trial planning should be clearly stated. For example, an increase in power is always expected when restricting the study population based on a prognostic marker, but enrichment on the basis of a predictive marker may be of main interest. i.e. whether there is a treatment*covariate interaction such that the treatment effect is larger in the subgroup defined by use of the z-N170L cut-off. A potential increase in power may be difficult to quantify, because this depends on assumptions that cannot be verified (without uncertainty). The proportion of explained variance can be evaluated; this should also be compared to a model including z-N170L as covariate, continuously or categorical. The objective of assessing N170 utility for future population enrichment (or use as a stratification factor) is only vaguely formulated in the study synopsis. The objectives (including success criteria) should be specified in more detail in advance.

In addition, Vineland™-3 will again be used instead of Vineland™-II.

In summary, the generation of further data is supported. Follow-up Qualification Opinion is recommended as soon as additional data become available.

Sincerely,

Emer Cooke
Executive Director

References:

Kang, E., Keifer, C. M., Levy, E. J., Foss-Feig, J. H., McPartland, J. C., & Lerner, M. D. (2018). Atypicality of the N170 Event-Related Potential in Autism Spectrum Disorder: A Meta-analysis. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *3*(8), 657–666. https://doi.org/10.1016/j.bpsc.2017.11.003

Loth, E., Spooren, W., Ham, L. M., Isaac, M. B., Auriche-Benichou, C., Banaschewski, T., … Murphy, D. G. M. (2016). Identification and validation of biomarkers for autism spectrum disorders. *Nature Reviews Drug Discovery*, *15*(1), 70–70. https://doi.org/10.1038/nrd.2015.7

McPartland, J., Dawson, G., Webb, S. J., Panagiotides, H., & Carver, L. J. (2004). Event-related brain potentials reveal anomalies in temporal processing of faces in autism spectrum disorder. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *45*(7), 1235–1245. https://doi.org/10.1111/j.1469-7610.2004.00318.x

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neuroscience & Biobehavioral Reviews*, *42*, 9–34. https://doi.org/10.1016/j.neubiorev.2014.01.009