

Published in final edited form as:

Alzheimers Dement. 2013 May ; 9(3): 251–261. doi:10.1016/j.jalz.2013.01.010.

CSF biomarker variability in the Alzheimer's Association quality control program

Niklas Mattsson^{a,b,*}, Ulf Andreasson^a, Staffan Persson^a, Maria C. Carrillo^c, Steven Collins^d, Sonia Chalbot^e, Neal Cutler^f, Diane Dufour-Rainfray^g, Anne M. Fagan^h, Niels H. H. Heegaardⁱ, Ging-Yuek Robin Hsiung^j, Bradley Hyman^k, Khalid Iqbal^e, D. Richard Lachno^l, Alberto Lleó^m, Piotr Lewczukⁿ, José L. Molinuevo^o, Piero Parchi^p, Axel Regeniter^q, Robert Rissman^r, Hanna Rosenmann^s, Giuseppe Sancesario^t, Johannes Schröder^u, Leslie M. Shaw^v, Charlotte E. Teunissen^w, John Q. Trojanowski^v, Hugo Vanderstichele^x, Manu Vandijck^y, Marcel M. Verbeek^z, Henrik Zetterberg^{a,aa}, Kaj Blennow^a, and Stephan A. Käser^{bb} on behalf of the Alzheimer's Association QC Program Work Group

^aClinical Neurochemistry Laboratory, Institute of Neuroscience and Physiology, Department of Psychiatry and Neurochemistry, the Sahlgrenska Academy at the University of Gothenburg, Sahlgrenska University Hospital, Mölndal, Sweden ^bSan Francisco VA Medical Center, Center for Imaging of Neurodegenerative Diseases, University of California San Francisco, San Francisco, CA, USA ^cThe Alzheimer's Association, Chicago, IL, USA ^dDepartment of Pathology, Florey Institute of Neuroscience and Mental Health, The University of Melbourne, Melbourne, Australia ^eNew York State Institute for Basic Research, Staten Island, NY, USA ^fWorldwide Clinical Trials, Beverly Hills, CA, USA ^gCHRU de Tours, Laboratoire de médecine nucléaire; Université François-Rabelais de Tours, PRES Centre-Val de Loire Université; and UMR INSERM U 930 Imagerie et cerveau, Université François-Rabelais de Tours, PRES Centre-Val de Loire Université, Tours, France ^hWashington University, St. Louis, MO, USA ⁱDepartment of Clinical Biochemistry, Immunology & Genetics, Statens Serum Institut, Copenhagen, Denmark ^jUniversity of British Columbia, Vancouver, Canada ^kMassGeneral Institute for Neurodegenerative Disease, Charlestown, MA, USA ^lEli Lilly and Company, Windlesham, UK ^mHospital de Sant Pau, Barcelona, Spain ⁿUniversitätsklinikum Erlangen, Erlangen, Germany ^oAlzheimer's Disease and Other Cognitive Disorders Unit, Neurology Service, ICN Hospital Clinic i Universitari and Pasqual Maragall Foundation, Barcelona, Spain ^pInstitute of Neurological Sciences and Department of Biomedical and Neuromotor Sciences, University of Bologna, Bologna, Italy ^qBasel University Hospital, Basel, Switzerland ^rSchool of Medicine, University California, San Diego, La Jolla, CA, USA ^sHadassah Hebrew University Medical Center, Jerusalem, Israel ^tGeneral Hospital of University of Rome Tor Vergata, Rome, Italy ^uSektion Gerontopsychiatrie Universitätsklinikum, Heidelberg, Germany ^vUniversity of Pennsylvania, Philadelphia, PA, USA ^wDepartment of Clinical Chemistry, VU University Medical Center, Amsterdam, Netherlands ^xADx NeuroSciences, Gent, Belgium ^yInnogenetics nv (part of Fujirebio), Gent, Belgium ^zDepartments of Neurology and Laboratory Medicine, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Centre, Nijmegen, Netherlands ^{aa}Institute of Neurology, University College London, Queen Square, London, UK ^{bb}University of Tübingen, Tübingen, Germany

Abstract

© 2013 The Alzheimer's Association. All rights reserved.

*Corresponding author. Tel.: +1-415-494-1139; Fax: +1-415-668-2864. niklas.mattsson@neuro.gu.se.

K. B., H. Z., N. M., and U. A. designed the study. N. M. and U. A. performed statistical analyses. N. M. drafted the manuscript. S. P. was the study coordinator. All authors participated in interpretation of data, revised the manuscript for intellectual content, and approved the final version.

Background—The cerebrospinal fluid (CSF) biomarkers amyloid beta 1–42, total tau, and phosphorylated tau are used increasingly for Alzheimer’s disease (AD) research and patient management. However, there are large variations in biomarker measurements among and within laboratories.

Methods—Data from the first nine rounds of the Alzheimer’s Association quality control program was used to define the extent and sources of analytical variability. In each round, three CSF samples prepared at the Clinical Neurochemistry Laboratory (Möln dal, Sweden) were analyzed by single-analyte enzyme-linked immunosorbent assay (ELISA), a multiplexing xMAP assay, or an immunoassay with electrochemoluminescence detection.

Results—A total of 84 laboratories participated. Coefficients of variation (CVs) between laboratories were around 20% to 30%; within-run CVs, less than 5% to 10%; and longitudinal within-laboratory CVs, 5% to 19%. Interestingly, longitudinal within-laboratory CV differed between biomarkers at individual laboratories, suggesting that a component of it was assay dependent. Variability between kit lots and between laboratories both had a major influence on amyloid beta 1–42 measurements, but for total tau and phosphorylated tau, between-kit lot effects were much less than between-laboratory effects. Despite the measurement variability, the between-laboratory consistency in classification of samples (using pre hoc-derived cutoffs for AD) was high (>90% in 15 of 18 samples for ELISA and in 12 of 18 samples for xMAP).

Conclusions—The overall variability remains too high to allow assignment of universal biomarker cutoff values for a specific intended use. Each laboratory must ensure longitudinal stability in its measurements and use internally qualified cutoff levels. Further standardization of laboratory procedures and improvement of kit performance will likely increase the usefulness of CSF AD biomarkers for researchers and clinicians.

Keywords

Alzheimer’s disease; Cerebrospinal fluid; Biomarkers; External assurance; Quality control; Proficiency testing

1. Introduction

Cerebrospinal fluid (CSF) examination in Alzheimer’s disease (AD) typically shows reduced levels of amyloid β 1–42 (A β 42), and increased levels of total tau (T-tau) and phosphorylated tau (P-tau) [1–3]. The presence of this CSF pattern has recently been proposed for use in the research diagnostic criteria for AD [4–7]. Clinical diagnostic testing of CSF samples is already available from several hospital laboratories as well as from commercial laboratories. The measured biomarker levels, however, differ among studies, which may be the result of a number of preanalytical, analytical, or assay-related factors [8–10]. To overcome this situation, several standardization efforts have been initiated to harmonize laboratory procedures [11], give guidelines on CSF collection and handling procedures [12], define reference measurement procedures [13], and construct reference materials for assay calibration [14].

The Alzheimer’s Association launched an international quality control (QC) program for CSF biomarkers in 2009 [15]. The program was established to monitor total analytical variability for A β and tau proteins in CSF, to provide a network where sources of variation could be identified, and to implement actions originating from standardization efforts. There are no requirements or obligations to become a participant for the QC program other than using a commercially available assay for A β or tau. Three complete rounds of samples, each including two round-specific samples and one longitudinal sample that remains the same over years, are prepared at the Clinical Neurochemistry Laboratory in Möln dal, Sweden, and

shipped yearly to participating laboratories. Moreover, five experienced laboratories that process large numbers of samples routinely serve as reference laboratories and analyze the samples multiple times. The results from the first two rounds, involving 40 laboratories, have been described previously [15].

Herein, we report the development of the program during 2010 to 2012, and describe results through to program round 9. During this time, the number of participating sites doubled, and the large amount of data collected increased our capability to identify sources of measurement variability, including differences between laboratories and between lots of analytical kits.

2. Methods

2.1. CSF samples and laboratory procedures

As reported previously [15], human CSF pools were prepared in Mölndal, Sweden, from a large number of fresh, de-identified samples obtained during routine clinical workflow (all samples underwent one freeze/thaw cycle before pooling). No extra amount (spiking) of analyte was added to the samples. The pools were prepared by experienced and certified laboratory technicians during continuous mixing to ensure homogeneity of the pools. The total volumes of the pools were 75 to 1500 mL. The pools were divided into 500- μ L aliquots in polypropylene screw-cap tubes (art. no. 72.692, 1.5 mL; Sarstedt AG & Co., Nümbrecht, Germany; except for samples 2011-6A, 2011-7A, 2012-8B, and 2012-9B, for which we used art. no. 72.730.007, 0.5 mL; Sarstedt AG & Co.). The samples were refrozen at -80°C , followed by distribution to the participating laboratories on dry ice by courier. All shipments included three samples. Two (blinded challenge samples) were specific to the round (designated 2009-1A, 2009-1B, 2010-2A, 2010-2B, and so forth), and one sample (quality control longitudinal sample [QC-L]) was from a pool used to evaluate longitudinal stability (used until round 7 [total shelf life of the sample, 26 months], when it was discontinued because of a supply shortage and was exchanged for a new longitudinal sample). The blinded challenge samples differed in their AD biomarker profiles (Fig. 1).

All laboratories verified that the samples had arrived frozen. The analyses were done by each participant in duplicate as part of their routine laboratory activities. No extra freeze/thawing of samples was allowed. The reference laboratories (located in Amsterdam, Mölndal, Erlangen, Ghent, and Pennsylvania) analyzed the samples six times (with one aliquot per run) using different plates to assess within-laboratory precision. All results were reported back to Mölndal for data analysis together with a questionnaire that gave an overview of the applied materials and handling procedures for the specific run for data analysis on the reported results.

2.2. Participating laboratories and assay systems

The size and exposure of the Alzheimer's Association QC program has grown continuously since its start in 2009. The majority of the participants use INNOTEST enzyme-linked immunosorbent assays (ELISAs; $n = 61$ in round 9) or bead-based xMAP platforms with the INNO-BIA AlzBio3 (both Innogenetics, Gent, Belgium; www.innogenetics.com; $n = 12$ in round 9) to quantify A β 42, T-tau and P-tau (181P) (or simply P-tau). Meso Scale Discovery (MSD; Gaithersburg, MD; www.mesoscale.com) technology was used by a smaller number of laboratories ($n = 8$ in round 9) for A β N-42, A β N-40, and A β N-38 (A β triplex). MSD A β triplex was used with either 4G8 (epitope A β 17–24) or 6E10 (epitope A β 9–12) as detection antibodies. The volume of provided samples (500 μ L) was sufficient to allow for duplicate analyses of the sample with ELISA (T-tau, $2 \times 25 \mu\text{L}$; A β 42, $2 \times 25 \mu\text{L}$; and P-tau, $2 \times 75 \mu\text{L}$), xMAP ($2 \times 75 \mu\text{L}$), MSD (A β triplex $2 \times 25 \mu\text{L}$), or combinations thereof. Several laboratories ($n = 9$, 13% in round 9) used multiple techniques. Note that samples were

analyzed as part of the laboratories' routine activities, and a large total number of different production lots of analytical kits were used throughout the program. The total numbers of different kit lots used were 44 for ELISA A β 42, 39 for ELISAT-tau, 33 for ELISA P-tau, 21 for xMAP, and 29 for MSD. However, some lots were overrepresented in the program (about 50% of measurements for each analyte were done using only seven different kit lots for ELISA A β 42, seven for ELISA T-tau, five for ELISA P-tau, five for xMAP, and eight for MSD).

2.3. Estimates of variability

The overall variability of attained results may be described by the coefficient of variation (CV; standard deviation \times 100 divided by the mean) for each sample and assay. Some of the variables are the responsibility of the vendors of the assays, whereas other variables are considered to be responsibility of the performing laboratory. The overall variability is affected by several different factors, including within-assay run variability (between duplicate samples), within-laboratory longitudinal variability, between-laboratory variability, and within- and between-assay kit lot variability. Variability depends also on a combination of trueness (bias, systematic deviation from a reference value) or precision (imprecision, random deviation from a value). In this study, we aimed to estimate the size and source of these different types of variability.

2.4. Statistical analysis

Biomarker results were analyzed statistically and grouped by rounds, samples, and analytical techniques. Mean levels, standard deviations, and CVs were calculated. Between-group differences were assessed using nonparametric tests (Mann-Whitney *U* or Kruskal-Wallis tests). Analysis of variance was performed using restricted maximum likelihood estimation of covariances (the estimated variance components were between-laboratory and between-batch lot variability). SPSS version 20 (IBM Corporation, Armonk, NY, USA) and GraphPad Prism 5 (Graph-Pad Software Inc., La Jolla, CA, USA) were used.

3. Results

3.1. Overall variability

The overall CV was 20% to 30% for most assays and samples. All mean levels, standard deviations, and CVs for blinded test samples are presented in Fig. 1. For ELISAs, mean CV was 23% (range, 17%–29%) for A β 42, 18% (range, 12%–27%) for T-tau, and 19% (range, 12%–28%) for P-tau. For xMAP, mean CV was 28% (range, 17%–38%) for A β 42, 20% (range, 13%–28%, after removal of one significant outlier, see Fig. 1) for T-tau, and 21% (range, 11%–30%) for P-tau. For MSD, mean CV was 24% (range, 13%–36%) for A β 42, 26% (range, 16%–37%) for A β 40, and 27% (range, 10%–60%) for A β 38. These data combined MSD assays using different A β detection antibodies (see Supplemental Fig. 1 for MSD data stratified by antibody).

3.2. Within-run variability

In rounds 4 to 7, the laboratories reported within-run variability as CV of duplicate measurements for the QC-L sample. Median within-run CV was less than 4% for ELISA, 1.9% to 7.4% for xMAP, and 1.5% to 17% for MSD assays (17% was an outlier for the MSD assays, for which most within-run CVs were less than 10%; see Table 1). No trend in the within-run variability over the study was noted, which could indicate that all laboratories, independent from their experience level, have a comparable within-run variability.

3.3. Longitudinal variability

Longitudinal variability was estimated separately at the five reference laboratories (using several different samples measured at six different time points) and at all laboratories (using the QC-L sample at laboratories participating in at least three rounds; Fig. 2).

At the reference laboratories, the mean longitudinal within-laboratory CV was 8% to 13% for ELISA ($n = 4$) and 5% to 17% for xMAP measurements ($n = 3$). There were differences in CVs between the reference laboratories and also between analytes at the same laboratory. For example, the variability for xMAP P-tau was very high at reference laboratory 3 (Fig. 2B). The cause for this is unknown, but we verified that it did not depend on single outliers, or errors in reporting results, and that the CVs for simultaneous measurements of xMAP A β 42 and T-tau were not elevated, the latter suggesting that assay-dependent factors rather than factors related to laboratory procedures were important.

The within-laboratory longitudinal CVs at all participating laboratories were often higher than the CVs at the reference laboratories (12%–19%, Fig. 2), with the highest CV seen for A β 42. The overall variability for the QC-L samples was approximately 20% to 30% (comparable with the blinded test sample results described earlier, Fig. 1), with no significant change over time in mean concentrations (Fig. 3). This result supports that the QC-L samples were stable during storage at -80°C for 26 months. However, we noted that the variability was lower among the reference laboratories than among all laboratories, especially for A β 42 (Fig. 3A, B).

3.4. Between-laboratory vs between-lot variability

It is important to establish how much of the overall variability is caused by differences between laboratories vs differences between manufactured lots. Analysis of variance was used to estimate the separate contributions of these components. For ELISA measurements of A β 42, between-laboratory and between-kit lot components demonstrated approximately equal contributions, but for T-tau and P-tau the between-laboratory component was much larger than the between-lot component (Fig. 4). For xMAP measurements, both components contributed to the A β 42 variability, but for T-tau and P-tau the between-lot component was redundant, suggesting that its contribution was very small. Because of the unbalanced design and limited amount of data per assay lot and laboratory, variance components were estimated with large uncertainties. The results should therefore be interpreted as rankings of the different factors rather than exact calculations of their contributions.

3.5. Bias vs imprecision

We next examined bias and imprecision, which are descriptions of systematic and random deviations from a reference value, respectively. As expected [15], there was a large bias in analyte concentrations when evaluated against the different assay formats (see Fig. 1, but for measurements correlated between assay formats, see Supplemental Fig. 2). All subsequent statistical analyses were therefore performed by comparing laboratories using identical instrument platforms. Because there are no available standardized reference methods for CSF AD biomarker measurements, mean concentrations were used as reference values. For each measurement, the relative difference from the corresponding reference value (mean of measurements in all laboratories) was calculated. For each laboratory, the average of those relative differences was used to calculate the bias, whereas the variance of the differences was used to calculate the imprecision. For example, if a laboratory systematically reported higher than average concentrations, it had a positive bias; if a laboratory had a large variability in reported concentrations, it had a high imprecision. The bias and imprecision were plotted for each individual analyte and assay format (Fig. 5). This depiction revealed differences between the technologies. Low imprecision (defined as less than 10%)

was common among laboratories using ELISA measurements ($n = 50$ [72%] for at least one analyte, $n = 10$ [15%] for all analytes), but not among laboratories using xMAP measurements ($n = 7$ [37%] for at least one analyte, $n = 0$ [0%] for all analytes).

Last, we tested the laboratories against prehoc-defined criteria for high bias ($>30\%$ or 30%) and imprecision ($>20\%$). These limits were broken (called “events”) 46 times for all laboratories ($n = 82$) and analytes (the maximum number of possible events was 264: 63 [laboratories using only ELISA] $\times 3$ [analytes] + 13 [laboratories using only xMAP] $\times 3$ [analytes] + 6 [laboratories using both ELISA and xMAP] $\times 6$ [analytes]). We noted that a small number of laboratories was overrepresented among the events. In all, events were seen at 32 (39%) laboratories, but most of these ($n = 20$) had single events only. The majority of events were seen at a small minority of laboratories ($n = 12$, 15%), with two to three events each (Fig. 5). These laboratories may have a significant influence on the overall variability of a testing program. Excluding these 12 laboratories reduced the average between-laboratory CV by 3.3% to 21% for the different analytes (relative reductions; absolute reductions, 0.7%–3.9%; see Supplemental Table 1).

3.6. Influence of experience

With the possible exception of ELISA P-tau (Fig. 1C), there was no general trend for lower CV during the program, notwithstanding the gradual increasing experience obtained by the participants. However, the fact that the different laboratories have joined the program at different times might be a confounding factor in this respect. To test for this, we compared variability at round 9 among all laboratories vs variability among those who had participated in at least six rounds in the program, but there were no significant differences between these groups in CV (data not shown). Because we did not have access to detailed data about the amount of samples handled at each participating laboratory, we were not able to do a more specific analysis of variability in relation to laboratory experience.

3.7. Laboratory procedures as confounding factors

It is known that laboratory procedures differ among centers, even when using commercially available assays [11]. To monitor these differences and to estimate their importance for the overall variability in the QC program, all participants were asked to complete an extensive checklist of laboratory procedures (see the QC program homepage at <http://neurochem.gu.se/TheAlzAssQCProgram> for full contents). At round 9, 54 laboratories using ELISA had provided answers. On most questions, they answered very uniformly. Questions for which answers varied noticeably ($<85\%$ agreement) were considered potential confounding factors for measurements and included the use of automatic plate washing (21 yes, 32 no); the duration of QC sample thawing at room temperature (5–150 min); the use of internal QC samples (21 used pooled CSF samples, 26 used other samples, and seven did not use internal QC samples); the use of polypropylene plates for preincubation (38 yes, 8 no) and, if yes, the use of polypropylene plates for preincubation of both standards and CSF samples (28 yes, 6 no); and last, the use of a four-parameter logistic equation to calculate the standard curve (37 yes, 11 no). However, no significant influence of any of these parameters on measurement results was found, either for accuracy (determined by testing for differences in measurements by the Mann-Whitney U test or the Kruskal-Wallis test) or precision (determined by testing for differences in variances by Levene statistics). Too few laboratories provided checklist data for xMAP ($n = 12$) or MSD ($n = 5$) to evaluate the responses.

3.8. Testing universal cutoffs

A key issue is to what extent the variability in biomarker measurement influences interpretation consistency between centers and hinders the introduction of universal cutoffs

for putative AD. To test issue, we carried out a pilot experiment using previously reported biomarker cutoffs on ELISA and xMAP measurements of the blinded challenge samples in rounds 1 through 9 (18 samples). For ELISA, we used cutoffs from Buchhave and colleagues [16], in which the combination of the reduced A β 42-to-P-tau ratio (<6.16) and elevated T-tau (>350 ng/L) had a positive predictive value of 94%, a negative predictive value of 82%, sensitivity of 82%, and specificity of 94% for early-stage AD (patients with mild cognitive impairment developing AD dementia during a median of 9.2 years of follow-up). For xMAP, we used cutoffs from Shaw and colleagues [17], in which the elevated T-tau-to-A β 42 ratio (>0.39) had a positive predictive value of 86%, a negative predictive value of 85%, a sensitivity of 86%, and a specificity of 85% for (autopsy-confirmed) AD patients vs control subjects. Using these cutoff values, we classified the reported results from the QC laboratories, the results of which are presented in Table 2. We were surprised to find that despite the large variability described earlier, the consistency between laboratories was remarkably high. This was true especially for ELISA measurements, for which the common cutoff resulted in a more than 90% between-laboratory consistency in 15 of 18 samples. The consistency was lower for xMAP (>90% consistency in 12 of 18 samples). For most samples, there was a high consistency between ELISA and xMAP interpretations. Exceptions included samples with reduced A β 42 and elevated T-tau levels, but not elevated P-tau (which were more likely to be classified as AD by the xMAP algorithm than by the ELISA algorithm because the xMAP algorithm tested here did not use P-tau measurements).

4. Discussion

As the largest international network for CSF AD biomarker measurements, the Alzheimer's Association QC program is a valuable tool for identifying sources of global measurement variability. In this study of QC program data encompassing rounds 1 through 9 (corresponding to a time period of 3 years), the overall variability was generally around 20% to 30%, with lower numbers for ELISA than for xMAP and MSD measurements. This result is comparable with what has been observed in a previous QC program report (including only the first two rounds [15]), and with other QC monitoring initiatives [18–20]. A small part of the overall variability was caused by within-run variability (CV, <5%–10%, which is in agreement with previous reports and information from kit vendors regarding assay performance [21–23]). The within-laboratory longitudinal variability was larger (CV, 5%–19%), which is also in agreement with previous reports [8,24]. In analysis of variance, between-laboratory variability was a major contributor (19%–28%) to the overall variability. Despite using checklists for laboratory procedures, we could not identify any single factor causing the variability (but, all the information regarding laboratory procedures was supplied by the participants, and was not validated externally). However, we did note that a small number of labs (n = 12, 15%) were overrepresented among those with high bias (systematic deviations) and imprecision (random deviations). We propose that laboratories follow published guidelines more strictly [11] and product inserts to harmonize test performance. Studies examining the impact of uniform standardized operating procedures for CSF biomarker analyses and assays are ongoing in the Joint Program for Neurodegenerative Diseases project. For some of the analytes, especially A β 42 analyzed by ELISA or xMAP, there was a significant impact from between-lot dependent variability (22% and 10%, respectively). Therefore, it is critically important that kit manufacturers improve the quality of their products to minimize lot-to-lot variations (caused by matrix effects, variations in production of different kit components, and so on) to facilitate wider use of these assays in the clinical setting.

Commonly used clinical CSF biomarkers such as levels of albumin or immunoglobulin G may achieve an overall variability of less than 10% in external QC programs, and this is a reasonable ultimate goal for CSF AD biomarkers as well. Despite the high measurement

variability described here, a test using prehoc-derived cutoff levels for AD showed surprisingly high consistency between laboratories in sample classification. This result is encouraging for the development of validated universal biomarker cutoffs. Also, we would like to emphasize that our results should not delay the implementation of CSF biomarkers for evaluation of patients with AD symptoms in clinical practice at individual centers, because the clinical value of these biomarkers has been established in multiple, independent studies [25]. Rather, the variability stresses (i) the need for all laboratories to strive for longitudinal stability and to use validated internal cutoff levels, and (ii) the need for vendors to deliver more robust assays. More important, although this study was conducted on pooled CSF samples handled and delivered under strict controlled conditions, in the daily execution of CSF analyses in the different laboratories there are additional factors that may affect measurement results—in particular, the sampling, handling, and delivery of CSF to the laboratory—factors that are not under the direct responsibility of the performing laboratory. Also important is the need to collect the CSF into polypropylene tubes (especially critical for A β 42). To aid this situation, recommendations on preanalytical aspects of AD biomarker testing in CSF were recently published [12,26]. Relevant International Organization for Standardization norms were reviewed recently in this context [27].

In parallel with the QC work, several researchers are developing new methods for absolute quantification of CSF biomarkers (especially A β 42) that may serve as reference measurement procedures (assays available at this point in time must be considered as relative quantitative immunoassays) [13]. This movement toward CSF biomarker standardization also includes the creation of a certified reference material, which is being carried out as a collaborative effort between the Alzheimer's Association, the International Federation of Clinical Chemistry and Laboratory Medicine, and the Institute for Reference Materials & Measurements [28]. Another possibility may be the development of a certified proficiency panel to evaluate the (analytical) performance characteristics of CSF immunoassays to obtain an objective, measurable quality label from a regulatory agency. In combination, these efforts may increase the availability and usefulness of CSFAD biomarkers as tools for researchers and clinicians.

In conclusion, in the current study, we demonstrate that the most significant source of the observed variability for CSF biomarkers is between-laboratory factors. Each laboratory procedure potentially contributing to variation needs to be examined in a specifically designed experimental study with a sufficiently large number of samples. In the end, the transfer of assays to fully automated instruments, and the reduction of kit lot-to-lot variability, may eventually reduce both within- and between-laboratory variations. The QC program continues with multiple test rounds each year and is still open for enrollment. Inquiries regarding participation can be made to the coordinator at NeurochemistryLab@neuro.gu.se (see <http://neurochem.gu.se/TheAlzAssQCProgram> for more information). Several future program extensions are possible, such as evaluation of new assays or assay concepts, if there is enough evidence that patient care may be improved by using the new tools. A recently added feature of the QC program is the monitoring of how biomarker results are interpreted at the individual laboratories. Future analyses may also examine whether the variability between certified clinical laboratories is low enough to allow introduction of AD cutoffs that are shared between designated sites that fulfill QC requirements (the current analysis included several different types of laboratories, spanning from certified clinical laboratories to laboratories at pharmaceutical companies, as well as small and large research laboratories). Such analyses of certified clinical laboratories may clarify the potential clinical implications of the measurement variability described in this study.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Åsa Källén, Monica Christiansson, Sara Hullberg, and Dzemila Secic for excellent technical assistance. A generous grant from an anonymous donor to the Alzheimer's Association supported this study.

References

1. Vandermeeren M, Mercken M, Vanmechelen E, Six J, van de Voorde A, Martin JJ, et al. Detection of tau proteins in normal and Alzheimer's disease cerebrospinal fluid with a sensitive sandwich enzyme-linked immunosorbent assay. *J Neurochem.* 1993; 61:1828–34. [PubMed: 8228996]
2. Blennow K, Wallin A, Agren H, Spenger C, Siegfried J, Vanmechelen E. Tau protein in cerebrospinal fluid: a biochemical marker for axonal degeneration in Alzheimer disease? *Mol Chem Neuropathol.* 1995; 26:231–45. [PubMed: 8748926]
3. Motter R, Vigo-Pelfrey C, Kholodenko D, Barbour R, Johnson-Wood K, Galasko D, et al. Reduction of beta-amyloid peptide42 in the cerebrospinal fluid of patients with Alzheimer's disease. *Ann Neurol.* 1995; 38:643–8. [PubMed: 7574461]
4. Dubois B, Feldman HH, Jacova C, Cummings JL, Dekosky ST, Barberger-Gateau P, et al. Revising the definition of Alzheimer's disease: a new lexicon. *Lancet Neurol.* 2010; 9:1118–27. [PubMed: 20934914]
5. Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging–Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement.* 2011; 7:270–9. [PubMed: 21514249]
6. McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR Jr, Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging–Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement.* 2011; 7:263–9. [PubMed: 21514250]
7. Isaac M, Vamvakas S, Abadie E, Jonsson B, Gispen C, Pani L. Qualification opinion of novel methodologies in the predementia stage of Alzheimer's disease: cerebrospinal fluid-related biomarkers for drugs affecting amyloid burden: regulatory considerations by European Medicines Agency focusing in improving benefit/risk in regulatory trials. *Eur Neuropsychopharmacol.* 2011; 21:781–8. [PubMed: 21903360]
8. Bjerke M, Portelius E, Minthon L, Wallin A, Anckarsater H, Anckarsater R, et al. Confounding factors influencing amyloid beta concentration in cerebrospinal fluid. *Int J Alzheimers Dis.* 2010; 21:221–8.
9. Andreasson U, Vanmechelen E, Shaw LM, Zetterberg H, Vanderstichele H. Analytical aspects of molecular Alzheimer's disease biomarkers. *Biomark Med.* 2012; 6:377–89. [PubMed: 22917140]
10. Sancesario GM, Esposito Z, Nuccetelli M, Bernardini S, Sorge R, Martorana A, et al. Abeta1-42 detection in CSF of Alzheimer's disease is influenced by temperature: indication of reversible Abeta1-42 aggregation? *Exp Neurol.* 2010; 223:371–6. [PubMed: 19664624]
11. 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; 4061/2010/635053
12. Vanderstichele H, Bibl M, Engelborghs S, Le Bastard N, Lewczuk P, Molinuevo JL, 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. *Alzheimers Dement.* 2012; 8:65–73. [PubMed: 22047631]
13. Mattsson N, Zegers I, Andreasson U, Bjerke M, Blankenstein MA, Bowser R, et al. Reference measurement procedures for Alzheimer's disease cerebrospinal fluid biomarkers: definitions and approaches with focus on amyloid beta42. *Biomark Med.* 2012; 6:409–17. [PubMed: 22917143]

14. Mattsson N, Zetterberg H. What is a certified reference material? *Biomark Med.* 2012; 6:369–70. [PubMed: 22917138]
15. Mattsson N, Andreasson U, Persson S, Arai H, Batish SD, Bernardini S, et al. The Alzheimer's Association external quality control program for cerebrospinal fluid biomarkers. *Alzheimers Dement.* 2011; 7:386–95. [PubMed: 21784349]
16. Buchhave P, Minthon L, Zetterberg H, Wallin AK, Blennow K, Hansson O. Cerebrospinal fluid levels of beta-amyloid 1-42, but not of tau, are fully changed already 5 to 10 years before the onset of Alzheimer dementia. *Arch Gen Psychiatry.* 2012; 69:98–106. [PubMed: 22213792]
17. Shaw LM, Vanderstichele H, Knapik-Czajka M, Clark CM, Aisen PS, Petersen RC, et al. Cerebrospinal fluid biomarker signature in Alzheimer's Disease Neuroimaging Initiative subjects. *Ann Neurol.* 2009; 65:403–13. [PubMed: 19296504]
18. Lewczuk P, Beck G, Ganslandt O, Esselmann H, Deisenhammer F, Regeniter A, et al. International quality control survey of neurochemical dementia diagnostics. *Neurosci Lett.* 2006; 409:1–4. [PubMed: 17045397]
19. Verwey NA, van der Flier WM, Blennow K, Clark C, Sokolow S, De Deyn PP, et al. A worldwide multicentre comparison of assays for cerebrospinal fluid biomarkers in Alzheimer's disease. *Ann Clin Biochem.* 2009; 46:235–40. [PubMed: 19342441]
20. Shaw LM, Vanderstichele H, Knapik-Czajka M, Figurski M, Coart E, Blennow K, et al. Qualification of the analytical and clinical performance of CSF biomarker analyses in ADNI. *Acta Neuropathol.* 2011; 121:597–609. [PubMed: 21311900]
21. Mattsson N, Portelius E, Rolstad S, Gustavsson M, Andreasson U, Stridsberg M, et al. Longitudinal cerebrospinal fluid biomarkers over four years in mild cognitive impairment. *J Alzheimers Dis.* 2012; 30:767–78. [PubMed: 22475796]
22. Olsson A, Vanderstichele H, Andreasen N, De Meyer G, Wallin A, Holmberg B, et al. Simultaneous measurement of beta-amyloid(1-42), total tau, and phosphorylated tau (Thr181) in cerebrospinal fluid by the xMAP technology. *Clin Chem.* 2005; 51:336–45. [PubMed: 15563479]
23. Schoonenboom NS, Mulder C, Vanderstichele H, Van Elk EJ, Kok A, Van Kamp GJ, et al. Effects of processing and storage conditions on amyloid beta (1-42) and tau concentrations in cerebrospinal fluid: implications for use in clinical practice. *Clin Chem.* 2005; 51:189–95. [PubMed: 15539465]
24. Zimmermann R, Lelental N, Ganslandt O, Maler JM, Kornhuber J, Lewczuk P. Preanalytical sample handling and sample stability testing for the neurochemical dementia diagnostics. *J Alzheimers Dis.* 2011; 25:739–45. [PubMed: 21593568]
25. Bloudek LM, Spackman DE, Blankenburg M, Sullivan SD. Review and meta-analysis of biomarkers and diagnostic imaging in Alzheimer's disease. *J Alzheimers Dis.* 2011; 26:627–45. [PubMed: 21694448]
26. del Campo M, Mollenhauer B, Bertolotto A, Engelborghs S, Hampel H, Simonsen AH, et al. Recommendations to standardize pre-analytical confounding factors in Alzheimer's and Parkinson's disease cerebrospinal fluid biomarkers: an update. *Biomark Med.* 2012; 6:419–30. [PubMed: 22917144]
27. Waedt J, Kleinow M, Kornhuber J, Lewczuk P. Neurochemical dementia diagnostics for Alzheimer's disease and other dementias: an ISO 15189 perspective. *Biomark Med.* 2012; 6:685–90. [PubMed: 23075248]
28. Mattsson N, Andreasson U, Carrillo MC, Persson S, Shaw LM, Zegers I, et al. Proficiency testing programs for Alzheimer's disease cerebrospinal fluid biomarkers. *Biomark Med.* 2012; 6:401–7. [PubMed: 22917142]

RESEARCH IN CONTEXT

1. **Systematic review:** Cerebrospinal fluid (CSF) biomarkers are used increasingly in Alzheimer's disease (AD) research, but there is an active discussion about measurement variability. We searched PubMed for published studies on biomarker variability and also reviewed specific studies of which we were previously aware.
2. **Interpretation:** To our knowledge, this is the most extensive study ever published on CSF AD biomarker variability. We provide new data on different sources of variability, including between- and within-laboratory, and between-kit lot variability for several different assay systems. We show that a minority of laboratories are overrepresented among those with high bias and/or imprecision, and that excluding them improves the overall results. Last, we demonstrate that, despite the variability, the consistency among laboratories in sample classification is relatively high.
3. **Future directions:** Critical steps are development of certified reference methods and standardization of laboratory procedures. Ultimately, this may lead to development of universal biomarker cutoffs.

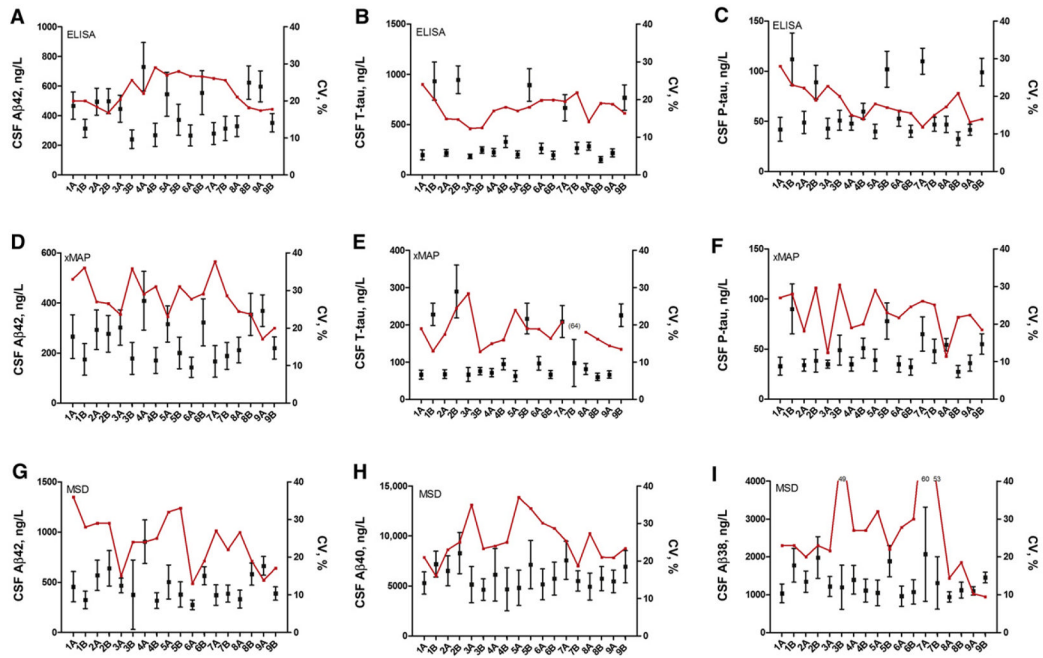
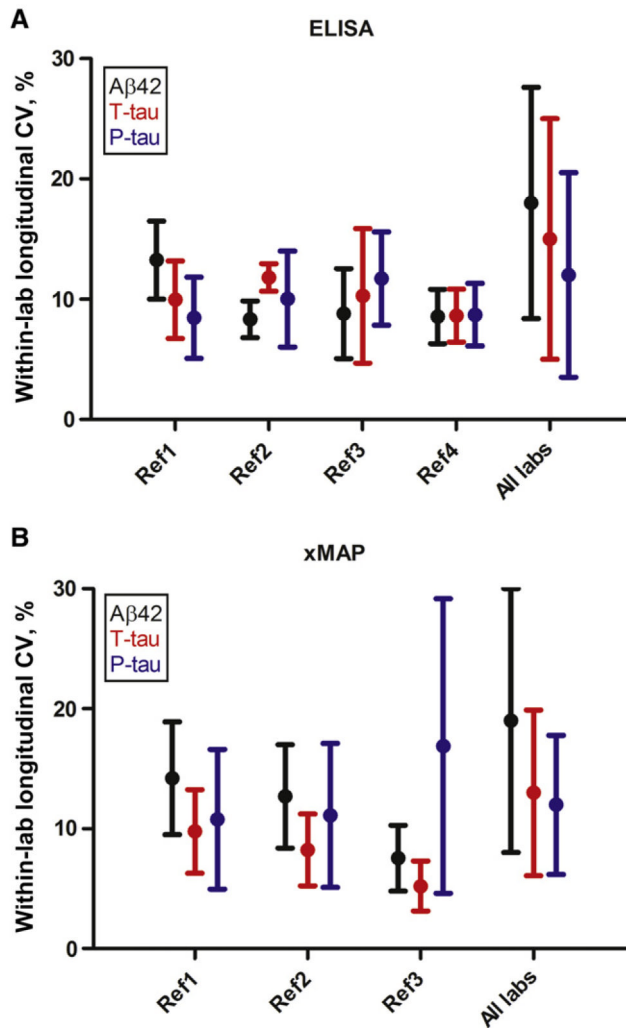


Fig. 1. Measurements of blinded quality control test samples. Dots with error bars show mean measured concentrations and standard deviation from all participating sites (the samples were made from different pools of cerebrospinal fluid [CSF], so constant concentrations were not expected). Connected lines show the coefficient of variation (CV; right-hand y-axes). (A–C) INNOTEST Enzyme-linked immunosorbent assay (ELISA). (D–F) INNO-BIA xMAP. (G–I) Meso Scale Discovery (MSD) amyloid beta (A β) triplex. The CV for xMAP total tau (T-tau) sample 7B (E) was very high (64%) because of a single extreme outlier (the CV was 22% after removal of this outlier). P-tau, phosphorylated tau.

**Fig. 2.**

(A, B) Within-laboratory longitudinal coefficients of variation (CVs) were calculated by repeated measurements at reference laboratories (Ref x ; using six measurements per sample, varying number of samples per laboratory) and in the whole program (using the quality control longitudinal sample at laboratories measuring the sample at least three times during rounds 1 through 7). Data are means of CV (error bars are standard deviations) for each biomarker, ordered by laboratory (x -axes). Enzyme-linked immunosorbent assay (ELISA) reference laboratory 4 (Ref 4) only used two lots of analytical kits for each analyte, which limits the influence of lot-dependent variability. Meso Scale Discovery (MSD) data are not included in the figure, since only one reference laboratory reported data for MSD within-laboratory longitudinal CV (mean CV at that laboratory was 11% to 17% (standard deviation, 4%–8%) for all amyloid beta (A β) triplex measurements using either 6E10 or 4G8 as the detection antibody. T-tau, total tau; P-tau, phosphorylated tau.

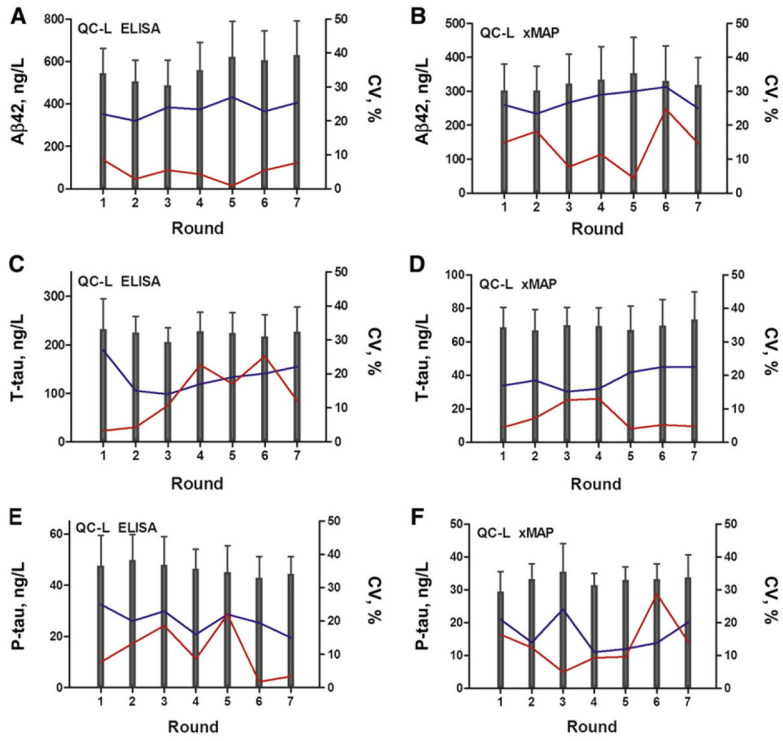


Fig. 3. Measurements of longitudinal quality control samples. Bars show mean measured concentrations (error bars are standard deviations) for the quality control longitudinal sample (constant concentrations expected). Connected lines show the coefficient of variation CV (right-hand y-axes) for all laboratories (blue) and reference laboratories (red). ELISA, enzyme-linked immunosorbent assay; Aβ, amyloid beta; T-tau, total tau; P-tau, phosphorylated tau.

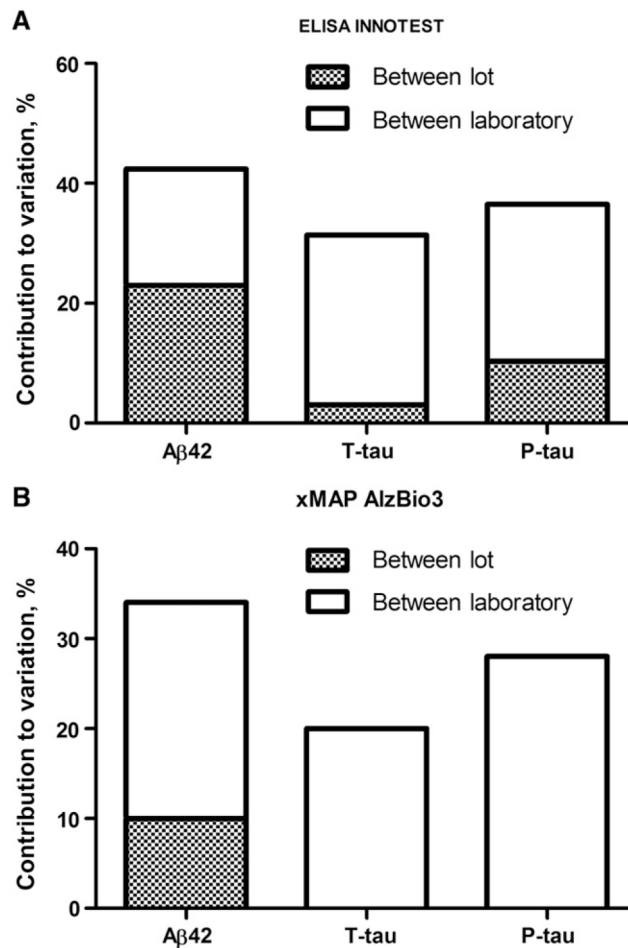


Fig. 4. Between-kit lot and between-laboratory contributions to variability. Bars show the contribution from the different components to the overall variability according to variance component analysis for enzyme-linked immunosorbent assay (ELISA) (A) and xMAP (B). The bars are stacked. Note that the components do not sum to 100% because there is also an influence of within-laboratory and within-kit lot variability that contributes to the total. A β , amyloid beta; T-tau, total tau; P-tau, phosphorylated tau.

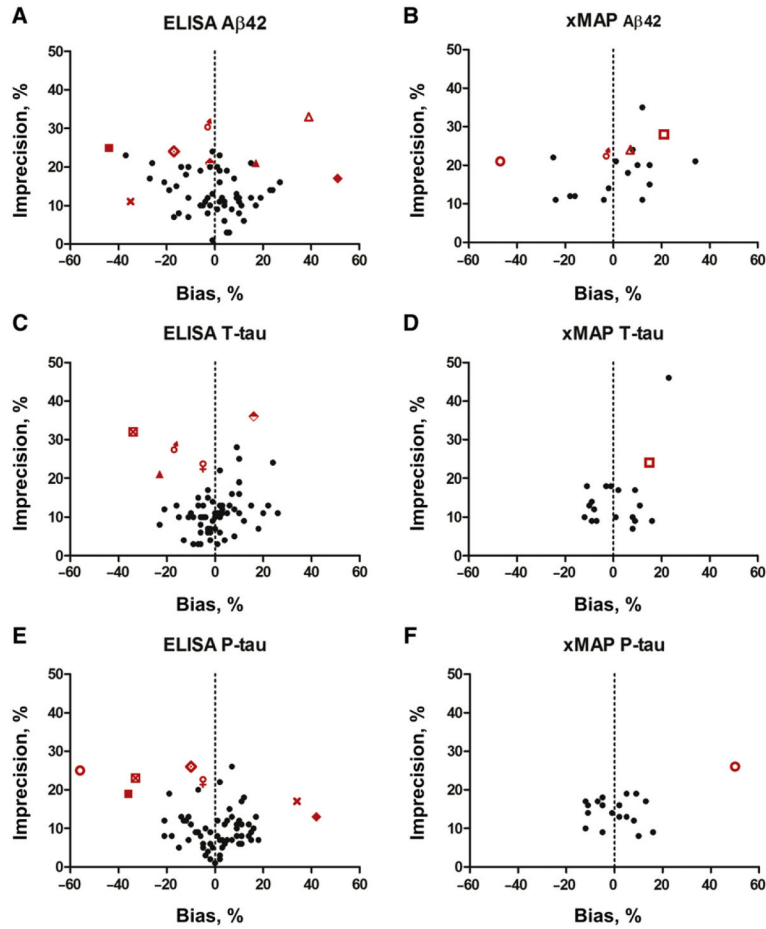


Fig. 5. Bias and imprecision plots. Bias (systematic variation) at a particular laboratory was estimated by comparing the measurement at that laboratory with the mean of all reported measurements for that particular sample. Imprecision (random variation) was estimated by calculating the variance of the results. Laboratories with high bias (defined as > 30% or 30% from the mean) or high imprecision (defined as > 20%) for at least two analytes are indicated by special symbols. Three of these laboratories (■, ◆, and ×) only provided data for one to two rounds, which make their estimates uncertain. Laboratory ○ provided data for three rounds, but on review it was found that they were all analyzed at the same time point, which may contribute to a strong bias. The remaining laboratories, indicated by special symbols, provided data for five to nine rounds. ELISA, enzyme-linked immunosorbent assay; Aβ, amyloid beta; T-tau, total tau; P-tau, phosphorylated tau.

Table 1

Within-run variability in cerebrospinal fluid measurements among laboratories

Round	ELISA		xMAP		MSD 6E10		MSD 4G8			
	A β 42	T-tau	P-tau	A β 42	T-tau	P-tau	A β 42	A β 38		
4	3.2 (1.2, 6.29)	3.0 (2.0, 5.2, 32)	2.1 (1.0, 3.7, 31)	2.5 (1.2, 4.8, 12)	3.6 (2.7, 6.1, 12)	3.9 (2.3, 4.1, 5)	5* (5.1, 6.8, 5)	4.2 (3.8, 4.4, 5)	5* (4.7, 9.1, 3)	17* (5.1, 13.3)
5	2.5 (1.7, 4.7, 43)	3.0 (1.4, 6.9, 46)	1.6 (0.5, 2.9, 44)	4.1 (2.6, 6.1, 14)	3.2 (2.3, 5.4, 14)	1.5 (1.0, 5.4, 5)	6.7 (3.8, 13, 5)	5.9 (5.4, 9, 5)	4.6 (2.7, 11, 3)	6.3 (4.7, 9.1, 3)
6	3.7 (1.7, 6.6, 47)	3.4 (1.9, 7.3, 50)	1.6 (0.9, 4.8, 48)	4.9 (1.4, 6.2, 15)	5.7 (2.9, 8.9, 15)	2.7 (2.0, 3.0, 4)	6.1 (4.5, 7.7, 4)	3 (1.2, 4.7, 4)	5.9 (5.4, 6.5, 2)	5.5 (3.7, 7.2, 2)
7	2.7 (1.7, 4.2, 52)	2.9 (0.9, 5.8, 52)	2.0 (0.7, 4.2, 53)	7.3 (2.9, 12, 14)	3.7 (1.7, 6.6, 13)	3.1 (2.0, 4.2, 4)	2.1 (0.7, 3.7, 4)	3.0 (1.8, 3.8, 4)	4.4 (2.8, 6.2, 3)	4.2 (3.1, 10, 3)

Abbreviations: ELISA, enzyme-linked immunosorbent assay; MSD, Meso Scale Discovery; A β , amyloid beta; T-tau, total tau; P-tau, phosphorylated tau.

NOTE. Within-run variability was calculated using duplicate measurements (two wells) of the quality control longitudinal sample. Data are expressed as median of the coefficient of variation (25th percentile, 75th percentile, n). The results were similar for the other samples (data not shown).

* Data available from one laboratory only.

Table 2

Consistency between laboratories in test sample AD profiles

Round	Sample	ELISA		xMAP		AD profile, %
		n	AD profile, %	n	AD profile, %	
1	A	25	0.0	14	21.4	
1	B	24	91.7	14	100.0	
2	A	25	0.0	16	12.5	
2	B	25	68.0	16	100.0	
3	A	25	0.0	15	13.3	
3	B	24	0.0	15	66.7	
4	A	35	0.0	14	0.0	
4	B	34	35.3	14	92.9	
5	A	43	0.0	16	0.0	
5	B	43	95.3	16	100.0	
6	A	48	2.1	16	100.0	
6	B	48	0.0	16	0.0	
7	A	54	100.0	15	100.0	
7	B	54	0.0	15	80.0	
8	A	53	0.0	11	54.5	
8	B	52	0.0	11	0.0	
9	A	54	0.0	12	0.0	
9	B	54	100.0	12	100.0	

Abbreviations: AD, Alzheimer's disease; ELISA, enzyme-linked immunosorbent assay.

NOTE. Percentages of laboratories' reporting values consistent with an AD biochemical profile for each sample. Percentages close to or reaching 100% or 0% indicate high consistency and are desirable. The AD profile was defined as an amyloid beta 42-to-phosphorylated tau ratio less than 6.16 combined with a total tau of more than 350 ng/L for ELISA measurements [16], and a total tau-to-amyloid beta 42 ratio of more than 0.39 for xMAP measurements [17].