




# Increasing the reproducibility of fluid biomarker studies in neurodegenerative studies

Niklas Mattsson-Carlgrén <sup>1,2,3✉</sup>, Sebastian Palmqvist<sup>1,4</sup>, Kaj Blennow <sup>5,6</sup> & Oskar Hansson <sup>1,4✉</sup>

Biomarkers have revolutionized scientific research on neurodegenerative diseases, in particular Alzheimer's disease, transformed drug trial design, and are also increasingly improving patient management in clinical practice. A few key cerebrospinal fluid biomarkers have been robustly associated with neurodegenerative diseases. Several novel biomarkers are very promising, especially blood-based markers. However, many biomarker findings have had low reproducibility despite initial promising results. In this perspective, we identify possible sources for low reproducibility of studies on fluid biomarkers for neurodegenerative diseases, with a focus on Alzheimer's disease. We suggest guidelines for researchers and journal editors, with the aim to improve reproducibility of findings.

**N**eurodegenerative diseases, including Alzheimer's disease (AD), account for significant morbidity, mortality, and costs worldwide. A major problem for research, clinical practice, and drug development is that diagnosis, prognosis and disease monitoring are difficult using clinical examination alone. Clinical examination is particularly problematic in early disease stages and cannot often on its own guide diagnosis or predict progression. Biomarkers have been introduced as a way to improve this, by providing objective measures of the underlying pathophysiology<sup>1</sup>. Very successful results have been achieved for fluid biomarkers, including both in cerebrospinal fluid (CSF) and blood<sup>2</sup>. For AD, biomarkers of relevant brain changes have even been incorporated into research definition of the disease (using biomarkers for  $\beta$ -amyloid [A $\beta$ ] pathology, tau pathology, and neurodegeneration)<sup>3</sup>. Examples of highly reproducible biomarkers that to some degree are used in clinical practice for diagnosis of neurodegenerative diseases include CSF A $\beta$ 42, the A $\beta$ 42/40 ratio, total-tau (T-tau) and phosphorylated tau (P-tau) for AD diagnosis<sup>4</sup>, and real-time quaking-induced conversion (RT-QuIC) assays on CSF for CJD<sup>5–7</sup>. CSF levels of neurofilament light (NFL) protein<sup>8</sup> is also sometimes used in clinical practice as a disease-unspecific biomarker for neuronal injury, to detect the presence of degeneration and thereby support a diagnosis of a neurodegenerative condition, e.g., Amyotrophic Lateral Sclerosis (ALS)<sup>9</sup>. When disease-modifying treatments for common neurodegenerative diseases become available, biomarkers may also have a role to guide usage of treatments in clinical practice. This may be relevant very soon, giving the promising recent results of certain immunotherapies against AD<sup>10–12</sup>.

However, for many very promising biomarker findings, including biomarker panels<sup>13–16</sup>, replication efforts have failed<sup>17</sup>. Poor reproducibility is, however, not a problem isolated to

<sup>1</sup>Clinical Memory Research Unit, Faculty of Medicine, Lund University, Lund, Sweden. <sup>2</sup>Department of Neurology, Skåne University Hospital, Lund University, Lund, Sweden. <sup>3</sup>Wallenberg Center for Molecular Medicine, Lund University, Lund, Sweden. <sup>4</sup>Memory Clinic, Skåne University Hospital, Malmö, Sweden. <sup>5</sup>Clinical Neurochemistry Laboratory, Sahlgrenska University Hospital, Mölndal, Sweden. <sup>6</sup>Institute of Neuroscience and Physiology, University of Gothenburg, Gothenburg, Sweden. ✉email: [niklas.mattsson@med.lu.se](mailto:niklas.mattsson@med.lu.se); [oskar.hansson@med.lu.se](mailto:oskar.hansson@med.lu.se)

biomarker research<sup>18,19</sup>. In 2016, the journal *Nature* published results from a survey taken by 1576 researchers from many scientific disciplines, where 52% thought that there was a “significant crisis” in reproducibility of published research, due to factors such as selective reporting, pressure to publish, low statistical power or poor analysis, too little replication in the original lab, publication bias, and other factors<sup>20</sup>. The problem is aggravated by the fact that small studies that overestimate effects (including for biomarker performance) may be more likely to get cited than larger studies with more sobering results<sup>21</sup>. Publication of biomarker findings with low reproducibility is a waste of time and money for researchers and assay developers aiming to replicate the results.

In this perspective article, we therefore analyze reproducibility issues for fluid biomarkers for neurodegenerative diseases. This is not a systematic quantitative review of all fluid biomarker studies that have been published for neurodegenerative diseases. We have selected examples of studies that represent different types of reproducibility problems. Most of our examples are from the AD-field, but we believe that the recommendations are applicable as guiding principles also for other neurodegenerative diseases. We explore several amendable sources of poor reproducibility, including cohort design, pre-analytical and analytical factors, and statistical procedures. We suggest factors that may be taken into consideration when designing and publishing biomarker studies. Our ambition is to help researchers, drug developers and scientific journals to achieve reproducible results and advance the field of biomarkers in neurodegenerative diseases.

### Factors affecting reproducibility

There are many potential sources of poor reproducibility for biomarker studies. These include cohort-related factors, pre-analytical factors, analytical and kit-related factors, biotemporal variability of the measured markers, insufficient statistical methods, lack of proper validation, and factors related to the decisions to submit or accept publications for publishing. Several of these factors contribute to poor reproducibility by either introducing imprecision (a random error) of measurements (increasing “variability”—the closeness of agreement between biomarker measurements obtained by replicate measurements on the same or similar objects under specified condition<sup>22</sup>), or introducing bias (a systematic error) of measurements (reducing “accuracy”—the closeness of agreement between a biomarker measurement and the true value of the biomarker<sup>22</sup>). We summarize examples of these factors and considerations in two figures. The first part of studies, from cohort recruitment up until biomarker measurements (including cohort-related, assay-related, pre-analytical and analytical factors) are summarized in Fig. 1. The later part of studies and validation efforts (including choices of relevant comparisons, statistical analyses, and different levels of validation) are summarized in Fig. 2. Each of these components are described in detail below.

**Cohort-related factors.** Suboptimal cohort design increases the likelihood of overoptimistic findings. First, studies with small patient and control samples typically overestimate performance compared to larger studies<sup>23,24</sup>. This is logical, since only large effects can be detected with a small sample. If by chance a larger effect is present in the smaller sample this may increase the likelihood of publication, overestimating the effect in an initial small pilot study and leading to publication bias. A funnel plot may indicate if publication bias is present across a number of studies<sup>25</sup>.

Second, if study participants are pre-selected, or if extensive inclusion or exclusion criteria are applied, findings may have

lower reproducibility compared to if participants are consecutively or randomly recruited. This is because in a more heterogeneous (more real-world like) sample, several factors may contribute to both the biomarker and the clinical endpoint, attenuating the biomarker effect. For biomarkers developed to detect and quantify a specific pathology, such as amyloid deposition, it is important to have a control group not harboring such brain changes (e.g., having negative amyloid positron emission tomography (PET) scans or CSF A $\beta$ 42/40 ratio). However, one should typically avoid to select a “super healthy” control population that not only differs substantially from subjects with neurodegenerative diseases but also from the general population without such diseases, in order to avoid bias of the results.

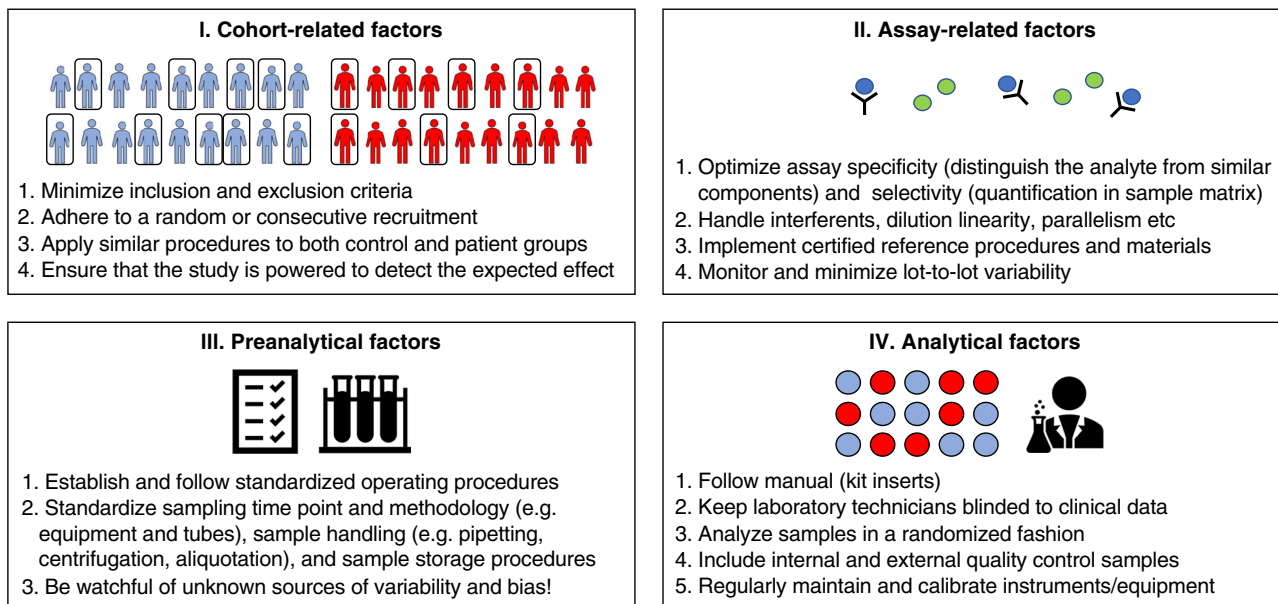
Third, study procedures should not differ between patients and controls. If patients and controls are recruited at different centers or during different time periods, known or unknown systematic differences in the procedures may lead to a bias in form of biomarker findings that are erroneously interpreted as disease-related. Optimally all groups are recruited at the same sites, during the same period using the same standardized operating procedures.

Fourth, confounding or modifying factors may impact reproducibility. This may include demographics, genetic factors, drugs, kidney and liver function, or presence of co-morbidities. For example, in highly educated patients, “cognitive reserve” may attenuate the relationship between a biomarker and clinical diagnosis. Many late-onset clinically diagnosed AD cases also have combinations of  $\alpha$ -synuclein pathology, TDP-43 deposits, microvascular changes and hippocampal sclerosis on top of plaque and tangle pathology<sup>26</sup>, which may attenuate associations between biomarkers and the studied disease (since different pathologies may affect biomarkers in different ways, e.g., white matter lesions may reduce CSF levels of several biomarkers<sup>27</sup>).

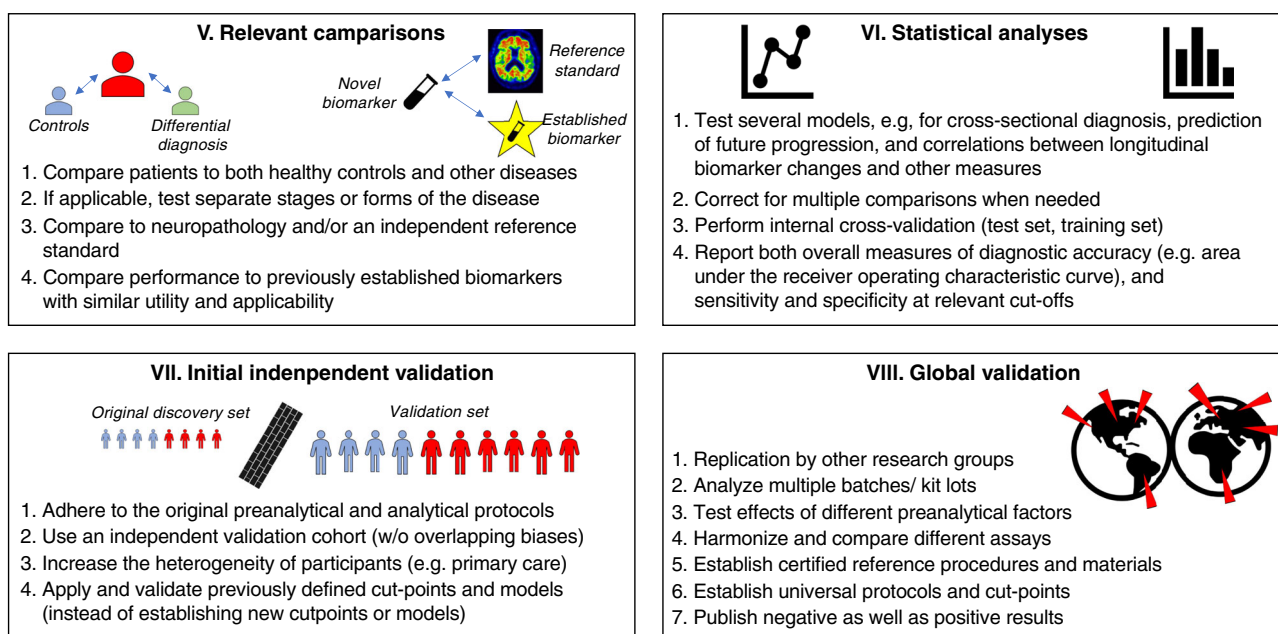
Lastly, we think that it is important to pre-register the cohort study (e.g., at [clinicaltrials.gov](http://clinicaltrials.gov)), both to increase transparency of the biomarker study, and to determine outcomes before data collection and analysis, to reduce selective reporting and p-hacking.

**Assay-related factors.** Analytical methods (assays) can be subject to both random and systematic measurement errors, which can all impact reproducibility of findings. The procedures to control assay-related factors are complex and technical, but necessary for the introduction of assays in routine diagnostic use in Clinical Chemistry. We will only briefly highlight a few aspects here.

For ligand binding assays (e.g., immunoassays), two key properties are specificity and selectivity<sup>28</sup>. The assay specificity refers to how well the assay (with its antibodies and other components) can distinguish between its intended analyte and other structurally similar components. Poor specificity leads to a systematic overestimation of biomarker levels. Assay specificity can be tested by evaluating the assay for cross-reactivity using similar material, e.g., a protein homolog such as A $\beta$ 40 in an A $\beta$ 42 assay or the medium or heavy subunits of neurofilament in an NFL assay, which it should not react towards. The assay selectivity refers to how well the assay measures the analyte in the sample matrix, with presence of other biological components. Selectivity can be tested with different spike-recovery experiments, where a known quantity of the measured analyte is added to a sample, and the assays’ ability to recover the known quantity is evaluated. One caveat is that for protein biomarkers, the spiking material is often recombinant and may differ from the endogenous form of the biomarker, e.g., by lacking post-translational modifications, truncations etc. Hypothetically, this



**Fig. 1 Finding reproducible biomarkers, from cohort recruitment to biomarker measurements.** A flowchart highlighting key points described in the review, with examples of key factors that influence reproducibility from the first design of the study, up to biomarker measurements.



**Fig. 2 Finding reproducible biomarkers, from relevant comparisons to validation.** A flowchart highlighting key points described in the review, with examples of key factors that influence reproducibility from including relevant comparisons to global validation.

may over- or underestimate the accuracy of the assay compared to when measuring the biomarker in its endogenous form.

Other key properties that needs to be controlled for in assay development and maintenance include dilution linearity (measurement levels should be proportional to sample dilution), and parallelism (standard reference and serially diluted sample curves should be parallel)<sup>29</sup>. In individual samples, assays may be sensitive to interferents such as lipids, hemolysate or heterophilic antibodies, which may result in both over- or underestimation of the measured concentrations of the analyte<sup>30</sup>.

Assay-related factors are not only important to control when developing and launching a new assay, but also over time when using an assay in research or clinical practice. As known or

unknown changes occur in production procedures or reagents, biomarker measurements may be affected by lot-to-lot variability. This can introduce differences both between studies, and within studies (e.g., if several lots of analytical kits are used within a large study). It is the responsibility of kit manufacturers and vendors to minimize lot-to-lot variability. Batch-bridging at individual labs can track this potential issue<sup>31</sup>, which can be controlled for by rejection of a batch or possibly by adjustments of calibrator levels.

If possible, novel assays should always be compared with certified reference procedures, using certified reference materials<sup>32–34</sup>. Unfortunately, reference measurement procedures (“gold standard” methods) or certified reference materials (“gold

standard” samples) are not available for most neurodegeneration biomarkers, with the exception of CSF A $\beta$ 42<sup>35</sup>.

**Pre-analytical factors.** Even with a validated assay, many factors can affect biomarker measurements even before the analytical phase starts. Examples of these include time-of-day for sampling, the technique used by the doctor/nurse/staff for phlebotomy or lumbar puncture, tube-related factors, and pure errors in tube handling and labeling<sup>36</sup>.

Some biomarkers may vary due to normal physiological processes, or in response to diet, stress factors, or health issues. For example, plasma T-tau (but not P-tau) may be affected by sleep loss<sup>37</sup> (which might contribute to the poor reproducibility of plasma T-tau as a biomarker, described below). Studies of test-retest variability with repeated measures (after hours, days or weeks) may quantify physiological, changes of the biomarker over time. One study tested variability over eight weeks for CSF A $\beta$ 38, A $\beta$ 40, A $\beta$ 42, T-tau, NFL, and panels of inflammatory, neurovascular, and metabolic biomarkers<sup>38</sup>. The variability was acceptable for most of biomarkers, but a few inflammatory markers showed instability over time, making them less suitable as CSF biomarkers (SAA, CRP, and IL-10). Some biomarkers may also have a circadian, cyclic, variability, where levels fluctuate over the course of a 24-hour cycle. This is difficult to study for CSF biomarkers, since the sampling procedure itself, with repeated CSF collection through an indwelling lumbar catheter, affects CSF turn-over and biomarker concentrations much more clearly than diurnal changes<sup>39</sup>. To minimize the influence of potential biotemporal variability, it is recommended that sampling is done at a consistent time of the day. However, one study even suggested that some biomarkers (CSF A $\beta$ 40, A $\beta$ 42, T-tau, and P-tau were tested) may fluctuate over the year<sup>40</sup>, which (if reproduced) could be an additional source of variability.

Beyond factors that temporarily offset biomarker levels, some biomarkers may have a large normal physiological range within the population, which could impact the likelihood that a biomarker reaches a threshold for positivity. For example, a person with a slightly higher than normal release of A $\beta$  peptides into CSF is more likely to be classified as normal for CSF A $\beta$ 42 compared to a person with slightly lower than normal release of A $\beta$  peptides. Most of the A $\beta$  peptides released into CSF are of the shorter variant, ending at the amino acid position 40 (A $\beta$ 40), which is not affected by AD. The associations for CSF A $\beta$ 42 with AD are, therefore, improved when adjusting for CSF A $\beta$ 40, usually in the form of the CSF A $\beta$ 42/A $\beta$ 40 ratio<sup>41</sup>, which adjusts for between-individual differences in overall (with A $\beta$ 40 serving as a proxy for total) A $\beta$  peptide levels in CSF, and potentially also for both within-individual and between-individual differences in production and clearance of CSF<sup>42</sup>. For this reason, we include CSF A $\beta$ 42/A $\beta$ 40 rather than just CSF A $\beta$ 42 in the clinical work-up of dementia patients at our centers. The usage of reference peptides may also improve the accuracy for other AD biomarkers in CSF and blood, but more work is needed in this area.

Beyond these biological factors, sampling and storage procedures may also affect biomarker levels, and thereby influence reproducibility for both blood-based and CSF-based biomarker<sup>43,44</sup>. For example, CSF A $\beta$ 42 is hydrophobic and especially sensitive for variations in pre-analytical handling, and may be affected by the type of tubes used for CSF sampling and storage, as well as freeze-thaw procedures<sup>45</sup>. As another example of a pre-analytical storage effect, a 12.5 kDa C-terminally cleaved fragment of cystatin C was proposed as a promising CSF biomarker for multiple sclerosis (MS), reaching 100% specificity for other neurological disorders<sup>46</sup>. However, a paper by another research group could not replicate the finding<sup>47</sup> and it also was

noted that this protein cleavage occurs as a storage artifact when samples are being stored at  $-20^{\circ}\text{C}$  instead of  $-80^{\circ}\text{C}$  (which had been shown before as well<sup>48</sup>).

To minimize all pre-analytical factors, it is important that the pre-analytical protocol is identical for patient and control groups, and across participating centers in multi-center studies. Further, it is important that systematic experiments are performed to evaluate the effects of different pre-analytical variables on the biomarker of interest, even though such publication seldom are published in higher impact journals. See e.g., ref. <sup>9</sup> for a standardized protocol applied in clinical practice to improve the performance of CSF A $\beta$ 42.

**Analytical factors.** Several analytical factors (beyond what we describe above under assay-related factors) may increase imprecision or bias and thereby contribute to poor reproducibility of studies.

Broadly, analytical imprecision (variability) can be described in terms of within-lab and between-lab variability. Within-lab variability can be further divided into intra-assay variability (precision in the same run, which can be monitored by measuring duplicates of all or selected samples), and inter-assay variability (precision across different assay runs, which can be monitored by running aliquots of the same internal control samples in every run). To minimize these sources of variability, laboratory staff must follow analytical protocols carefully, and have sufficient control systems to detect and quantify variability. Significant between-lab, inter-assay, or lot-to-lot variability makes it difficult to reproduce results at specific cut-points or decision thresholds, and is a hurdle towards widespread use of the biomarker<sup>49</sup>. Significant variability also makes it difficult to use longitudinal rates of change as indicators of incipient pathology. Longitudinal testing of biomarkers with low variability may provide meaningful information, even before critical thresholds are reached, as shown for A $\beta$  PET imaging<sup>50</sup>. An international quality control program has been established to monitor measurement variability for different assays and platforms for key CSF biomarkers for AD<sup>49,51</sup>. This program has demonstrated that variability can be reduced by transferring assays from manual to fully automated methods, as shown for the CSF AD biomarkers A $\beta$ 42<sup>52</sup>, T-tau and P-tau<sup>53</sup> (see updated info at <http://www.neurochem.gu.se/theAlzAssQCprogram>).

To minimize analytical bias, technicians running the assays must be blinded to all clinical data. One consequence of this is that all samples become randomized, to make sure that patients and controls are spread out over the plates. In the best scenario the whole lab analyzing the samples should not have access to any clinical information until the data is finalized.

**Statistical methods associated with poor reproducibility.** A typical situation with a high risk for false-positive findings is when a large number of biomarkers are tested without a grounded theory. This is especially common in “omics”-studies (e.g., proteomics or metabolomics) with hundreds or thousands of molecules. A statistical adjustment for multiple comparisons is always required at some step for these studies, but if two cohorts are available, we can accept false-positive findings in the discovery cohort and apply a strict adjustment when testing the biomarker in the validation cohort. We also note the risk for overcorrection (leading to “type II error”) when several findings have nominal uncorrected significance close to an a priori threshold, and they are all ruled out as non-significant after correction for multiple comparisons.

A special scenario with high risk for poor reproducibility is when multiple biomarkers are tested simultaneously in a panel

selected from a large set of originally available biomarkers. Such panels form complex models, which are liable to overfitting (thus representing noise in the data rather than true patterns of interest). Again the most robust strategy to avoid false-positive findings is to use a large external validation cohort to test the specific biomarker panel identified in the discovery cohort. When only one cohort is available, validation methods should be applied within the single cohort, as explained below.

To elucidate if biomarker changes are specific to a certain neurodegenerative disease or if alterations are non-specific in response to brain injury, we suggest that studies include comparisons between as many different diagnostic groups as possible. For example, a study of a novel AD biomarker should optimally compare A $\beta$ -positive AD dementia, other neurodegenerative diseases, A $\beta$ -positive MCI, A $\beta$ -negative MCI, A $\beta$ -positive cognitively unimpaired controls, and A $\beta$ -negative cognitively unimpaired controls<sup>54</sup>. If the biomarker is specific for AD, we expect alterations both in A $\beta$ -positive AD dementia and MCI (and potentially also in A $\beta$ -positive cognitively unimpaired controls if changes come early). If the biomarker is altered also in other neurodegenerative diseases and A $\beta$ -negative MCI and dementia, it indicates that it responds non-specifically to brain injury. If a biomarker is altered only in A $\beta$ -positive MCI but not in AD dementia, we consider it likely to be a false-positive finding (although a transient peak in biomarker levels in earlier stages is possible and has been suggested<sup>55</sup>, our experience is that this has not been reproduced for a biomarker of neurodegeneration).

We also suggest studies to not only report a measure of overall performance to separate between groups, e.g., AUC, which may give a “numerical” impression of high performance, since it cannot be lower than 0.5, but also sensitivities and specificities (and possibly positive and negative predictive values depending on the generalizability of the disease prevalence in the present cohort) preferably using cut-offs defined in another population. This is since sensitivity and specificity figures are important for the clinical application, either identifying as many as possible with a disease (high sensitivity) but accepting false positives, or ruling out a disease (high specificity). It is also valuable to show cut-points defined using different methods, including for example at optimal specificity (such as the mean plus two standard deviations in A $\beta$ -negative cognitively unimpaired individuals for an AD-biomarker), a combined optimal sensitivity and specificity (Youden index), or a “natural” cut-points identified by Gaussian mixture models, which can provide robust cut-points especially for biomarkers with a bimodal distribution<sup>56–58</sup>. For clinical chemistry tests used in clinical practice, reference intervals are most often based on findings in healthy individuals, and for biomarkers where changes in both directions are clinically relevant (such as plasma glucose) the cut-offs corresponds to the central 95% of the distribution, or for biomarkers where only a change in one direction (either an increase or a decrease) is clinically relevant (such as increased serum Troponin-T) the single cut-off corresponds to the upper 95% of the distribution<sup>59</sup>.

**Validation methods.** All studies with novel or unexpected biomarker results should have a validation section. The strongest possible validation is to demonstrate robustness of results in a separate validation cohort. Robustness should be shown both for the overall continuous associations between the biomarker and the main clinical endpoints, and if possible for performance at specific cut-points (e.g., for classification). The validation cohort must be sufficiently large to be powered to detect the biomarker effect found in the discovery cohort.

If an independent validation cohort is not available, validation is often done within the original cohort<sup>60</sup>. One possibility is to

split the available sample into a training set (e.g., 80%, but may be higher if the sample size is small) where the biomarker is “trained” to predict the outcome, and a test set (e.g., 20%), for validation. The partitioning of the cohort should be done before any analyses, to avoid leakage of information from the test to the training set. If researchers first find effects for a biomarker in the whole population, and then post hoc perform a training/test split, the risk is high for overoptimistic estimates of performance. A preferable alternative to a training/test split is to perform *k*-fold cross-validation (CV). The data is portioned into *k* bins of equal size (usually *k* = 10, but may be lower with a small sample size). In an iterative procedure, the biomarker model (for example a logistic regression model for binary classification) is trained sequentially in all bins except one, and evaluated in the remaining single bin. The result is a string of *k* measures of performance. This is preferable to a simple training/test set split, because it reduces the impact of the random grouping of subjects into one of the sets, and gives a distribution of the test effect, rather than just one test performance value. The robustness of the analysis can be further increased by repeating the *k*-fold CV (e.g., a five times iterated 10-fold CV). Confounding factors can be balanced across bins. Some statistical models include “hyperparameters” (e.g., the regularization constant in LASSO regression, or the number of clusters in *k*-means clustering). Hyperparameters should be estimated separately from the model performance. For this, nested CV may be used, with two layers of CV (“inner” and “outer”), where the hyperparameters are tuned in one layer, and the model performance is estimated in the second layer. However, if the discovery and validation cohorts are based just on dividing the total cohort into two, possible systematic bias between patients and controls in terms of, e.g., differences in pre-analytical procedures or cohort specific biases will remain. Consequently, there are previous examples of convincing internal cross-validations that have failed when replicated in an independent cohort<sup>15,17</sup>. Therefore, the discovery and validation cohorts need to be independent and come from different studies.

One type of desirable (but rarely available) validation, is towards neuropathological data<sup>61</sup>. This is valuable since clinical diagnosis is imperfect, with about 70% sensitivity and specificity<sup>62</sup>. However, neuropathological validation also has its caveats. First, the size of cohorts with autopsy data are often small. Second, there is also almost always a lag between fluid biomarker sampling and death, which may underestimate associations between biomarker levels and brain changes that likely continue to develop, or other types of pathology (e.g., ischemic lesions) may appear. Third, most neuropathological examinations does not include detailed quantification of brain changes across the whole brain, but is focused on particular tissue sections, which may be more or less representative of the pathologies that produce altered biomarker levels. Moreover validation against neuropathology most often restricts the analysis to generating data on biomarker performance in the very latest stages of the disease. A proxy for pathological validation of A $\beta$  and tau pathology may in some cases be molecular imaging, using well-validated PET tracers<sup>63</sup>.

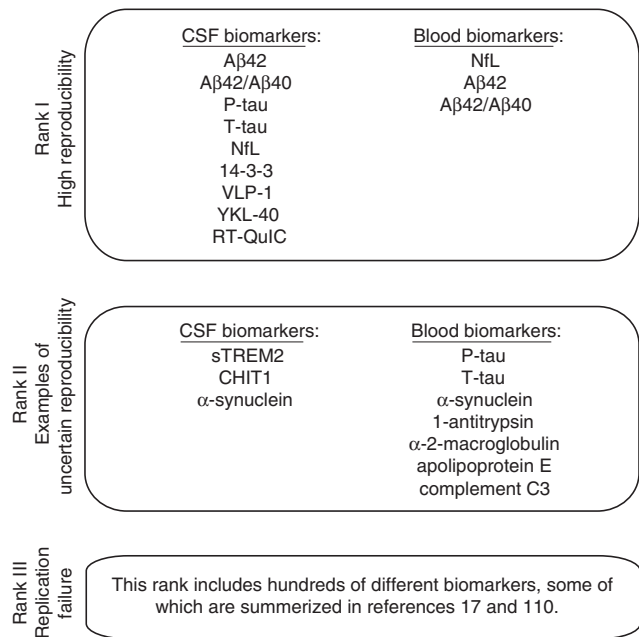
Finally, validation of a biomarker with a secondary, independent assay, increases the chance that the biomarker signal is true, i.e., that the assay actually measures what it is intended to measure. Correlations between different assays, or assay formats, are typically evaluated in Round Robin or Commutability studies, in which aliquots of the same samples are analyzed using several different analytical methods<sup>32</sup>. There are several examples of biomarkers that correlate poorly when measured with different assays<sup>17</sup>, making links to the underlying biological processes unclear. Associations should also be tested towards previously validated biomarkers of pathology (including for example CSF A $\beta$ 42/A $\beta$ 40 ratio or A $\beta$  PET imaging for A $\beta$  pathology).

Researchers and Editors could work together to define community standards for the type of validation that is necessary for biomarkers detected through unbiased methods. Completely novel exploratory markers should have extensive validation (including validation in an independent cohort, which should be large enough to have sufficient power). For more established biomarkers, an internal validation procedure may be sufficient. Complex models, especially including panels of biomarkers, or biomarkers used together with demographic factors, need more extensive validation than a single a priori defined biomarker. Single biomarkers selected post hoc from a panel of many exploratory biomarkers needs careful validation, specifically that the exact composition of the panel is set in the discovery cohort, and optimally that the performance of the same panel is evaluated in a fully independent validation cohort. Extensive validation is necessary if the cohort is at risk for bias. Finally, to our knowledge, almost all biomarkers that have been reproduced sufficiently to be used in clinical practice (e.g., CSF A $\beta$ 42 and P-tau for AD, and RT-QuIC for CJD) have been discovered with a grounded theory, based on a clear hypothesis about disease mechanisms. Despite decades of research, unbiased methods, including proteomics and metabolomics have still not resulted in biomarkers that have been sufficiently reproduced for use in clinical practice. We therefore argue that biomarkers detected through unbiased methods should have extensive validation before publication.

**Levels of reproducibility.** We have reviewed the literature for reproducibility of fluid biomarkers in neurodegenerative diseases, including AD, Parkinson's disease (PD) and related conditions, frontotemporal lobe dementia, and motor neuron disease. In summary, we found that a few biomarkers have had very high reproducibility with almost unanimously converging results. These biomarkers are summarized below ("Rank I"). There is a second large group of biomarkers with variable results, which may be considered to have uncertain reproducibility. We present a few examples of such biomarkers, with suggested explanations for the variable results ("Rank II"). Finally, for many biomarkers replication has been attempted and failed. Again, we present a few examples, together with a discussion about the reasons for the failed replications ("Rank III"). These are summarized in Fig. 3.

**Rank I: high reproducibility.** A few fluid biomarkers have been robustly associated with neurodegenerative diseases. These biomarkers are often incorporated in clinical trials, either at study inclusion to enrich for participants with AD pathology, or as exploratory secondary outcomes. Some of these biomarkers are also used in clinical practice for dementia work-up.

CSF A $\beta$ 42, T-tau, and P-tau are altered in AD, as reviewed extensively before (e.g., in the database AlzBiomarker<sup>25,64</sup>). CSF T-tau has been tested in AD dementia versus controls in at least 188 studies, including over 12000 patients and over 8000 controls. All but two of these found significantly higher CSF T-tau in AD, while the two negative studies found non-significant increases. However, CSF T-tau is also increased non-specifically due to brain injury in other neurological diseases. CSF P-tau (mainly using the phosphorylation variant P-tau181) has been tested in AD dementia versus controls in at least 116 studies. All but three of these found significantly higher CSF P-tau in AD (the three negative studies found non-significant increases). CSF A $\beta$ 42 has been tested in AD dementia versus controls in at least 168 studies. All but seven of these found reduced levels in AD; and only one of those seven studies found a significant increase in AD (see <https://www.alzforum.org/alzbiomarker>). Several studies have also found that CSF A $\beta$ 42, T-tau, and P-tau are altered prior to dementia in



**Fig. 3 Ranks of reproducibility.** A brief summary with examples of fluid biomarkers for neurodegenerative diseases, by their level of reproducibility.

AD, demonstrating that the biomarker changes are robust at all clinically relevant stages. The findings have also been replicated with several different independent assays.

Plasma (or serum) biomarkers for A $\beta$  have developed significantly over the years. The first generation of studies on plasma A $\beta$ 42 (and the A $\beta$ 42/40 ratio) showed no change in clinically diagnosed AD cases versus cognitively unimpaired elderly (for review see<sup>25</sup>). However, newer studies with improved assays have repeatedly shown that plasma or serum measures of A $\beta$ 42 (or A $\beta$ 42/A $\beta$ 40) are altered in AD, although typically with much lower effect size than corresponding CSF biomarkers (see e.g., refs. 54,65–69). This may be due to that ultrasensitive Simoa immunoassay or immunoprecipitation combined with mass spectrometry have better analytical performance than the ELISA or Lumindex methods used in the first generation of studies. However, a contributing factor is likely that in the older studies, a proportion of clinically diagnosed AD cases may have been misdiagnosed and some of the unimpaired elderly may have had clinically silent AD pathology, thereby limiting performance of plasma A $\beta$  biomarkers. In contrast, more recent studies dichotomize AD patients and controls based on A $\beta$  PET positivity and negativity (and compare A $\beta$  PET-positive AD with PET-negative controls), which improves the possibility to find a high performance of any candidate AD biomarker.

Several proteins potentially related to synaptic or neuronal injury have been tested as biomarkers. One example is the intracellular structural protein NFL, which is increased in CSF and blood after neuronal injury, and therefore increased non-specifically in many neurological diseases (as shown for example in a large meta-analysis on 8727 patients with different diseases and 1332 healthy controls<sup>70</sup>). Very high NFL levels are seen in conditions with rapidly progressive neurodegenerative diseases, such as CJD (at least six studies found this, see e.g., ref. 71) and motor neuron disease (a meta-analysis showed increased levels in 15/16 studies<sup>72</sup>), but levels are also increased in conditions with more slowly progressive injury, such as frontotemporal lobe dementia (a meta-analysis showed increased levels in 26/26 studies<sup>72</sup>), atypical parkinsonian disorders (at least four studies found this, e.g., refs. 70,73), and vascular dementia (six

studies are summarized in a meta-analysis<sup>74</sup>). CSF and blood levels of NFL are also increased in AD (tested in at least 29 studies, see refs. <sup>64,72,75–77</sup>). Another non-specific injury marker is CSF 14-3-3, which has traditionally primarily been used in work-up of CJD<sup>78</sup> (a meta-analysis summarized results from 13 CJD studies<sup>79</sup>). Another injury-related marker, which potentially is more specific to AD, is the postsynaptic protein neurogranin<sup>80</sup> (increased CSF neurogranin levels in AD were seen in 10/10 studies included in a meta-analysis, using at least three different assays<sup>64</sup>). CSF neurogranin is also increased in MCI due to AD compared to other MCI, and has been validated for AD neuropathology using autopsy data<sup>81</sup>. Another example is the neuronal protein VLP-1, which has been shown to be increased in CSF and blood after brain injury<sup>82</sup> (increased CSF VLP-1 levels in AD were seen in 8/8 studies included in a meta-analysis<sup>64</sup>).

A few markers of inflammation and astroglial activation have also been reproduced. The strongest data appears to exist for YKL-40, which is a glycoprotein that can be released by several different cell-types in the body, but in the CNS it is mainly related to activated astrocytes<sup>83</sup>. CSF YKL-40 is increased in several neurodegenerative diseases. At least 14 studies have compared AD dementia versus controls, and all but two found increased levels in AD, but the fold change in AD versus controls in minor<sup>64</sup>. CSF YKL-40 was also increased in MCI due to AD compared to other MCI patients.

With regards to specific markers for non-AD pathologies, very few reproduced biomarkers have been reported, but aggregation assays for prion protein using RT-QuIC assays on CSF have high diagnostic accuracy for CJD (at least six different studies have found this, summarized in ref. <sup>79</sup>) and are used in clinical practice for CJD diagnosis in several countries<sup>5–7</sup> (e.g., US, UK, and Sweden). Similarly, RT-QuIC assays for CSF  $\alpha$ -synuclein have been shown in several studies to have high sensitivity and specificity for PD and dementia with Lewy bodies (at least eight studies, e.g., refs. <sup>84,85</sup>).

**Rank II: uncertain reproducibility.** Many biomarkers for neurodegenerative diseases have either had few attempts of replication, or the evidence has been conflicting, without consensus on how to optimally use them in practice. These biomarkers may potentially be useful in the future, possibly after further standardization work. We discuss a few examples here.

One group of promising plasma-based biomarkers are P-tau with different phosphorylation variants (threonine 181 and 217)<sup>86–90</sup>. These correlate well with CSF P-tau and Tau PET, can differentiate AD from other neurodegenerative diseases, and predict future conversion to AD dementia. The results from the published studies on plasma P-tau are very encouraging, but we believe that the number of studies is still too small for plasma P-tau to qualify for Rank I.

sTREM-2 is released from microglia. Most studies (4/6 studies in a meta-analysis<sup>64</sup>) show increased CSF sTREM-2 in AD versus controls. Levels are also slightly increased in AD when subjects are enriched for AD pathology using other biomarker data<sup>91</sup>. Several other inflammatory markers in blood or CSF have also shown variable, but overall slight associations with AD<sup>92</sup>. One example is CHIT1, which has been found to be increased in CSF in both AD<sup>93,94</sup> and other neurodegenerative diseases<sup>95</sup>. These results demonstrate that although there can be a statistically significant difference in a biomarker in research studies, the degree of change can be too low to fill the demands for a useful diagnostic test in the clinic. Such biomarkers roles are mainly limited to pointing at involvement of inflammatory processes in a complex multifactorial disease. With specific treatments, for

example directed against microglia, some of these markers could also be explored as outcome measures.

CSF  $\alpha$ -synuclein has been tested as a biomarker for PD, with varying results. In a meta-analysis of 34 studies, CSF  $\alpha$ -synuclein was slightly reduced in PD, but the diagnostic performance was considered too poor for clinical practice, and most studies were at risk for bias<sup>96</sup>. Results for CSF  $\alpha$ -synuclein in AD dementia have included both reductions, increases, or no effects in AD compared to controls<sup>64</sup>. One possible explanation for the poor reproducibility may be that CSF  $\alpha$ -synuclein can be affected by multiple pathological processes. Hypothetically, the presence of Lewy bodies (a common co-pathology in AD patients) could reduce CSF  $\alpha$ -synuclein (as in PD), while the presence of synaptic or neuronal degeneration could increase CSF  $\alpha$ -synuclein as an injury response<sup>97</sup>. A pre-analytical factor affecting CSF  $\alpha$ -synuclein is that it is sensitive to blood-contamination, which gives false high levels<sup>98</sup>. The inconsistent results, the unclear biological role, and the pre-analytical variability may make it difficult to apply CSF  $\alpha$ -synuclein as a biomarker. However, we note that CSF  $\alpha$ -synuclein may potentially be useful to separate AD from dementia with Lewy bodies<sup>99</sup>. There are also a few studies on  $\alpha$ -synuclein in blood with varying results<sup>100–103</sup>.

Plasma or serum levels of T-tau have been measured in AD in several studies, with conflicting results. Most studies have found slightly increased levels in AD (with varying effect sizes<sup>54,104–106</sup>), but there are also studies without difference between AD and controls<sup>107</sup>, or with lower levels in AD<sup>108</sup>. One explanation for the poor reproducibility is that the results may be platform- or assay-dependent. Another possibility is that T-tau in blood may have rapid kinetics and is degraded, or is sensitive to processes beyond neurodegeneration<sup>37</sup>. In our opinion, this argues against using T-tau in blood as an AD biomarker.

Several other blood-based biomarkers have also shown slight associations with different AD-phenotypes (e.g., clinical diagnosis, brain A $\beta$  burden, or atrophy measures) in several studies, for example 1-antitrypsin,  $\alpha$ -2-macroglobulin, apolipoprotein E and complement C3<sup>109,110</sup>. Such reproducible associations all point to possible involvement of these proteins in the pathogenesis of AD, although the associations are generally too weak to provide clinical utility.

**Rank III: replication failure.** Many biomarkers for neurodegenerative diseases have failed replication. Here we can only discuss a few examples. In general, biomarkers discovered through exploratory methods, e.g., proteomics studies without a grounded hypothesis about links to specific disease mechanisms, have low reproducibility. One study attempted to replicate associations with AD for 94 candidate plasma proteins that had been described previously in at least one (out of a total of 21) discovery/panel-based studies (each protein had most often only been associated with AD in one previous study)<sup>110</sup>. Only nine proteins had significant effects in the validation cohort, meaning that they were associated with at least one of several possible AD phenotypes (including diagnosis, cognitive measures, or measures of brain structure). However, we note that it was rare for a protein to be associated with more than one AD phenotype, and some associations were even in the opposite direction compared to the original studies. Another study (on the AIBL cohort) aimed to validate blood-based proteins related to A $\beta$  PET positivity<sup>17</sup>. Thirty-five proteins that had been described in at least one of four previous proteomics studies (including two AIBL studies) were tested. Only two proteins were associated with A $\beta$  PET in the validation study. The same study also highlighted that different multiplex proteomics platforms (in this case, SOMAscan and Myriad's Rules-Based Medicine Multi-analytes Profile) may give

different results for putatively similar biomarkers. Note that panel-based biomarker discovery work still offers powerful ways to study a broad range of biological processes in neurodegenerative diseases, but findings need careful validation to avoid poor reproducibility.

Studies that build a classifier from a panel of biomarkers also have low reproducibility. One study published in 2007 used a proteomics approach for AD diagnosis<sup>111</sup>. Out of 120 available proteins, 18 proteins were selected (several of these were involved in immune response). Used together, the proteins had an overall accuracy of 89% for AD dementia versus controls, and successfully identified most MCI patients who later converted to AD dementia. The study included 259 individuals, recruited from seven centers. This study has been cited by over 700 papers (according to [www.nature.com](http://www.nature.com), October 2020). In 2012, a replication attempt was conducted in a cohort of controls, AD patients, and other patients (total  $N = 433$ )<sup>112</sup>. The performance of the tested biomarkers was considerably lower in the replication study than in the original study (AUC 0.63). For several of the tested proteins, the associations were also reversed in the second study (e.g., lower plasma levels in AD in the original study, but higher in AD in the replication study).

One study published in 2014 used a plasma lipidomics approach to detect preclinical AD<sup>60</sup>. A panel of ten metabolites had an impressive performance for AD diagnosis (AUC 0.92). However, in 2016 another group presented a failed attempt to replicate the performance<sup>113</sup>. In a subsequent debate, the authors of the original study underscored differences between the original study and the replication to explain the discrepancies. They pointed to differences in sample matrix (plasma versus serum), sample storage time (longer in the replication study), and frequency of clinical follow-up for endpoints (lower in the replication study)<sup>114</sup>. The authors of the replication study responded that they only found minor differences between serum and plasma, that they had included a second validation cohort to have more similar storage times as the original study, and that they considered the follow-up designs to be comparable between the studies<sup>115</sup>. They also underscored that the original study used a small sample for validation (21 patients and 20 controls), which they thought had overinflated the effect size. There is still no replication study published having validated the results on the diagnostic performance for AD for these ten metabolites.

### Future outlook

As the fluid biomarker field advances, we will see more novel biomarkers that are reproducible between different cohorts and methods of quantification. To guide both clinicians and policymakers it will be important to compare the value of these novel biomarkers to already established state-of-the-art diagnostic methods. For example, if a new CSF biomarker for prediction of development of AD dementia is found it needs to be evaluated against already established biomarker (A $\beta$ 42/A $\beta$ 40, P-tau, T-tau), but also magnetic resonance imaging measures such as cortical thickness of the medial temporal lobe and memory function. Similarly, if a new blood-based biomarker for prediction of ALS is discovered it should be compared to plasma NFL<sup>116</sup>.

It will also be important to show that biomarkers are reproducible not only within a highly specialized setting (tertiary or secondary referral center), but also in a primary care setting<sup>117,118</sup>. Poor reproducibility in primary care may be partly due to differences in pre-analytical factors, which may be more difficult to standardize in a primary care setting than in a specialized setting. Poor reproducibility could also be due to cohort-related factors, since the demographics of a primary care setting

may be different than in the often highly selected population in a specialized setting (which may affect the relationship between a biomarker and an underlying pathology). For practical purposes, blood-based biomarkers will be more relevant than CSF-based biomarkers for primary care. Reproducibility for primary care will therefore require strict control of factors that can influence blood measurements. Low analytical variability will be key especially to detect subtle longitudinal changes in blood-based biomarkers for aggregation of pathologies. Despite the many obstacles and challenges in validating biomarkers in primary care, we greatly encourage this type of validation since a biomarker that passes this replication test most likely has proved to have a high level of robustness in terms of the influence from pre-analytical factors, co-morbidities and a variety of demographic factors. Such a successful validation also potentially makes the biomarker accessible for a much larger group of people.

**Summary recommendations for high-quality publications.** All stakeholders need to define what is needed in terms of validation and reporting to improve reproducibility in biomarker research (a summary of our recommendations is outlined in Figs. 1 and 2). Authors should to the best of their capacity provide comprehensive and detailed reports of their findings. Authors should aim to describe as many potential usages as possible for a novel biomarker to demonstrate convergence of findings. The most common applications are: (1) cross-sectional identification of diagnosis or another disease feature, (2) forward-looking prognostication of a feature after diagnosis is established, and (3) longitudinal biomarker measurements to track disease changes over time. A complex biomarker study may include all of these different aspects, with multiple cross-sectional comparisons and longitudinal prediction of several different variables. Naturally, an isolated positive finding together with several clearly non-significant results is a warning that the finding may be a false-positive result with low reproducibility. In contrast, multiple, convergent and logical associations for a biomarker are very encouraging. Besides testing multiple aspects of the biomarker, another hallmark of a high-quality biomarker publication is that it includes a sufficient validation section, the best being validation of results in an independent cohort of patients and controls. Another hallmark is that the biomarker is thoroughly compared with relevant state-of-the-art methods. A high-quality study should also report as much data as possible on variability of the biomarker (including analytical variability and biological variability). Finally, we encourage editors to also accept well-performed negative (failed) replications (both in the same or another journal) to counteract positive publication bias. To minimize publication bias, journals may also consider accepting papers as “registered reports”, where a proposed set of analyses are reviewed and can be provisionally accepted for publication before data collection has begun.

It is our hope that the recommendations in this perspective may help to improve reproducibility of future research on fluid biomarkers for neurodegenerative diseases.

Received: 6 May 2020; Accepted: 9 November 2020;

Published online: 07 December 2020

### References

1. Jack, C. R. et al. A/T/N: an unbiased descriptive classification scheme for Alzheimer disease biomarkers. *Neurology* **87**, 539–547 (2016).
2. Blennow, K., Hampel, H., Weiner, M. & Zetterberg, H. Cerebrospinal fluid and plasma biomarkers in Alzheimer disease. *Nat. Rev. Neurol.* **6**, 131–144 (2010).
3. Jack, C. R. et al. NIA-AA research framework: toward a biological definition of Alzheimer's disease. *Alzheimers Dement.* **14**, 535–562 (2018). **This is a**



- commonly used framework for Alzheimer's disease research, which makes biomarker integral to disease definitions.**
- Shaw, L. M. et al. Appropriate use criteria for lumbar puncture and cerebrospinal fluid testing in the diagnosis of Alzheimer's disease. *Alzheimers Dement.* **14**, 1505–1521 (2018).
  - Centers for Disease Control and Prevention, U.S. Department of Health & Human Services. *CDC's Diagnostic Criteria for Creutzfeldt-Jakob Disease (CJD)* (Centers for Disease Control and Prevention, U.S. Department of Health & Human Services, 2018).
  - The National Creutzfeldt-Jakob Disease Research & Surveillance Unit. *Diagnostic Criteria for Human Prion Disease*. Version 4 (University of Edinburgh, 2017). <https://www.cjd.ed.ac.uk/sites/default/files/NCJDRSU%20surveillance%20protocol-april%202017%20rev2.pdf>.
  - Green, A. J. E. RT-QuIC: a new test for sporadic CJD. *Pr. Neurol.* **19**, 49–55 (2019).
  - Khalil, M. et al. Neurofilaments as biomarkers in neurological disorders. *Nat. Rev. Neurol.* **14**, 577–589 (2018).
  - Rosengren, L. E., Karlsson, J. E., Karlsson, J. O., Persson, L. I. & Wikkelso, C. Patients with amyotrophic lateral sclerosis and other neurodegenerative diseases have increased levels of neurofilament protein in CSF. *J. Neurochem.* **67**, 2013–2018 (1996).
  - Abbasi, J. Promising results in 18-month analysis of Alzheimer drug candidate. *JAMA* **320**, 965–965 (2018).
  - Biogen Plans Regulatory Filing for Aducanumab in Alzheimer's Disease Based on New Analysis of Larger Dataset from Phase 3 Studies*. (2019) Biogen <https://investors.biogen.com/news-releases/news-release-details/biogen-plans-regulatory-filing-aducanumab-alzheimers-disease>.
  - Sevigny, J. et al. The antibody aducanumab reduces A $\beta$  plaques in Alzheimer's disease. *Nature* **537**, 50–56 (2016).
  - Thambisetty, M. et al. Proteome-based plasma markers of brain amyloid- $\beta$  deposition in non-demented older individuals. *J. Alzheimers Dis.* **22**, 1099–1109 (2010).
  - Kiddle, S. J. et al. Plasma based markers of [11C] PiB-PET brain amyloid burden. *PLoS ONE* **7**, e44260 (2012).
  - Burnham, S. C. et al. A blood-based predictor for neocortical A $\beta$  burden in Alzheimer's disease: results from the AIBL study. *Mol. Psychiatry* **19**, 519–526 (2014).
  - Ashton, N. J. et al. Blood protein predictors of brain amyloid for enrichment in clinical trials? *Alzheimers Dement. (Amst.)* **1**, 48–60 (2015).
  - Voyle, N. et al. Blood protein markers of neocortical amyloid- $\beta$  burden: a candidate study using SOMAscan Technology. *J. Alzheimers Dis.* **46**, 947–961.
  - Begley, C. G. & Ellis, L. M. Drug development: raise standards for preclinical cancer research. *Nature* **483**, 531–533 (2012).
  - Prinz, F., Schlange, T. & Asadullah, K. Believe it or not: how much can we rely on published data on potential drug targets? *Nat. Rev. Drug Discov.* **10**, 712–712 (2011).
  - Baker, M. 1,500 scientists lift the lid on reproducibility. *Nature* **533**, 452 (2016).
  - Ioannidis, J. P. A. & Panagiotou, O. A. Comparison of effect sizes associated with biomarkers reported in highly cited individual articles and in subsequent meta-analyses. *JAMA* **305**, 2200–2210 (2011). **This study showed that highly cited strong biomarker effects are often overestimated compared to larger meta-analyses of the same associations.**
  - JCGM. *International vocabulary of metrology—Basic and general concepts and associated terms (VIM), 3rd edn, 2008 version with minor corrections*. (JCGM, 2012).
  - Button, K. S. et al. Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* **14**, 365–376 (2013).
  - Vabalas, A., Gowen, E., Poliakoff, E. & Casson, A. J. Machine learning algorithm validation with a limited sample size. *PLoS ONE* **14**, e0224365 (2019).
  - Olsson, B. et al. CSF and blood biomarkers for the diagnosis of Alzheimer's disease: a systematic review and meta-analysis. *Lancet Neurol.* **15**, 673–684 (2016).
  - Kovacs, G. G. et al. Non-Alzheimer neurodegenerative pathologies and their combinations are more frequent than commonly believed in the elderly brain: a community-based autopsy series. *Acta Neuropathol.* **126**, 365–384 (2013).
  - van Westen, D. et al. Cerebral white matter lesions—associations with A $\beta$  isoforms and amyloid PET. *Sci. Rep.* **6**, 20709 (2016).
  - Tu, J. & Bennett, P. Parallelism experiments to evaluate matrix effects, selectivity and sensitivity in ligand-binding assay method development: pros and cons. *Bioanalysis*. <https://doi.org/10.4155/bio-2017-0084>.
  - Plikaytis, B. D. et al. Determination of parallelism and nonparallelism in bioassay dilution curves. *J. Clin. Microbiol.* **32**, 2441–2447 (1994).
  - Tate, J. & Ward, G. Interferences in Immunoassay. *Clin. Biochem. Rev.* **25**, 105–120 (2004).
  - Palmqvist, S. et al. Accuracy of brain amyloid detection in clinical practice using cerebrospinal fluid  $\beta$ -amyloid 42: a cross-validation study against amyloid positron emission tomography. *JAMA Neurol.* **71**, 1282–1289 (2014).
  - Andreasson, U. et al. Commutability of the certified reference materials for the standardization of  $\beta$ -amyloid 1–42 assay in human cerebrospinal fluid: lessons for tau and  $\beta$ -amyloid 1–40 measurements. *Clin. Chem. Lab. Med.* **56**, 2058–2066 (2018).
  - Mattsson, N. et al. Reference measurement procedures for Alzheimer's disease cerebrospinal fluid biomarkers: definitions and approaches with focus on amyloid beta42. *Biomark. Med.* **6**, 409–417 (2012).
  - Leinenbach, A. et al. Mass spectrometry-based candidate reference measurement procedure for quantification of amyloid- $\beta$  in cerebrospinal fluid. *Clin. Chem.* **60**, 987–994 (2014). **This is a description of a reference measurement procedure for a CSF biomarker, which is a key step towards fluid biomarker reproducibility.**
  - Kuhlmann, J. et al. CSF A $\beta$ 1–42 - an excellent but complicated Alzheimer's biomarker - a route to standardisation. *Clin. Chim. Acta* **467**, 27–33 (2017).
  - Caruso, B., Bovo, C. & Guidi, G. C. Causes of preanalytical interferences on laboratory immunoassays—a critical review. *EJIFCC* **31**, 70–84 (2020).
  - Benedict, C., Blennow, K., Zetterberg, H. & Cedernaes, J. Effects of acute sleep loss on diurnal plasma dynamics of CNS health biomarkers in young men. *Neurology* <https://doi.org/10.1212/WNL.0000000000008866> (2020).
  - Trombetta, B. A. et al. The technical reliability and biotemporal stability of cerebrospinal fluid biomarkers for profiling multiple pathophysiologies in Alzheimer's disease. *PLoS ONE* **13**, e0193707 (2018).
  - Lucey, B. P., Fagan, A. M., Holtzman, D. M., Morris, J. C. & Bateman, R. J. Diurnal oscillation of CSF A $\beta$  and other AD biomarkers. *Mol. Neurodegener.* **12**, 36 (2017).
  - Lim, A. S. P. et al. Seasonal plasticity of cognition and related biological measures in adults with and without Alzheimer disease: analysis of multiple cohorts. *PLoS Med.* **15**, e1002647 (2018).
  - Hansson, O., Lehmann, S., Otto, M., Zetterberg, H. & Lewczuk, P. Advantages and disadvantages of the use of the CSF Amyloid  $\beta$  (A $\beta$ ) 42/40 ratio in the diagnosis of Alzheimer's Disease. *Alzheimers Res. Ther.* **11**, 34 (2019).
  - Janelidze, S. et al. CSF A $\beta$ 42/A $\beta$ 40 and A $\beta$ 42/A $\beta$ 38 ratios: better diagnostic markers of Alzheimer disease. *Ann. Clin. Transl. Neurol.* **3**, 154–165 (2016).
  - Janelidze, S., Stomrud, E., Brix, B. & Hansson, O. Towards a unified protocol for handling of CSF before  $\beta$ -amyloid measurements. *Alzheimers Res. Ther.* **11**, 63 (2019).
  - Hansson, O. et al. The impact of preanalytical variables on measuring cerebrospinal fluid biomarkers for Alzheimer's disease diagnosis: a review. *Alzheimers Dement.* **14**, 1313–1333 (2018).
  - Rozga, M., Bittner, T., Höglund, K. & Blennow, K. Accuracy of cerebrospinal fluid A $\beta$ 1–42 measurements: evaluation of pre-analytical factors using a novel Elecsys immunoassay. *Clin. Chem. Lab. Med.* **55**, 1545–1554 (2017). **This study includes preanalytical factors that may be controlled to increase reproducibility of CSF biomarker measurements.**
  - Irani, D. N. et al. Cleavage of cystatin C in the cerebrospinal fluid of patients with multiple sclerosis. *Ann. Neurol.* **59**, 237–247 (2006).
  - Hansson, S. F. et al. Cystatin C in cerebrospinal fluid and multiple sclerosis. *Ann. Neurol.* **62**, 193–196 (2007).
  - Carrette, O., Burkhard, P. R., Hughes, S., Hochstrasser, D. F. & Sanchez, J. C. Truncated cystatin C in cerebrospinal fluid: Technical [corrected] artefact or biological process? *Proteomics* **5**, 3060–3065 (2005).
  - Mattsson, N. et al. CSF biomarker variability in the Alzheimer's Association quality control program. *Alzheimers Dement.* **9**, 251–261 (2013).
  - Landau, S. M., Horng, A. & Jagust, W. J. Memory decline accompanies subthreshold amyloid accumulation. *Neurology* **90**, e1452–e1460 (2018).
  - Mattsson, N. et al. The Alzheimer's Association external quality control program for cerebrospinal fluid biomarkers. *Alzheimers Dement.* **7**, 386–395 e6 (2011).
  - Bittner, T. et al. Technical performance of a novel, fully automated electrochemiluminescence immunoassay for the quantitation of  $\beta$ -amyloid (1–42) in human cerebrospinal fluid. *Alzheimer's Dement.* **12**, 517–526 (2016).
  - Hansson, O. et al. CSF biomarkers of Alzheimer's disease concord with amyloid- $\beta$  PET and predict clinical progression: A study of fully automated immunoassays in BioFINDER and ADNI cohorts. *Alzheimers Dement.* **14**, 1470–1481 (2018).
  - Palmqvist, S. et al. Performance of Fully Automated Plasma Assays as Screening Tests for Alzheimer Disease-Related  $\beta$ -Amyloid Status. *JAMA Neurol.* <https://doi.org/10.1001/jamaneurol.2019.1632> (2019).
  - McDade, E. et al. Longitudinal cognitive and biomarker changes in dominantly inherited Alzheimer disease. *Neurology* **91**, e1295–e1306 (2018).
  - Bertens, D., Tijms, B. M., Scheltens, P., Teunissen, C. E. & Visser, P. J. Unbiased estimates of cerebrospinal fluid  $\beta$ -amyloid 1–42 cutoffs in a large memory clinic population. *Alzheimers Res. Ther.* **9**, 8 (2017).
  - Palmqvist, S., Mattsson, N. & Hansson, O. Alzheimer's Disease Neuroimaging Initiative . Cerebrospinal fluid analysis detects cerebral amyloid- $\beta$  accumulation earlier than positron emission tomography. *Brain* **139**, 1226–1236 (2016).

58. Villeneuve, S. et al. Existing Pittsburgh compound-B positron emission tomography thresholds are too high: statistical and pathological evaluation. *Brain* **138**, 2020–2033 (2015).
59. Horowitz, G. L. Reference intervals: practical aspects. *EJIFCC* **19**, 95–105 (2008).
60. Mapstone, M. et al. Plasma phospholipids identify antecedent memory impairment in older adults. *Nat. Med.* **20**, 415–418 (2014).
61. Irwin, D. J. Neuropathological validation of cerebrospinal fluid biomarkers in neurodegenerative diseases. *J. Appl. Lab. Med.* **5**, 232–238 (2020).
62. Beach, T. G., Monsell, S. E., Phillips, L. E. & Kukull, W. Accuracy of the clinical diagnosis of Alzheimer disease at national institute on aging Alzheimer's Disease Centers, 2005–2010. *J. Neuropathol. Exp. Neurol.* **71**, 266–273 (2012). **This study shows that clinical diagnosis of Alzheimer's disease often disagree with neuropathological confirmation of the disease.**
63. Schöll, M. et al. Biomarkers for tau pathology. *Mol. Cell. Neurosci.* <https://doi.org/10.1016/j.mcn.2018.12.001> (2018).
64. Zetterberg, H. AlzForum. AlzBiomarker version 2.1. (University of Gothenburg, 2018). <https://www.alzforum.org/alzbiomarker>.
65. Schindler, S. E. et al. High-precision plasma  $\beta$ -amyloid 42/40 predicts current and future brain amyloidosis. *Neurology* <https://doi.org/10.1212/WNL.0000000000008081> (2019).
66. Nakamura, A. et al. High performance plasma amyloid- $\beta$  biomarkers for Alzheimer's disease. *Nature* **554**, 249–254 (2018).
67. Fandos, N. et al. Plasma amyloid  $\beta$  42/40 ratios as biomarkers for amyloid  $\beta$  cerebral deposition in cognitively normal individuals. *Alzheimers Dement (Amst.)* **8**, 179–187 (2017).
68. Janelidze, S. et al. Plasma  $\beta$ -amyloid in Alzheimer's disease and vascular disease. *Sci. Rep.* **6**, 26801 (2016).
69. Devanand, D. P. et al. Plasma A $\beta$  and PET PiB binding are inversely related in mild cognitive impairment. *Neurology* **77**, 125–131 (2011).
70. Bridel, C. et al. Diagnostic value of cerebrospinal fluid neurofilament light protein in neurology: a systematic review and meta-analysis. *JAMA Neurol.* **76**, 1035–1048 (2019).
71. Abu-Rumeileh, S. et al. Diagnostic value of surrogate CSF biomarkers for Creutzfeldt-Jakob disease in the era of RT-QuIC. *J. Neurol.* **266**, 3136–3143 (2019).
72. Forgrave, L. M., Ma, M., Best, J. R. & DeMarco, M. L. The diagnostic performance of neurofilament light chain in CSF and blood for Alzheimer's disease, frontotemporal dementia, and amyotrophic lateral sclerosis: a systematic review and meta-analysis. *Alzheimers Dement (Amst.)* **11**, 730–743 (2019).
73. Parnetti, L. et al. CSF and blood biomarkers for Parkinson's disease. *Lancet Neurol.* **18**, 573–586 (2019).
74. Zhao, Y., Xin, Y., Meng, S., He, Z. & Hu, W. Neurofilament light chain protein in neurodegenerative dementia: A systematic review and network meta-analysis. *Neurosci. Biobehav. Rev.* **102**, 123–138 (2019).
75. Skillbäck, T. et al. CSF neurofilament light differs in neurodegenerative diseases and predicts severity and survival. *Neurology* **83**, 1945–1953 (2014).
76. Mattsson, N., Andreasson, U., Zetterberg, H. & Blennow, K. Alzheimer's Disease Neuroimaging Initiative . Association of plasma neurofilament light with neurodegeneration in patients With Alzheimer disease. *JAMA Neurol.* **74**, 557–566 (2017).
77. Bacioglu, M. et al. Neurofilament light chain in blood and CSF as marker of disease progression in mouse models and in neurodegenerative diseases. *Neuron* **91**, 56–66 (2016).
78. Connor, A., Wang, H., Appleby, B. S. & Rhoads, D. D. Clinical laboratory tests used to aid in diagnosis of human prion disease. *J. Clin. Microbiol.* **57**, e00769–19 (2019).
79. Behaeghe, O., Mangelschots, E., De Vil, B. & Cras, P. A systematic review comparing the diagnostic value of 14-3-3 protein in the cerebrospinal fluid, RT-QuIC and RT-QuIC on nasal brushing in sporadic Creutzfeldt-Jakob disease. *Acta Neurologica Belgica* **118**, 395–403 (2018).
80. Thorsell, A. et al. Neurogranin in cerebrospinal fluid as a marker of synaptic degeneration in Alzheimer's disease. *Brain Res.* **1362**, 13–22 (2010).
81. Portelius, E. et al. Cerebrospinal fluid neurogranin concentration in neurodegeneration: relation to clinical phenotypes and neuropathology. *Acta Neuropathol.* **136**, 363–376 (2018).
82. Laterza, O. F. et al. Identification of novel brain biomarkers. *Clin. Chem.* **52**, 1713–1721 (2006).
83. Llorens, F. et al. YKL-40 in the brain and cerebrospinal fluid of neurodegenerative dementias. *Mol. Neurodegeneration* **12**, 83 (2017).
84. Saijo, E. et al. in *Protein Misfolding Diseases: Methods and Protocols* (ed. Gomes, C. M.) 19–37 (Springer, 2019).
85. Fairfoul, G. et al. Alpha-synuclein RT-QuIC in the CSF of patients with alpha-synucleinopathies. *Ann. Clin. Transl. Neurol.* **3**, 812–818 (2016).
86. Mielke, M. M. et al. Plasma phospho-tau181 increases with Alzheimer's disease clinical severity and is associated with tau- and amyloid-positron emission tomography. *Alzheimers Dement.* **14**, 989–997 (2018).
87. Janelidze, S. et al. Plasma P-tau181 in Alzheimer's disease: relationship to other biomarkers, differential diagnosis, neuropathology and longitudinal progression to Alzheimer's dementia. *Nat. Med.* **26**, 379–386 (2020).
88. Thijssen, E. H. et al. Diagnostic value of plasma phosphorylated tau181 in Alzheimer's disease and frontotemporal lobar degeneration. *Nat. Med.* **26**, 387–397 (2020).
89. Janelidze, S. et al. Cerebrospinal fluid p-tau217 performs better than p-tau181 as a biomarker of Alzheimer's disease. *Nat. Commun.* **11**, 1683 (2020).
90. Palmqvist, S. et al. Discriminative accuracy of plasma phospho-tau217 for Alzheimer disease vs other neurodegenerative disorders. *JAMA* <https://doi.org/10.1001/jama.2020.12134> (2020).
91. Ewers, M. et al. Increased soluble TREM2 in cerebrospinal fluid is associated with reduced cognitive and clinical decline in Alzheimer's disease. *Sci. Transl. Med.* **11** eaav6221 (2019).
92. Shen, X.-N. et al. Inflammatory markers in Alzheimer's disease and mild cognitive impairment: a meta-analysis and systematic review of 170 studies. *J. Neurol. Neurosurg. Psychiatry* **90**, 590–598 (2019).
93. Whelan, C. D. et al. Multiplex proteomics identifies novel CSF and plasma biomarkers of early Alzheimer's disease. *Acta Neuropathol. Commun.* **7**, 169 (2019).
94. Mattsson, N. et al. Cerebrospinal fluid microglial markers in Alzheimer's disease: elevated chitotriosidase activity but lack of diagnostic utility. *Neuromolecular Med* **13**, 151–159 (2011).
95. Steinacker, P. et al. Chitotriosidase (CHIT1) is increased in microglia and macrophages in spinal cord of amyotrophic lateral sclerosis and cerebrospinal fluid levels correlate with disease severity and progression. *J. Neurol. Neurosurg. Psychiatry* **89**, 239–247 (2018).
96. Eusebi, P. et al. Diagnostic utility of cerebrospinal fluid  $\alpha$ -synuclein in Parkinson's disease: a systematic review and meta-analysis. *Mov. Disord.* **32**, 1389–1400 (2017).
97. Mackin, R. S. et al. Cerebrospinal fluid  $\alpha$ -synuclein and Lewy body-like symptoms in normal controls, mild cognitive impairment, and Alzheimer's disease. *J. Alzheimers Dis.* **43**, 1007–1016 (2015).
98. Barkovits, K. et al. Blood contamination in CSF and its impact on quantitative analysis of alpha-synuclein. *Cells* **9**, 370 (2020).
99. Hall, S. et al. Accuracy of a panel of 5 cerebrospinal fluid biomarkers in the differential diagnosis of patients with dementia and/or parkinsonian disorders. *Arch. Neurol.* **69**, 1445–1452 (2012).
100. Chang, C.-W., Yang, S.-Y., Yang, C.-C., Chang, C.-W. & Wu, Y.-R. Plasma and serum alpha-synuclein as a biomarker of diagnosis in patients with Parkinson's disease. *Front. Neurol.* **10**, 1388 (2019).
101. Fan, Z. et al. Systemic activation of NLRP3 inflammasome and plasma  $\alpha$ -synuclein levels are correlated with motor severity and progression in Parkinson's disease. *J. Neuroinflammation* **17**, 11 (2020).
102. Ng, A. S. L. et al. Plasma alpha-synuclein detected by single molecule array is increased in PD. *Ann. Clin. Transl. Neurol.* **6**, 615–619 (2019).
103. Tian, C. et al. Erythrocytic  $\alpha$ -Synuclein as a potential biomarker for Parkinson's disease. *Transl. Neurodegener.* **8**, 15 (2019).
104. Fossati, S. et al. Plasma tau complements CSF tau and P-tau in the diagnosis of Alzheimer's disease. *Alzheimers Dement. (Amst.)* **11**, 483–492 (2019).
105. Mattsson, N. et al. Plasma tau in Alzheimer disease. *Neurology* **87**, 1827–1835 (2016).
106. Zetterberg, H. et al. Plasma tau levels in Alzheimer's disease. *Alzheimers Res. Ther.* **5**, 9 (2013).
107. Wang, T. et al. The efficacy of plasma biomarkers in early diagnosis of Alzheimer's disease. *Int J. Geriatr. Psychiatry* **29**, 713–719 (2014).
108. Sparks, D. L. et al. Tau is reduced in AD plasma and validation of employed ELISA methods. *Am. J. Neurodegener. Dis.* **1**, 99–106 (2012).
109. Talwar, P. et al. Meta-analysis of apolipoprotein E levels in the cerebrospinal fluid of patients with Alzheimer's disease. *J. Neurol. Sci.* **360**, 179–187 (2016).
110. Kiddle, S. J. et al. Candidate blood proteome markers of Alzheimer's disease onset and progression: a systematic review and replication study. *JAD* **38**, 515–531 (2013). **This is a large scale effort to replicate blood-based biomarkers for Alzheimer's disease, illustrating the poor reproducibility for many biomarkers.**
111. Ray, S. et al. Classification and prediction of clinical Alzheimer's diagnosis based on plasma signaling proteins. *Nat. Med.* **13**, 1359–1362 (2007).
112. Björkqvist, M., Ohlsson, M., Minthon, L. & Hansson, O. Evaluation of a previously suggested plasma biomarker panel to identify Alzheimer's disease. *PLoS ONE* **7**, e29868 (2012).
113. Casanova, R. et al. Blood metabolite markers of preclinical Alzheimer's disease in two longitudinally followed cohorts of older individuals. *Alzheimers Dement.* **12**, 815–822 (2016).
114. Mapstone, M., Cheema, A. K., Zhong, X., Fiandaca, M. S. & Federoff, H. J. Biomarker validation: Methods and matrix matter. *Alzheimer's Dement.* **13**, 608–609 (2017).
115. Thambisetty, M., Casanova, R., Varma, S. & Legido Quigley, C. Peril beyond the winner's curse: a small sample size is the bane of biomarker discovery. *Alzheimer's Dement.* **13**, 606–607 (2017).

116. Benatar, M., Wu, J., Andersen, P. M., Lombardi, V. & Malaspina, A. Neurofilament light: a candidate biomarker of presymptomatic amyotrophic lateral sclerosis and phenoconversion. *Ann. Neurol.* **84**, 130–139 (2018).
117. Bhatia, G. S. et al. Evaluation of B-type natriuretic peptide for validation of a heart failure register in primary care. *BMC Cardiovasc. Disord.* **7**, 23 (2007).
118. Galvani, L., Flanagan, J., Sargazi, M. & Neithercut, W. D. Validation of serum free light chain reference ranges in primary care. *Ann. Clin. Biochem.* **53**, 399–404 (2016). *This study shows that biomarker findings may not easily reproduce in a primary care setting.*

### Acknowledgements

K.B. is supported by the Swedish Research Council (#2017-00915), the Alzheimer Drug Discovery Foundation (ADDF), USA (#RDAPB-201809-2016615), the Swedish Alzheimer Foundation (#AF-742881), Hjärnfonden, Sweden (#FO2017-0243), the Swedish state under the agreement between the Swedish government and the County Councils, the ALF-agreement (#ALFGBG-715986). Work at the authors' research center was supported by the Swedish Research Council, the Knut and Alice Wallenberg foundation, the Marianne and Marcus Wallenberg foundation, the Strategic Research Area MultiPark (Multidisciplinary Research in Parkinson's disease) at Lund University, the Swedish Alzheimer Foundation, the Swedish Brain Foundation, The Swedish Alzheimer Association, The Swedish Medical Association, The Parkinson foundation of Sweden, The Parkinson Research Foundation, the Skåne University Hospital Foundation, The Bundy Academy, The Konung Gustaf V:s och Drottning Victorias Frimurarestiftelse, and the Swedish federal government under the ALF agreement.

### Author contributions

N.M.C. drafted the manuscript. S.P., K.B., and O.H. revised the manuscript for intellectual content.

### Competing interests

N.M.C. and S.P. report no competing interest. K.B. has served as a consultant or at advisory boards for Abcam, Axon, Biogen, Lilly, MagQu, Novartis and Roche Diagnostics, and is a co-founder of Brain Biomarker Solutions in Gothenburg AB, a GU

Venture-based platform company at the University of Gothenburg. O.H. has acquired research support (for the institution) from Roche, Pfizer, GE Healthcare, Biogen, AVID Radiopharmaceuticals and Euroimmun. In the past 2 years, he has received consultancy/speaker fees (paid to the institution) from Biogen and Roche.

### Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41467-020-19957-6>.

Correspondence and requests for materials should be addressed to N.M.-C. or O.H.

Peer review information *Nature Communications* thanks Ivan Koychev, Jens Wiltfang, and other, anonymous, reviewers for their contributions to the peer review of this work. Peer review reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020