

## RESEARCH ARTICLE

# Measurement batch differences and between-batch conversion of Alzheimer's disease cerebrospinal fluid biomarker values

Yue Ma<sup>1,2</sup> | Derek L. Norton<sup>2,3</sup> | Carol A. Van Hulle<sup>1,2</sup> | Richard J. Chappell<sup>3,4</sup> |  
 Karen K. Lazar<sup>1,2,5</sup> | Erin M. Jonaitis<sup>1,6</sup> | Rebecca L. Kosciak<sup>1,6</sup> | Lindsay R. Clark<sup>1,2,5,6</sup> |  
 Rachel Krause<sup>1,2</sup> | Ulf Andreasson<sup>7,8</sup> | Nathaniel A. Chin<sup>1,2</sup> | Barbara B. Bendlin<sup>1,2,6</sup> |  
 Sanjay Asthana<sup>1,2,5</sup> | Ozioma C. Okonkwo<sup>1,2,6</sup> | Carey E. Gleason<sup>1,2,5,6</sup> |  
 Sterling C. Johnson<sup>1,2,5,6</sup> | Henrik Zetterberg<sup>7,8,9,10</sup> | Kaj Blennow<sup>7,8</sup> |  
 Cynthia M. Carlsson<sup>1,2,5,6</sup>

<sup>1</sup> Division of Geriatrics and Gerontology, Department of Medicine, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin, USA

<sup>2</sup> Wisconsin Alzheimer's Disease Research Center, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin, USA

<sup>3</sup> Department of Biostatistics and Medical Informatics, University of Wisconsin, Madison, Wisconsin, USA

<sup>4</sup> Department of Statistics, University of Wisconsin, Madison, Wisconsin, USA

<sup>5</sup> Geriatric Research Education and Clinical Center, William S. Middleton Memorial Veterans Hospital, Madison, Wisconsin, USA

<sup>6</sup> Wisconsin Alzheimer's Institute, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin, USA

<sup>7</sup> Department of Psychiatry and Neurochemistry, Institute of Neuroscience and Physiology, the Sahlgrenska Academy at the University of Gothenburg, Mölndal, Sweden

<sup>8</sup> Clinical Neurochemistry Laboratory, Sahlgrenska University Hospital, Mölndal, Sweden

<sup>9</sup> Department of Neurodegenerative Disease, UCL Institute of Neurology, London, UK

<sup>10</sup> UK Dementia Research Institute at UCL, London, UK

## Correspondence

Yue Ma, University of Wisconsin School of Medicine and Public Health, J5/1 Mezzanine, 600 Highland Avenue, Madison, WI 53792, USA.

Email: yma@medicine.wisc.edu

## Present address

Rachel Krause, Calimetrix, 505 S. Rosa Rd, Suite 35F, Madison WI, USA

## Funding information

National Institutes of Health; National Institute on Aging; National Institute on Aging Swedish Research Council; European Research Council; Swedish State Support for Clinical Research; Alzheimer Drug Discovery Foundation, USA; UK Dementia Research Institute at UCL European Union Joint Program for Neurodegenerative Disorders; Swedish Alzheimer Foundation; Hjärnfonden

## Abstract

**Introduction:** Batch differences in cerebrospinal fluid (CSF) biomarker measurement can introduce bias into analyses for Alzheimer's disease studies. We evaluated and adjusted for batch differences using statistical methods.

**Methods:** A total of 792 CSF samples from 528 participants were assayed in three batches for 12 biomarkers and 3 biomarker ratios. Batch differences were assessed using Bland-Altman plot, paired *t* test, Pitman-Morgan test, and linear regression. Generalized linear models were applied to convert CSF values between batches.

**Results:** We found statistically significant batch differences for all biomarkers and ratios, except that neurofilament light was comparable between batches 1 and 2. The conversion models generally had high  $R^2$  except for converting P-tau between batches 1 and 3.

**Discussion:** Between-batch conversion allows harmonized CSF values to be used in the same analysis. Such method may be applied to adjust for other sources of variability in measuring CSF or other types of biomarkers.

## KEYWORDS

Alzheimer's disease, batch difference, biomarker, cerebrospinal fluid, conversion, generalized linear model

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2021 The Authors. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring* published by Wiley Periodicals, LLC on behalf of Alzheimer's Association

## 1 | BACKGROUND

Cerebrospinal fluid (CSF) biomarkers play an increasingly important role in the early diagnosis of Alzheimer's disease (AD),<sup>1-8</sup> and have the potential to improve risk estimates of AD,<sup>9</sup> identify individuals for early intervention who are in the pre-clinical period of AD,<sup>10</sup> and monitor response to interventions.<sup>11</sup> A challenge in using CSF biomarkers for longitudinal AD research is the variability in CSF biomarker measurements.<sup>12-15</sup> CSF quality control (QC) studies have identified various factors contributing to measurement variability.<sup>5,12,15-21</sup> International collaborations are working to minimize such variability by developing standardized reference materials, guidelines, and procedures for CSF collection, processing, storage, and assay techniques.<sup>11,13</sup> However, even when most factors are addressed by the application of standardized procedures and methods, between-batch differences, that is, differences between analytical runs at varied time points, can still contribute to non-trivial variability in CSF results,<sup>5,12,21-24</sup> possibly due to subtle assay variations in reagents and in the technical performance of the instrument.<sup>5,12,21,23</sup> When data generated from CSF samples assayed in different batches are combined for analysis, between-batch differences can confound inter-individual and longitudinal intra-individual differences, leading to bias in estimating AD risks and longitudinal trajectories in CSF biomarker changes.

Statistical approaches have been proposed for assessing measurement differences between methods and can be applied to evaluating the batch differences. These approaches include Bland-Altman plots,<sup>25-28</sup> testing differences in means, testing ratios of variances,<sup>29-31</sup> and linear regressions.<sup>32,33</sup> However, these methods are generally limited to the clinical chemistry and statistics literature, and are not widely known in the AD research field. A possible reason could be the underestimation of the severity of measurement variation in CSF biomarkers. Thus the first pair of the study aims were to (1) increase the awareness of this measure variability issue and (2) illustrate how to use these approaches to evaluate the batch differences.

Few statistical approaches have been developed to adjust for batch differences. A traditional approach to adjust for batch differences is to include the batch indicator as a covariate in the analysis.<sup>34</sup> However, this approach is applicable only when CSF biomarker variables serve as outcomes (not as predictors). Moreover, it is valid only when the batch difference is constant for all CSF samples. To overcome these barriers, linear regression can be used to convert CSF values between batches.<sup>34-36</sup> However, such analysis commonly violates model assumptions on normality and homoscedasticity (ie., homogeneity of variance).<sup>34</sup> Therefore, logarithm or square root transformations can be applied to the CSF batch values used as the regression outcomes to solve this problem.<sup>34,35</sup> When the converted values are included in the analysis, raw values measured in the base batch need transformation as well to be on the same scale with the converted values. In studies that include more than two batches, a unique transformation likely exists for each additional batch relative to the original batch. As a result, the values converted from different batches would be on dif-

### RESEARCH IN CONTEXT

- 1. Systematic Review:** We conducted literature review using PubMed and identified several studies that evaluated batch differences in amyloid beta ( $A\beta$ ) and tau cerebrospinal fluid (CSF) biomarker measurement. Few studies addressed how to adjust for batch differences. These studies have been cited.
- 2. Interpretation:** Consistent with previous findings on  $A\beta$  and tau, we found significant batch differences in 12 CSF biomarkers and 3 biomarker ratios. With the application of generalized linear models (GLMs), we developed conversion models to adjust for batch differences. GLMs have advantages over linear regression and with batch indicator included as a covariate.
- 3. Future Directions:** The conversion models could be improved by increasing the comparability between samples used for model development and samples to be converted. Simulation studies could be used for evaluating the performance of conversion models under various situations in terms of accuracy and efficiency.

### HIGHLIGHTS

- We evaluated batch differences using four statistical approaches.
- We found significant batch differences for 12 CSF biomarkers with the least for neurofilament light (NfL).
- We developed between-batch conversion models using generalized linear models (GLMs).
- We compared strength and limitations of GLMs with other adjustment methods.
- We provided recommendations for how to develop and apply conversion models.

ferent scales. Furthermore, data on transformed scales are difficult to interpret, and more so for more complex models with interactive or nonlinear terms. Thus the second pair of the study aims were to (3) implement a statistical approach to adjust for batch differences which can overcome these limitations and (4) make recommendations on how to apply this approach. We pursued these aims through our experience with evaluating and adjusting for batch differences for an extended panel of CSF biomarkers in a typical AD research sample.

## 2 | METHODS

### 2.1 | Participants and CSF samples

CSF samples included in this analysis were from participants of six National Institutes of Health (NIH)-funded AD studies. Informed

consent was obtained from each participant. A total of 792 CSF samples were collected from 528 participants (mean age = 61.2 years, SD = 8.7, range 40.8 to 93.1). A total of 351 had one CSF sample, 98 had two, 71 had three, and 8 had four. Multiple samples from the same participant were collected at different ages. The clinical diagnosis associated with the CSF sample collection included 707 (89.3%) cognitively unimpaired, 44 (5.6%) dementia due to clinical AD, 35 (4.4%) mild cognitive impairment (MCI) due to clinical AD, 4 (0.5%) MCI due to other causes, and 2 (0.3%) cognitively impaired but not MCI.

## 2.2 | CSF sample collection, processing, and storage

Following a standard pre-analytic protocol across all included studies, we collected all CSF samples from participants in the morning after an overnight fast using a Sprotte 24- or 25-gauge spinal needle to extract 22 mL CSF into polypropylene syringes. The CSF was then combined, gently mixed, and centrifuged at 2000 *g* for 10 minutes at 4°C. Supernatants were frozen in 0.5 mL aliquots in polypropylene tubes and stored in a –80°C freezer. Samples remained frozen until assayed.

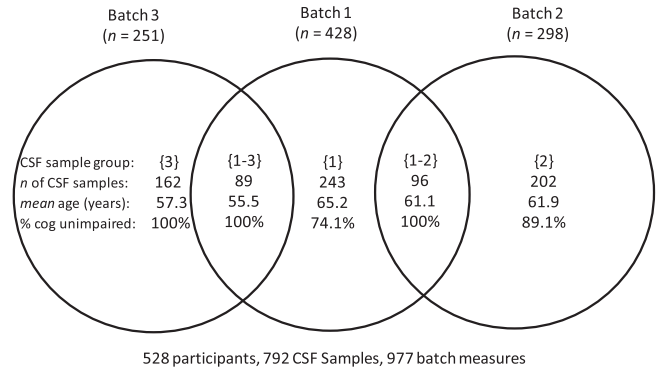
## 2.3 | CSF sample assay batches and methods

Over a 3-year period, the CSF samples were assayed in three batches at the Clinical Neurochemistry Laboratory, Sahlgrenska Academy, University of Gothenburg, Sweden, with a subset of samples re-assayed, which resulted in 977 batch measures. CSF samples were categorized into five groups based on their processing batch(es). Groups {1}, {2}, {3} represent the CSF samples assayed in a single batch 1, 2, and 3, respectively. Group {1-2} represents the CSF samples assayed in both batches 1 and 2, and group {1-3} represents the CSF samples assayed in both batches 1 and 3. Figure 1 and Table S1 summarize the distribution of CSF samples between the groups, and age and diagnosis for each group. Twelve CSF biomarkers were assayed, including  $\beta$ -amyloid 1 to 42 peptide (A $\beta$ 1-42), total tau (T-tau), phosphorylated tau (P-tau) (assayed by INNOTEST for batches 1 and 2, Luminex xMAP for batch 3); amyloid- $\beta$  peptides ending at the 42nd, 40th, 38th amino acid (A $\beta$ X-42, A $\beta$ X-40, A $\beta$ X-38, MSD Triplex); soluble amyloid precursor protein- $\alpha$  (sAPP- $\alpha$ ) and protein- $\beta$  (sAPP- $\beta$ ), chitinase-3-like protein 1 (YKL-40), monocyte chemoattractant protein-1 (MCP-1, R&D), neurofilament light chain (NfL), and neurogranin (Ng, not assayed in batch 3). In addition, three biomarker ratios were calculated: A $\beta$ X-42/A $\beta$ X-40, T-tau/A $\beta$ 1-42, and P-tau/A $\beta$ 1-42.

## 2.4 | Statistical analyses

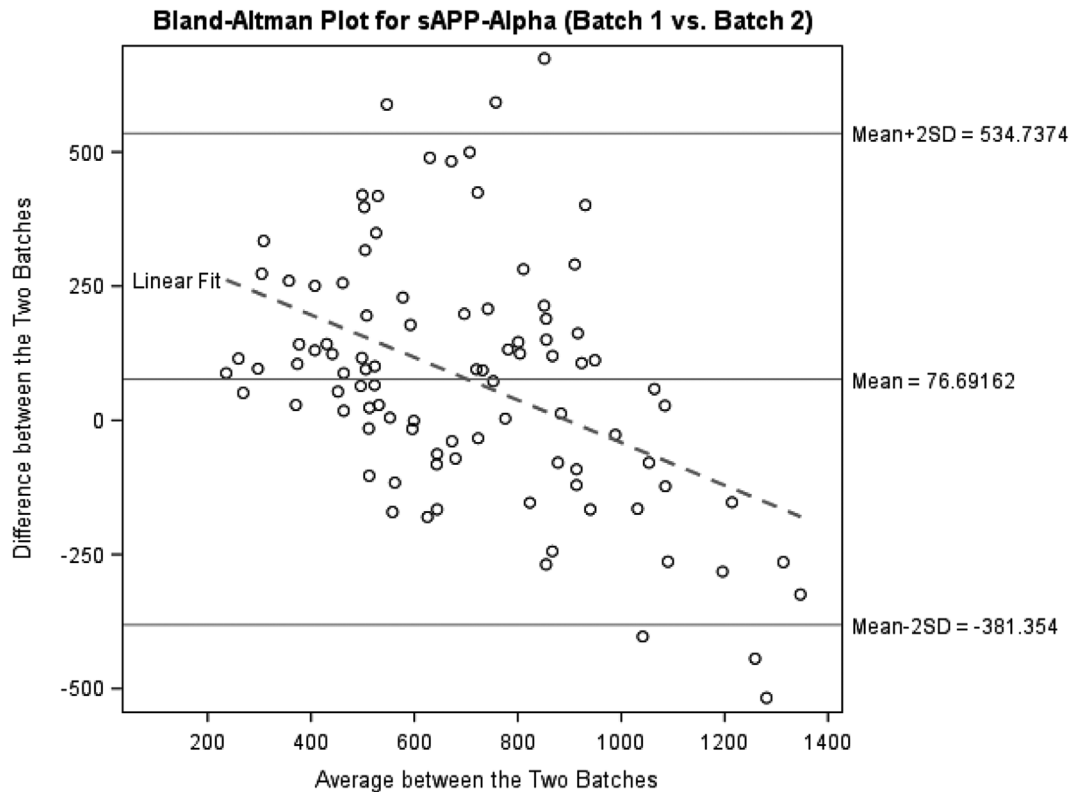
### 2.4.1 | Assessment of between-batch differences

Between-batch differences were assessed for CSF groups {1-2} (*n* = 96) and {1-3} (*n* = 89) separately with a series of four analyses, including



**FIGURE 1** Distribution of cerebrospinal fluid (CSF) samples between measurement batches. CSF samples included in this analysis were from participants of six National Institutes of Health (NIH)-funded Alzheimer's disease (AD) studies: the Wisconsin Alzheimer's Disease Research Center (ADRC) Clinical Core,<sup>45</sup> the Wisconsin Registry for Alzheimer's Prevention (WRAP),<sup>46</sup> the Longitudinal Early Alzheimer's Detection (LEAD) study, The Longitudinal Course of Neural Function and Amyloid in People At Risk for Alzheimer's Disease (PREDICT) study, the PIB Imaging in People at Risk for Alzheimer's Disease (PIPR) study, and the Statins in Healthy, At Risk Adults: Impact on Amyloid and Regional Perfusion (SHARP) study.<sup>47</sup> A total of 792 CSF samples collected from 528 participants were assayed in three batches, with a subset of samples re-assayed, which resulted in 977 batch measures. Groups {1}, {2}, {3} represent the CSF samples assayed in a single batch 1, 2, and 3, respectively. Group {1-2} represents the CSF samples assayed in both batches 1 and 2, and group {1-3} represents the CSF samples assayed in both batches 1 and 3. The groups with CSF samples assayed in two batches had smaller sample sizes than the groups with CSF samples assayed in a single batch. The mean age for each group ranged from 55.5 to 65.2 years. The clinical diagnoses associated with the CSF samples were all cognitively unimpaired in groups {1-2}, {1-3}, {3}, and were mostly cognitively unimpaired in groups {1}, {2}. Types and counts (percentages) of impaired diagnoses for these two groups are summarized in Table S1

Bland-Altman plots,<sup>25-28</sup> testing differences in means with the paired *t* test, testing ratios of variances with the Pitman-Morgan test,<sup>29-31</sup> and linear regressions. As depicted in Figure 2, the Bland-Altman approach plotted the difference between the two batch values of the same CSF sample against the average of the two values to detect any heterogeneity or trend. Linear regressions tested  $y = x$ , that is,  $y = 0 + 1x$ , where *x* was the CSF biomarker value assayed in one batch and *y* was the CSF biomarker value assayed in the other batch of the same sample. Batch values were considered equivalent if a linear regression met all the following criteria: (1) the test for intercept = 0 was not statistically significant; (2) the absolute value of the intercept was small, that is, within 5% of the difference between the minimum and maximum values of *y*, analogous to the conventional 5% type I error rate; (3) the test for slope = 1 was not statistically significant; (4) the absolute value of the slope was small, that is, within 5% of 1 (0.95 to 1.05); (5) model  $R^2 > .0.90$ , analogous to the cutoff value for excellent reliability<sup>37</sup>; and (6) plotted residuals showed random variation around 0 with constant variance. Regressions were performed in both directions by switching *x* and *y*.



**FIGURE 2** Bland-Altman plot for evaluating between-batch differences. We present this example to illustrate how to evaluate and interpret between-batch differences using the Bland-Altman plot. The differences between the two batch values of the biomarker sAPP- $\alpha$  for the same cerebrospinal fluid (CSF) sample (Batch 2–Batch 1) are plotted against the average of the two batch values. The dashed line represents the linear regression fit of the difference on the average. Mean  $\pm$  2 SD are limits of agreement (LOA). The mean of the differences is above zero, which indicates that batch 2 on average yielded higher values than batch 1. The negative slope of the linear fit indicates less variability in batch 2, and its crossing over the zero line indicates that batch 2 measurements exceeded batch 1 at the lower end of the range, but at higher levels, batch 1 exceeded batch 2. The variance of these batch differences was non-constant at different biomarker levels, with the largest variance observed in the middle of the biomarker level range. Under ideal circumstances, one can evaluate the significance of batch differences by comparing the LOA against an a priori threshold for a clinically significant difference. However, in this case we had no such a priori thresholds. Furthermore, the observed non-constant variance of the differences and its marked trend complicate such an interpretation. Consequently we focus on these descriptive aspects in our report

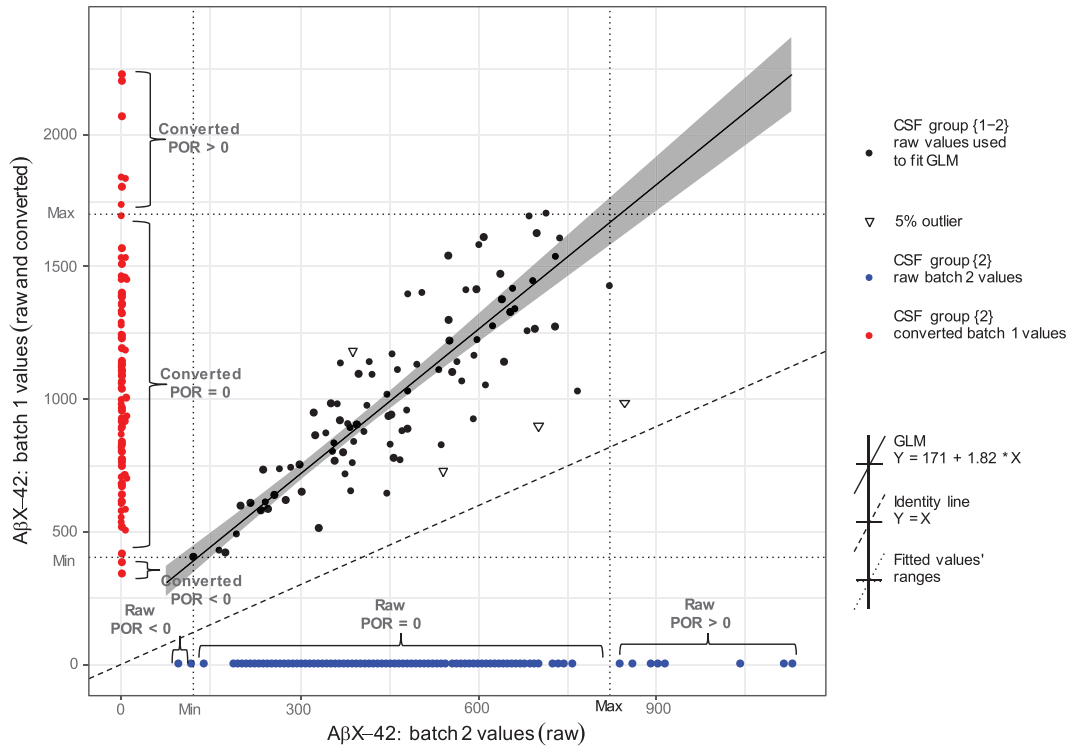
## 2.4.2 | Development of batch-to-batch conversion models

Conversion models were developed on CSF groups {1-2} and {1-3} separately, using generalized linear models (GLMs)<sup>38</sup> with the identity link function and linear relation between batches. The predictor was the value assayed in the batch to be converted *from*, and the outcome was the value assayed in the batch to be converted *to*. Conversion models were developed for both directions by switching the predictor and outcome to allow flexibility in choosing the converted batch as needed for a study. We applied three candidate distributions of the outcome, with each specifying a different variance function, including constant variance (normal), variance proportional to mean (VPM), and constant coefficient of variance (gamma), to accommodate non-normality and heteroscedasticity (ie., heterogeneity of variance).<sup>38</sup> Overdispersion was adjusted for VPM using the deviance statistic. Pearson residuals were plotted against the predicted values and fitted a locally estimated scatterplot smoothing (LOESS) line. The best distribution was chosen

by inspection of the residuals such that the LOESS line was most close to being flat at zero and the residuals were randomly scattered around zero with constant variance. Outliers were identified as having the top 5% largest Pearson residuals in the absolute value, and models were refitted with the outliers excluded. The model  $R^2$  was calculated following Zhang's approach based on the mean and variance functions.<sup>39</sup>

## 2.4.3 | Application of the conversion models and checking out of range values

Using A $\beta$ X-42 converting from batch 2 to batch 1 as an example (Figure 3), only CSF samples assayed in batch 2 but not batch 1 (group {2}) were converted. Conversion was done by applying the mean structure from the final GLM [converted A $\beta$ X-42 batch 1 value] = 171 + 1.82  $\times$  [raw A $\beta$ X-42 batch 2 value]. Converted values were checked for being "out of range" by examining if raw batch 2 values to convert (group {2}) were outside the range of raw batch 2 values used to fit the final GLM (group {1-2}), and if converted batch 1 values were outside the range



**FIGURE 3** AβX-42 converting from batch 2 to batch 1 and checking out of range values. Only samples assayed in batch 2 but not batch 1 (group {2}) were converted. Conversion was done by applying the mean structure from the final generalized linear model (GLM): [converted AβX-42 batch 1 value, represented with red dot] = 171 + 1.82 × [raw AβX-42 batch 2 value, represented with blue dot]. The final GLM model was developed using the raw batch 1 and raw batch 2 values of the group {1-2} (represented with black dot), with 5% outliers excluded (represented with triangle). The final GLM line deviated from the identity line Y = X, which indicated that raw batch 1 and batch 2 values were not comparable and conversion was important. We use x to represent the cerebrospinal fluid (CSF) value being checked, and use *min* and *max* to represent the minimum and maximum values of the range being compared to, that is, the range of the raw batch values used to fit the final GLM. POR indicates the extent a value is out of the range, calculated as a proportion of that range's width. A negative value represents below the range, a positive value represents above the range, and zero represents in the range. A larger absolute value indicates being further beyond the range. Thus,

$$\text{If } \min \leq x \leq \max, \text{ then } \text{POR} = 0$$

$$\text{If } x < \min, \text{ then } \text{POR} = \frac{(x - \min)}{(\max - \min)} \times 100\%, \text{ POR} < 0$$

$$\text{If } x > \max, \text{ then } \text{POR} = \frac{(x - \max)}{(\max - \min)} \times 100\%, \text{ POR} > 0$$

For each CSF value being converted, there are two sets of POR values, one for the raw value being converted and the other for the converted value. The two horizontal dotted lines parallel to the x-axis represent the minimum (*Min*) and maximum (*Max*) of the raw batch 1 values of the group {1-2}. If a converted value of the group {2} (represented with red dot) is between these two lines, it is in the range of the raw batch 1 values of the group {1-2}, Converted POR value = 0. If a converted value is below the *Min* line, it is below the range, Converted POR value < 0. If a converted value is above the *Max* line, it is above the range, Converted POR value > 0. The two vertical dotted lines parallel to the y-axis represent the *Min* and *Max* of the raw batch 2 values of the group {1-2}. If a raw value of the group {2} (represented with blue dot) is between these two lines, it is in the range of the raw batch 2 values of the group {1-2}, Raw POR = 0. If a raw value is below the *Min* line, it is below the range, Raw POR < 0. If a raw value is above the *Max* line, it is above the range, Raw POR > 0

of raw batch 1 values used to fit the model. Using a similar approach, conversions from batch 2 to batch 1, from 1 to 2, from 3 to 1, and from 1 to 3 were performed for each biomarker and ratio, by applying the corresponding GLM model to the CSF samples in each batch that were measured in the batch converting from but not in the batch converting to. Raw and converted values were next checked for being out of range.

Multiple longitudinal CSF samples from the same participant were treated as independent observations in the analyses, because this clustering would not impact the relationship between the two batch measures. All analyses were completed using SAS 9.4.<sup>40</sup>

### 3 | RESULTS

#### 3.1 | Between-batch differences

As summarized in Table 1, almost all 29 between-batch comparisons had significant differences as identified by at least two types of analyses, with 17 showing differences in all four analyses. The only exception was that no significant difference was found for NfL between batch 1 and batch 2 in any analyses. More detailed results are provided in Tables S2-S4 and Figure S1.

**TABLE 1** Between-batch comparison of the assayed CSF values for the same CSF sample group

CSF variable	n	Bland-Altman plot	Difference in means	Ratio of variances	Linear regression analyses for testing $y = x$				No. of methods	
					Batch 1 is y		Batch 1 is x			Model $R^2$
					Intercept	Slope	Intercept	Slope		
Comparison between batch 1 versus batch 2 for the CSF sample group {1-2}										
sAPP- $\alpha$	95	!	*	*	+		*+	*+	0.52	4
sAPP- $\beta$	95	!		*			*+	*+	0.45	3
A $\beta$ X-38	96	!	*	*			*+	*+	0.77	4
A $\beta$ X-40	96	!	*	*	+		*+	*+	0.70	4
A $\beta$ X-42	96	!	*	*	*+		*+	+	0.66	4
MCP-1	96	!	*	*			*+	*+	0.70	4
YKL-40	96		*				+	+	0.92	2
NfL	95								0.95	0
A $\beta$ 1-42	96	!					*+	*+	0.76	2
T-tau	95	!	*	*			*+	*+	0.96	4
P-tau	94		*		+	+	*+	*+	0.64	3
Ng	77	!	*	*	+	+	+	*+	0.66	4
A $\beta$ X-42/A $\beta$ X-40	96	!	*		*+	+		+	0.71	3
T-tau/A $\beta$ 1-42	95	!	*	*			*+	*+	0.97	4
P-tau/A $\beta$ 1-42	94	!	*	*	*+		*+	+	0.74	3
Comparison between batch 1 versus batch 3 for the CSF sample group {1-3}										
sAPP- $\alpha$	89	!	*		+		*+	*+	0.51	3
sAPP- $\beta$	89	!		*	+	+	*+	*+	0.49	3
A $\beta$ X-38	89	!	*	*			*+	+	0.88	4
A $\beta$ X-40	89	!	*	*			*+	*+	0.77	4
A $\beta$ X-42	89	!	*	*	*+		*+	+	0.72	4
MCP-1	89	!	*	*			*+	*+	0.71	4
YKL-40	89	!					*+	*+	0.84	2
NfL	89	!	*	*			*+	+	0.94	4
A $\beta$ 1-42	89	!	*	*	+		*+	*+	0.53	4
T-tau	89	!	*	*			*+	*+	0.93	4
P-tau	88	!	*		*+		*+	*+	0.16	3
A $\beta$ X-42/A $\beta$ X-40	89	!	*		*+		*+	+	0.54	3
T-tau/A $\beta$ 1-42	89	!	*	*			*+	*+	0.88	4
P-tau/A $\beta$ 1-42	88	!	*	*	*+		*+	+	0.44	4

NOTE Difference in means was tested using paired  $t$ -test with the null hypothesis  $M2 - M1 = 0$ . Ratio of variances was tested using the Pitman-Morgan test with the null hypothesis  $V2/V1 = 1$ . Model  $R^2$  is the same between the two regression models switching between  $x$  and  $y$ . Ng was not available for comparison between batch 1 versus batch 3 because it was not assayed in batch 3. No. of methods indicates the number of methods that have found batch differences.

!Batch differences were found in Bland-Altman plots.

\*Statistical significance, that is,  $P < (.05/29)$  for testing difference in means and ratio of variances, and  $P < (.05/58)$  for linear regression analyses testing intercept = 0 or slope = 1. Bonferroni correction was applied to adjust for the inflation of type I error rate due to multiple testing. The critical  $P$ -value was calculated by dividing the conventional  $P$ -value .05 by the number of tests.

+Effect size significance for linear regression analyses, that is,  $|\text{intercept}-0| > 5\%$  range of  $y$ , or  $|\text{slope}-1| > 5\%$  of 1. The statistical significance of a test depends on sample size, and the sample sizes were small. It was possible that a large deviation from 0 for the intercept or from 1 for the slope did not yield a significant  $P$ -value. Thus, the effect sizes of these deviations were also evaluated, using 5% as the cut-off value for significance, analogous to using 0.05 for the conventional cut off  $P$ -value, that is, 5% type I error rate. For the intercept, because the value range (ie., the difference between the minimum and maximum values) of  $y$  varies a lot between different biomarker variables, the same absolute amount of deviation can represent different extents of deviation for different biomarkers. Therefore, the intercept was compared against 5% of the value range of  $y$ .



**TABLE 2** Generalized linear models for between-batch conversion of CSF values

CSF variable	n	r	Distribution	Model R <sup>2</sup>	Distribution	Model R <sup>2</sup>
CSF sample group {1-2}			From batch 2 to batch 1		From batch 1 to batch 2	
sAPP- $\alpha$	95	0.72	Gamma	0.50	Normal	0.62
sAPP- $\beta$	95	0.67	Gamma	0.43	Gamma	0.50
A $\beta$ X-38	96	0.88	Gamma	0.86	Gamma	0.82
A $\beta$ X-40	96	0.84	Gamma	0.73	Normal	0.78
A $\beta$ X-42	96	0.81	Gamma	0.69	Normal	0.77
MCP-1	96	0.84	VPM	0.74	Gamma	0.79
YKL-40	96	0.96	Normal	0.95	Gamma	0.95
NfL	95	0.98	Normal	0.96	Normal	0.96
A $\beta$ 1-42	96	0.87	Gamma	0.74	Normal	0.81
T-tau	95	0.98	Normal	0.97	Gamma	0.96
P-tau	94	0.80	Gamma	0.66	Gamma	0.71
Ng	77	0.81	Gamma	0.70	VPM	0.75
A $\beta$ X-42/A $\beta$ X-40	96	0.85	Gamma	0.77	Gamma	0.77
T-tau/A $\beta$ 1-42	95	0.99	Normal	0.98	Normal	0.98
P-tau/A $\beta$ 1-42	94	0.86	Normal	0.70	Normal	0.75
CSF sample group {1-3}			From batch 3 to batch 1		From batch 1 to batch 3	
sAPP- $\alpha$	89	0.71	VPM	0.57	Normal	0.56
sAPP- $\beta$	89	0.70	VPM	0.59	Normal	0.57
A $\beta$ X-38	89	0.94	Normal	0.91	Gamma	0.88
A $\beta$ X-40	89	0.88	Gamma	0.81	Normal	0.82
A $\beta$ X-42	89	0.85	Normal	0.79	Normal	0.79
MCP-1	89	0.84	Normal	0.78	Normal	0.81
YKL-40	89	0.91	Normal	0.94	Normal	0.93
NfL	89	0.97	Gamma	0.96	Normal	0.96
A $\beta$ 1-42	89	0.73	Normal	0.57	Gamma	0.56
T-tau	89	0.97	Normal	0.95	Gamma	0.94
P-tau	88	0.40	Normal	0.25	Normal	0.17
A $\beta$ X-42/A $\beta$ X-40	89	0.74	Normal	0.64	Normal	0.65
T-tau/A $\beta$ 1-42	89	0.94	Normal	0.86	Normal	0.91
P-tau/A $\beta$ 1-42	88	0.66	Normal	0.52	Normal	0.35

NOTE. The conversion models were not developed for Ng for CSF sample {1-3} because it was not assayed in batch 3. *r* is Pearson correlation. All *r*s are statistically significant at  $P < (.05/29)$ .

Abbreviations: VPM, variance proportional to mean.

### 3.2 | Generalized linear models (GLMs) for conversion between batches

As summarized in Table 2, Pearson correlations were high and ranged from 0.66 to 0.99 with an average of 0.85, except for P-tau between batches 1 and 3,  $r = 0.40$ . For the 58 conversion models, 33 were fit with normal, 4 were fit with VPM, and 21 were fit with gamma distributions. Model  $R^2$  ranged between 0.17 and 0.98 with an average of 0.75. Model  $R^2$ s were very high for all four between-batch conversions for YKL-40 (0.93 to 0.95), NfL (all  $R^2 = 0.96$ ), T-tau (0.94 to 0.97), and T-tau/A $\beta$ 1-42 (0.86 to 0.98). The lowest Model  $R^2$ s ( $< 0.50$ ) were seen for P-tau conversion between batches 1 and 3 bi-directionally (0.17, 0.25),

P-tau/A $\beta$ 1-42 conversion from batches 1 to 3 (0.35), and sAPP- $\beta$  conversion from batches 2 to 1 (0.43). Conversion model intercepts, slopes, and 95% confidence intervals are provided in Table S5.

### 3.3 | Out of range raw and converted values

As shown in Table 3, of all 58 conversions, only 1 had neither raw nor converted values out of range (conversion of YKL-40 from batches 3 to 1), 10 had only raw but no converted values out of range, and the remaining 47 had both raw and converted values out of range. Among the four sets of conversions, conversion of the CSF group {3}

**TABLE 3** Counts and percentages of the raw and converted batch values that were out of range for the CSF samples being converted

CSF variable	n of converted	Raw	Converted	n of converted	Raw	Converted
Conversion from batch 2 to batch 1			Conversion from batch 1 to batch 2			
sAPP- $\alpha$	202	9 (4.5)	1 (0.5)	332	3 (0.9)	0 (0.0)
sAPP- $\beta$	202	7 (3.5)	0 (0.0)	332	5 (1.5)	1 (0.3)
A $\beta$ X-38	201	12 (6.0)	7 (3.5)	332	27 (8.1)	18 (5.4)
A $\beta$ X-40	201	9 (4.5)	8 (4.0)	332	14 (4.2)	12 (3.6)
A $\beta$ X-42	201	10 (5.0)	9 (4.5)	332	35 (10.5)	7 (2.1)
MCP-1	200	4 (2.0)	0 (0.0)	332	2 (0.6)	1 (0.3)
YKL-40	202	5 (2.5)	4 (2.0)	332	9 (2.7)	10 (3.0)
NfL	199	10 (5.0)	4 (2.0)	330	15 (4.5)	20 (6.1)
A $\beta$ 1-42	202	7 (3.5)	0 (0.0)	330	12 (3.6)	7 (2.1)
T-tau	202	21 (10.4)	21 (10.4)	330	39 (11.8)	38 (11.5)
P-tau	202	23 (11.4)	23 (11.4)	330	43 (13.0)	23 (7.0)
Ng	202	37 (18.3)	5 (2.5)	326	37 (11.3)	4 (1.2)
A $\beta$ X-42/A $\beta$ X-40	201	3 (1.5)	0 (0.0)	332	13 (3.9)	13 (3.9)
T-tau/A $\beta$ 1-42	202	14 (6.9)	14 (6.9)	330	38 (11.5)	33 (10.0)
P-tau/A $\beta$ 1-42	202	21 (10.4)	19 (9.4)	330	32 (9.7)	25 (7.6)
Conversion from batch 3 to batch 1			Conversion from batch 1 to batch 3			
sAPP- $\alpha$	162	3 (1.9)	0 (0.0)	338	23 (6.8)	0 (0.0)
sAPP- $\beta$	162	1 (0.6)	1 (0.6)	338	19 (5.6)	6 (1.8)
A $\beta$ X-38	162	3 (1.9)	3 (1.9)	339	14 (4.1)	11 (3.2)
A $\beta$ X-40	162	1 (0.6)	1 (0.6)	339	12 (3.5)	13 (3.8)
A $\beta$ X-42	162	3 (1.9)	0 (0.0)	339	40 (11.8)	28 (8.3)
MCP-1	162	9 (5.6)	6 (3.7)	339	7 (2.1)	7 (2.1)
YKL-40	162	0 (0.0)	0 (0.0)	339	20 (5.9)	20 (5.9)
NfL	162	4 (2.5)	3 (1.9)	337	48 (14.2)	52 (15.4)
A $\beta$ 1-42	162	4 (2.5)	5 (3.1)	337	55 (16.3)	6 (1.8)
T-tau	162	2 (1.2)	2 (1.2)	337	39 (11.6)	35 (10.4)
P-tau	162	1 (0.6)	1 (0.6)	337	28 (8.3)	7 (2.1)
A $\beta$ X-42/A $\beta$ X-40	162	2 (1.2)	0 (0.0)	339	33 (9.7)	2 (0.6)
T-tau/A $\beta$ 1-42	162	7 (4.3)	4 (2.5)	337	41 (12.2)	41 (12.2)
P-tau/A $\beta$ 1-42	162	3 (1.9)	0 (0.0)	337	55 (16.3)	40 (11.9)

NOTE. The conversions from and to batch 3 were not performed for Ng, because it was not assayed in batch 3 and thus no conversion models were developed.

from batches 3 to 1 generally had the lowest out of range percentages. On average, raw values had higher out of range percentages than converted values. Table S6 additionally summarizes the magnitude for being out of range.

## 4 | DISCUSSION

### 4.1 | Findings of between-batch differences

Almost all CSF biomarkers and ratios had non-trivial between-batch differences except for NfL being comparable between batches 1

and 2. Between-batch differences may be caused by subtle assay variations.<sup>5,12,21,23</sup> Assay variations can be minimized using specified acceptance criteria for the release of new kit lots, laboratories lot-to-lot bridging, and maintenance of instruments and internal and external QC samples. Nevertheless, subtle drift may occur before action is taken, due to replacement of kit lot, recalibration, or service of instrument, and so on. Moreover, for many biomarkers there are no certified reference materials that can be used to recalibrate assays rendering longitudinal stability of the measurements difficult to maintain. Between-batch conversions are thus important.



## 4.2 | Recommendations for conversion models

We provide the following recommendations with the application of the conversion models developed in this study to our center's research as an example.

First, researchers should evaluate and understand the batch differences in the assayed CSF biomarker values. Re-assaying a subset of samples in subsequent batches would allow such evaluation. A comprehensive evaluation can be performed with a combination of examining Bland-Altman plots, testing differences in means with the paired *t* test, testing ratios of variances using the Pitman-Morgan test, and testing value equivalence with linear regressions. Each method assesses the batch differences from a varying perspective. Whenever possible, biomarker measurement should be performed using internal QC samples that are the same between runs. This will allow for detection of larger drifts in the measurement (often  $\pm 10\%$  but this may vary across biomarkers and laboratories).

Second, when CSF biomarker values are comparable across batches, raw values can be combined directly in analyses. For example, in this study, raw NFL values from batches 1 and 2 can be combined given their batch comparability.

Third, for CSF biomarker values that have significant batch differences, statistical conversion models can be developed using GLMs to adjust for the batch differences. Some conversion models may yield low  $R^2$ , warranting caution when conducting analyses with these converted values included. When possible, a sub-sample consisting of raw values from only one batch can be analyzed instead, or at least serve as a sensitivity analysis.

Fourth, the conversion direction (eg., either from batch 1 to batch 2, or from 2 to 1) should be chosen such that more raw batch values will be included in the analysis, because converted values include conversion model prediction errors even if those errors are minimal, whereas raw values do not. In this study, the conversion models were developed for both directions to allow flexibility, given that studies may have more raw values in different batches.

Fifth, for an analysis that includes CSF samples from multiple study cohorts, it is possible that no CSF samples are repeatedly assayed in a pair of batches, and thus conversion models cannot be developed between these batches. If conversion is necessary, a chained indirect conversion that involves the application of more than one conversion model can be applied. For example, in this study no direct conversion was developed between batches 2 and 3 because no CSF samples were assayed in both of the batches. Conversion from batches 2 to 3 could be achieved by converting from 2 to 1, and then converting from 1 to 3. However, such indirect conversion is not recommended because of multiplied chances in prediction errors.

Finally, GLMs may have limited predictive ability beyond the data range of the sample that the conversion model is based on. Thus the extent to which it is out of range can be provided for each value being converted and used for assessing the utility of the converted values.

## 4.3 | Strengths

Previous studies<sup>12,17,22,23</sup> examining CSF measurement variability focused mainly on three core AD biomarkers— $A\beta_{1-42}$ , T-tau, P-tau, whereas we studied an extended panel of 12 AD biomarkers and 3 ratios. In addition to comprehensively evaluating between-batch differences with multiple approaches, we further developed statistical conversion models to adjust for batch differences using GLMs. Such harmonization would allow valid prediction of AD risks and longitudinal trajectories in CSF biomarker changes. Compared to including the batch indicator as a covariate, the GLM is applicable when the CSF biomarker variable serves as a predictor and when the batch differences are not constant. Compared to linear regression, GLM allows non-normality and heteroscedasticity,<sup>38</sup> and thus does not require transformation, which eases data interpretation and is applicable to more than two batches.

## 4.4 | Limitations

Three CSF biomarkers including  $A\beta_{1-42}$ , T-tau, and P-tau were measured using different assay methods (INNOTEST vs XMap) between batches 1 and 3. Unfortunately, the assay difference could not be separately assessed from the batch difference in the current study, because each batch was measured with only one assay method and vice versa. Of interest, the conversion models  $R^2$ s were high for T-tau (0.95 and 0.94), but were the lowest for P-tau (0.17 and 0.25). Measuring factors that contribute to the measurement variability within the same batch and evaluating how they behave differently between T-tau and P-tau may shed light on this finding and improve the conversion models for P-tau.

Similar to any other regression-based approaches, GLM suffers reduced variability in the converted values associated with regression toward the mean. A non-regression approach is z-score transformation,<sup>41</sup> which would convert each CSF sample's raw value to standardized value using the mean and standard deviation of all CSF samples included in the batch the sample was assayed in. Ideally, CSF samples measured in different batches could have similar distributions (e.g., similar means and standard deviations) of the underlying true biomarker levels. Correspondingly, a sample would stand at a similar percentile in the distribution and have a similar z-score if it were measured in a different batch, allowing the possible harmonization using z-score. However, in reality, batches usually include samples collected from different studies and from participants with different ages and diagnoses (eg., as shown in the current study), and thus have different distributions, thereby limiting the application of this method. Nonetheless, this approach can be applicable for harmonizing different measurement methods for large epidemiological studies, in which participants measured by each method are approximately a random sample from the population. In summary, continued research is needed to develop better solutions for adjusting batch differences.

Compared to the CSF groups used for developing conversion models, the CSF groups being converted had larger sample sizes (ratio of ns ranged from 1.9 to 4.4), had older mean ages with greater variability, and had mixed clinical diagnoses (except for group [3]), and thus tended to have more variability in the CSF biomarker levels and yielded out of range values. Conversion accuracy may be reduced for out of range values, as it is functionally extrapolating beyond the range of the fitted data. In future study design, it will be helpful to expand the value range of the CSF samples used for developing conversion models by including samples collected at broader age span and diverse clinical diagnosis, in order to make the samples for developing conversion models more comparable with the samples to be converted.

Ideally, we would have additional CSF samples that were measured in both batches but not used for model development to apply the conversion model and evaluate prediction accuracy. However, because the number of repeatedly measured CSF samples was limited, all such samples were used for model development to minimize random errors associated with small sample size and best capture the between-batch relations. Such validation is encouraged for future conversion model development if sample size is sufficient to allow the split between the training (ie., model development) versus testing (ie., model validation) subsamples. In addition, the performance of conversion models could be evaluated using simulation studies, which would generate multiple samples for each scenario of interests, and estimate a conversion model on each sample to produce an empirical distribution of model parameter estimates. Performance of the conversion would then be evaluated based on statistical properties of this empirical distribution in terms of accuracy (ie., bias) and efficiency (ie., empirical standard error).

## 5 | CONCLUSIONS

In summary, researchers should first try their best to assay CSF samples in the same batch and minimize factors that contribute to measurement variability by following standard procedures and methods for CSF collection, processing, and assay. Between-batch conversion using statistical models can serve as a post hoc treatment to harmonize the CSF values assayed in different batches to be included in the same analysis if necessary. However, such statistical harmonization does not lessen the importance of the aforementioned effort to minimize batch differences upfront. The fully automated platforms such as the Elecsys immunoassays are promising in controlling for batch variability, as evidence has shown for Abeta1-42, T-tau, and P-tau.<sup>42-44</sup> However, as novel CSF biomarkers are being developed, and it takes time for such platforms to become available for all biomarkers, statistical conversion models remain important for AD studies that include CSF data. Furthermore, the approaches for assessing and accounting for batch differences in the current study have the potential application for evaluating and adjusting for other sources of variability in measuring CSF biomarkers beyond batch difference, or studying other types of biomarkers such as imaging data.

## ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health (NIH) National Institute on Aging (NIA) [P50AG033514 (PI Sanjay Asthana, MD, FRCP), P30AG062715 (PI Sanjay Asthana, MD, FRCP), R01AG027161 (PI Sterling C. Johnson, PhD), UL1TR000427 (PI Sterling C. Johnson, PhD), R01AG021155 (PI Sterling C. Johnson, PhD), R01AG037639 (PI Barbara B. Bendlin, PhD), R01AG031790 (PI Cynthia M. Carlsson, MD, MS), and R01AG054059 (PI Carey E. Gleason, PhD)]. The NIH did not play a role in the study design, collection, analysis and interpretation of data, writing of the report, or decision to submit this manuscript for publication. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. Dr. Henrik Zetterberg is a Wallenberg Scholar supported by grants from the Swedish Research Council (#2018-02532), the European Research Council (#681712), Swedish State Support for Clinical Research (#ALFGBG-720931), the Alzheimer Drug Discovery Foundation (ADDF), USA (#201809-2016862), and the UK Dementia Research Institute at UCL. Dr. Kaj Blennow 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), and European Union Joint Program for Neurodegenerative Disorders (JPND2019-466-236). We gratefully acknowledge the participants and research teams of the ADRC, WRAP, PREDICT, PIPR, LEAD, and SHARP studies for their contributions. We thank the Geriatric Research, Education and Clinical Center (GRECC) at the William S. Middleton Memorial Veterans Hospital for providing support and resources for this study. This is GRECC manuscript number 007-2020.

## CONFLICTS OF INTEREST

Dr. Cynthia Carlsson receives grant support from NIH/Lilly, NIH, Veterans Affairs, and Bader Philanthropies. Dr. Sterling Johnson previously served on the advisory board for Roche Diagnostics. Dr. Sanjay Asthana serves as a site PI for pharmaceutical trials funded by Merck Pharmaceuticals, Lundbeck, NIH/UCSD, Eisai, and Genentech Inc. Dr. Richard Chappell serves on data safety monitoring boards for Axsome Pharmaceuticals and TGR Pharmaceuticals and has recently had a speaking engagement at Merck, Inc. Dr. Henrik Zetterberg has served at scientific advisory boards for Denali, Roche Diagnostics, Wave, Samumed, and CogRx; has given lectures in symposia sponsored by Fujirebio, Alzecure, and Biogen; and is a co-founder of Brain Biomarker Solutions in Gothenburg AB (BBS), which is a part of the GU Ventures Incubator Program. Dr. Kaj Blennow has served as a consultant or on advisory boards for Abcam, Axon, Biogen, Lilly, MagQu, Novartis, and Roche Diagnostics, and is a co-founder of Brain Biomarker Solutions in Gothenburg AB (BBS), which is a part of the GU Ventures Incubator Program.

## REFERENCES

1. Paterson RW, Slattery CF, Poole T, et al. Cerebrospinal fluid in the differential diagnosis of Alzheimer's disease: clinical utility of an extended panel of biomarkers in a specialist cognitive clinic. *Alzheimers Res Ther.* 2018;10:32.
2. Olsson B, Lautner R, Andreasson U, et al. CSF and blood biomarkers for the diagnosis of Alzheimer's disease: a systematic review and meta-analysis. *Lancet Neurol.* 2016;15:673-684.
3. Jack CR, Knopman DS, Jagust WJ, et al. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *Lancet Neurol.* 2010;9:119-128.
4. Jack CR, Knopman DS, Jagust WJ, et al. Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. *Lancet Neurol.* 2013;12:207-216.
5. Mattsson N, Zetterberg H, kaj Blennow. Lessons from multicenter studies on CSF biomarkers for Alzheimer's disease. *Int J Alzheimers Dis.* 2010;2010:610-613.
6. Blennow K, Shaw LM, Stomrud E, et al. Predicting clinical decline and conversion to Alzheimer's disease or dementia using novel Elecsys A $\beta$ (1-42), pTau and tTau CSF immunoassays. *Sci Rep.* 2019;9:19024.
7. Jack CR, Bennett DA, Blennow K, et al. NIA-AA research framework: toward a biological definition of Alzheimer's disease. *Alzheimers Dement.* 2018;14:535-562.
8. Blennow K, Zetterberg H. Biomarkers for Alzheimer's disease: current status and prospects for the future. *J Inter Med.* 2018;284:643-663.
9. Brookmeyer R, Abdalla N. Estimation of lifetime risks of Alzheimer's disease dementia using biomarkers for preclinical disease. *Alzheimer's Dement.* 2018;14:981-988.
10. Sperling RA, Aisen PS, Beckett LA, et al. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's Dement.* 2011;7:280-292.
11. Carrillo MC, Blennow K, Soares H, et al. Global standardization measurement of cerebral spinal fluid for Alzheimer's disease: an update from the Alzheimer's Association Global Biomarkers Consortium. *Alzheimer's Dement.* 2013;9:137-140.
12. Mattsson N, Andreasson U, Persson S, et al. The Alzheimer's Association external quality control program for cerebrospinal fluid biomarkers. *Alzheimer's Dement.* 2011;7:386-395.e6.
13. Vanderstichele H, Bibl M, Engelborghs S, et al. Standardization of preanalytical aspects of cerebrospinal fluid biomarker testing for Alzheimer's disease diagnosis: a consensus paper from the Alzheimer's Biomarkers Standardization Initiative. *Alzheimer's Dement.* 2012;8:65-73.
14. Lewczuk P, Beck G, Ganslandt O, et al. International quality control survey of neurochemical dementia diagnostics. *Neurosci Lett.* 2006;409:1-4.
15. Shaw LM, Vanderstichele H, Knapik-Czajka M, et al. Qualification of the analytical and clinical performance of CSF biomarker analyses in ADNI. *Acta Neuropathol.* 2011;121:597-609.
16. Dumurgier J, Vercruyse O, Paquet C, et al. Intersite variability of CSF Alzheimer's disease biomarkers in clinical setting. *Alzheimer's Dement.* 2013;9:406-413.
17. Mattsson N, Andreasson U, Persson S, et al. CSF biomarker variability in the Alzheimer's Association quality control program. *Alzheimers Dement.* 2013;9:251-261.
18. Teunissen CE, Verwey NA, Kester MI, van Uffelen K, Blankenstein MA. Standardization of assay procedures for analysis of the CSF biomarkers amyloid  $\beta$  (1-42), tau, and phosphorylated tau in Alzheimer's disease: report of an International Workshop. *Int J Alzheimers Dis.* 2010;2010:635053.
19. Lewczuk P, Kornhuber J, Vanderstichele H, et al. Multiplexed quantification of dementia biomarkers in the CSF of patients with early dementias and MCI: a multicenter study. *Neurobiol Aging.* 2008;29:812-818.
20. Vos SJB, Visser PJ, Verhey F, et al. Variability of CSF Alzheimer's disease biomarkers: implications for clinical practice. *PLoS One.* 2014;9:e100784.
21. Verwey NA, van der Flier WM, Blennow K, et al. A worldwide multicentre comparison of assays for cerebrospinal fluid biomarkers in Alzheimer's disease. *Ann Clin Biochem.* 2009;46:235-240.
22. Bouwman FH, van der Flier WM, Schoonenboom NS, et al. Usefulness of longitudinal measurements of {szligbeta}-Amyloid sub(1-42) in cerebrospinal fluid of patients with various cognitive and neurologic disorders. *Clin Chem.* 2006;52:1604-1606.
23. Verwey NA, Bouwman FH, Van der Flier WM, Veerhuis R, Scheltens P, Blankenstein MA. Variability in longitudinal cerebrospinal fluid tau and phosphorylated tau measurements. *Clin Chem Lab Med.* 2008;46:1300-1304.
24. Kuhlmann J, Andreasson U, Pannee J, et al. CSF A $\beta$ 1-42—an excellent but complicated Alzheimer's biomarker—a route to standardisation. *Clin Chim Acta.* 2017;467:27-33.
25. Altman DG, Bland JM. Measurement in medicine: the analysis of method comparison studies. *J Royal Stat Soc Series D.* 1983;32:307-317.
26. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet North Am Ed.* 1986;1:307-310.
27. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res.* 1999;8:135-160.
28. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Int J Nurs Stud.* 2010;47:931-936.
29. Pitman EJG. A note on normal correlation. *Biometrika.* 1939;31:9-12.
30. Morgan WA. A test for the significance of the difference between the two variances in a sample from a normal bivariate population. *Biometrika.* 1939;31:13-19.
31. Lee J. Comparison of variance between correlated samples. *Bioinformatics.* 1992;8:405-456.
32. Ludbrook J. Linear regression analysis for comparing two measurers or methods of measurement: but which regression?: linear regression for comparing methods. *Clin Exp Pharmacol Physiol.* 2010;37:692-629.
33. Zaki R, Bulgiba A, Ismail R, Ismail NA. Statistical methods used to test for agreement of medical instruments measuring continuous variables in method comparison studies: a systematic review. *PLoS One.* 2012;7:e37908.
34. Norton D. CSF predicted values between batch 1 and batch 2; 2016.
35. Wang L-S, Leung YY, Chang S-K, et al. Comparison of xMAP and ELISA assays for detecting cerebrospinal fluid biomarkers of Alzheimer's disease. *J Alzheimers Dis.* 2012;31:439-445.
36. Mattsson N, Zetterberg H, Hansson O, et al. CSF biomarkers and incipient Alzheimer disease in patients with mild cognitive impairment. *JAMA.* 2009;302:385-393.
37. George D, Mallery P. *SPSS for Windows Step by Step: A Simple Guide and Reference, 11.0 Update.* 4th Ed. Boston, MA: Allyn & Bacon; 2003.
38. McCullagh P, Nelder JA. *Generalized Linear Models.* Second Edition. London NY: Chapman and Hall; 1989.
39. Zhang D. A coefficient of determination for generalized linear models. *Am Stat.* 2017;71:310-316.
40. SAS. Cary, NC: SAS Institute Inc.; 2014.
41. Burke Ó, Benton S, Szafranski P, et al. Extending the scope of pooled analyses of individual patient biomarker data from heterogeneous laboratory platforms and cohorts using merging algorithms. *Pregnancy Hypertens.* 2016;6:53-59.
42. Gille B, Dedeene L, Stoops E, et al. Automation on an open-access platform of Alzheimer's disease biomarker immunoassays. *SLAS Technol.* 2018;23:188-197.

43. Bittner T, Zetterberg H, Teunissen CE, 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*. 2016;12:517-526.
44. Lifke V, Kollmorgen G, Manuilova E, et al. Elecsys® Total-Tau and Phospho-Tau (181P) CSF assays: analytical performance of the novel, fully automated immunoassays for quantification of tau proteins in human cerebrospinal fluid. *Clin Biochem*. 2019;72:30-38.
45. Morris JC, Weintraub S, Chui HC, et al. The Uniform Data Set (UDS): clinical and cognitive variables and descriptive data from Alzheimer disease centers. *Alzheimer Dis Assoc Disord*. 2006;20:210-216.
46. Johnson SC, Kosik RL, Jonaitis EM, et al. The Wisconsin Registry for Alzheimer's Prevention: a review of findings and current directions. *Alzheimer's Dement*. 2018;10:130-142.
47. Gepner AD, Lazar K, Hulle CV, Korcarz CE, Asthana S, Carlsson CM. Effects of simvastatin on augmentation index are transient: out-

comes from a randomized controlled trial. *J Am Heart Assoc*. 2019;8:e009792.

#### SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**How to cite this article:** Ma Y, Norton DL, Van Hulle CA, et al. Measurement batch differences and between-batch conversion of Alzheimer's disease cerebrospinal fluid biomarker values. *Alzheimer's Dement*. 2021;13:e12194. <https://doi.org/10.1002/dad2.12194>