Hypocretin-1 measurements in cerebrospinal fluid using radioimmunoassay: within and between assay reliability and limit of quantification

Hoeven, A.E. van der; Waaij, K. van; Bijlenga, D.; Roelandse, F.W.C.; Overeem, S.; Bakker, J.A.; ... ; Lammers, G.J.

SLEEP, 2022, 1–8

https://doi.org/10.1093/sleep/zsac103
Advance Access Publication Date: 5 May 2022
Original Article

ORIGINAL ARTICLE

Hypocretin-1 measurements in cerebrospinal fluid using radioimmunoassay: within and between assay reliability and limit of quantification

Adrienne Elisabeth van der Hoeven[1,2], Kevin van Waaij[1], Denise Bijlenga[1,2], Frederik Willem Cornelis Roelandse[3], Sebastiaan Overeem[4], Jaap Adriaan Bakker[3], Rolf Fronczek[1,2] and Gert Jan Lammers[1,2,*]

[1]Department of Neurology, Leiden University Medical Center, Leiden, the Netherlands, [2]Sleep-Wake Center, Stichting Epilepsie Instellingen Nederland (SEIN), Heemstede, the Netherlands, [3]Department of Clinical Chemistry and Laboratory Medicine, Leiden University Medical Center, Leiden, the Netherlands and [4]Sleep Medicine Center, Kempenhaeghe, Heeze, the Netherlands

*Corresponding author. Gert Jan Lammers, Department of Neurology, Leiden University Medical Center, Leiden, the Netherlands. Email: g.j.lammers@lumc.nl.

Abstract

Study Objectives: The most sensitive and specific investigative method for the diagnosis of narcolepsy type 1 (NT1) is the determination of hypocretin-1 (orexin-A) deficiency (≤110 pg/mL) in cerebrospinal fluid using a radioimmunoassay (RIA). We aimed to assess the reliability of the Phoenix Pharmaceuticals hypocretin-1 RIA, by determining the lower limit of quantification (LLOQ), the variability around the cutoff of 110 pg/mL, and the inter- and intra-assay variability.

Methods: Raw data of 80 consecutive hypocretin-1 RIAs were used to estimate the intra- and inter-assay coefficient of variation (CV). The LLOQ was established and defined as the lowest converted concentration with a CV <25%; the conversion is performed using a harmonization sample which is internationally used to minimize variation between RIAs.

Results: The mean intra-assay CV was 4.7%, while the unconverted inter-assay CV was 28.3% (18.5% excluding 2 outliers) and 7.5% when converted to international values. The LLOQ was determined as 27.9 pg/mL. The intra-assay CV of RIAs with lower specific radioactive activity showed a median of 5.6% (n = 41, range 1.6%–17.0%), which was significantly higher than in RIAs with higher specific activity (n = 36; median 3.2%, range 0.4%–11.6%, p = .013). The CV around the 110 pg/mL cutoff was <7%.

Conclusions: Hypocretin-1 RIAs should always be harmonized using standard reference material. The specific activity of an RIA has a significant impact on its reliability, because of the decay of ^{125}I radioactivity. Values around the hypocretin-1 cut-off can reliably be measured. Hypocretin-1 concentrations below 28 pg/mL should be reported as "undetectable" when measured with the Phoenix Pharmaceuticals RIA.

Clinical Trial Information: This study is not registered in a clinical trial register, as it has a retrospective database design

Statement of significance

The reliability of radioimmunoassay to determine cerebrospinal fluid hypocretin-1 levels has not systematically been evaluated. We found high intra-assay reliability and low inter-assay reliability when concentrations were not converted to international standard values. Harmonization using a commonly used reference sample from Stanford greatly improved inter-assay reliability. Additionally, there is a clear lower limit of quantification. In spite of this, numerous previous studies have reported very low hypocretin-1 values, which were used in various analyses. These reported values lie far below the lower limit of quantification and should be presented as such. The outcomes of this evaluation of 20 years of experience have implications for both previously published and future research.

Key words: orexin-A; narcolepsy; intra-assay coefficient of variation; inter-assay coefficient of variation

Submitted: 15 October, 2021; **Revised:** 30 March, 2022

Introduction

Narcolepsy type 1 (NT1) is a rare chronic neurological sleep-wake disorder, characterized by excessive daytime sleepiness, sleep fragmentation, sleep paralysis, hypnagogic hallucinations, and cataplexy [1]. In addition, a decreased or undetectable concentration of hypocretin-1 (also known as orexin-A) in cerebrospinal fluid (CSF) can be used to confirm a NT1 diagnosis. This hypocretin-1 deficiency is hypothesized to be caused by the autoimmune destruction of hypocretin-producing neurons [2]. Hypocretin-1 deficiency is found in a few other disorders [3, 4] and is considered the gold standard for diagnosing NT1 [5].

The cutoff value for hypocretin deficiency is 110 pg/mL (adjusted for Stanford values) [6, 7], and is commonly determined using the radioimmunoassay (RIA) kit from Phoenix Pharmaceuticals Inc. (Burlingame, CA, USA). To determine the hypocretin-1 concentration, the RIA uses a known quantity of radioactive (^{125}I) labeled hypocretin-1, competing for a known quantity of hypocretin-1 antibody binding sites with the unlabeled hypocretin-1 in the sample. Then, the unbound antigens are washed away, and the radioactivity of the precipitate is measured [8]. More unlabeled antigens result in less specific activity (i.e. the activity per mass of radionuclide [9], expressed as counts per minute [CPM]). Thus, the hypocretin concentration is determined using a calibration curve based on standards with known concentrations.

However, RIAs have limitations. As radioactive materials are involved, precautions have to be taken concerning the use and disposal of the materials. Furthermore, antibody batches can vary greatly in their binding potential [8]. Other factors to consider include differences in experience using RIAs to measure hypocretin-1 between laboratories, and possible cross-reactivity with matrix constituents [10, 11]. The fast decay of ^{125}I radioactivity is another limitation, leading to short expiration dates [12]. Differences in shelf life thus cause variability in specific radioactivity between RIAs, potentially impacting their reliability. The extent of the variability in outcomes caused by these factors has not been systematically evaluated.

Test duplicate samples are used to estimate for each pair the coefficient of variation (CV, i.e. the ratio of the standard deviation to the mean value of the measured concentrations). The CV gives an indication of the variability and can be used to determine the intra-assay variability of the RIA [13]. The inter-assay variation (i.e. inter-assay CV) can be estimated by analyzing aliquots of the same sample in different assay runs [14]. Harmonization samples from Stanford are used to verify the reliability of an RIA and to correct the measured values [15]. These reference samples have a known concentration of hypocretin-1 and can be used for many RIAs due to the fact that hypocretin-1 concentrations in CSF remain very stable over time even after freezing and thawing [16]. With the use of these reference samples, a conversion factor can be estimated for the harmonization of the individual RIAs [15].

The reliability of measurements around the cutoff point of 110 pg/mL is of clinical relevance but has not been determined yet. Also, even though most studies report CSF hypocretin-1 concentrations in absolute numbers, the lowest hypocretin-1 concentration that can be reliably determined (i.e. lower limit of quantification [LLOQ]) is yet unknown [17, 18]. It is internationally agreed that the LLOQ should be within 25% of the nominal value, while the measurement of quality control (QC) samples (or standards used to create a calibration curve) need to have a repeatability of 20% or less [19, 20].

To further validate the Phoenix Pharmaceuticals hypocretin-1 RIA kit, we established the inter-assay variability and the intra-assay variability in general and around the cutoff of 110 pg/mL, and the LLOQ. We also assessed the impact of differences in specific radioactivity between different antibody batches on the reliability of the measured hypocretin-1 levels.

Methods

Data collection

The Leiden University Medical Center (LUMC) Department of Neurology has an international reference center and provides a service for hypocretin-1 measurement. The LUMC Department of Clinical Chemistry has more than 20 years of experience in measuring hypocretin-1 in CSF, using the Phoenix Pharmaceuticals RIA kit for the quantification of hypocretin-1. Between October 2001 and December 2021, 80 RIAs were performed to determine the CSF hypocretin-1 concentrations of people with suspected NT1. We analyzed raw data from the 80 consecutive RIAs.

During validation or verification of new analytical methods determination, the intra- and inter-assay CV is required. The Guideline on bioanalytical method validation of the European Medicines Agency (EMA), like others [20], recommends measuring multiple QC samples at five concentration levels covering the calibration curve range: around the anticipated LLOQ, a low, medium, and high concentration and around the anticipated ULOQ (upper limit of quantification). To determine the inter-assay CV, these QC samples need to be measured on different assay runs. We could not apply this preferred procedure due to the retrospective nature of our study. We therefore mimicked this using available data. To enhance feasibility and to reduce selection bias we choose to limit the collection of raw data (duplicate CPM measurements) to:

- The RIA calibration curve standards, which have the following standard hypocretin-1 concentrations: 10, 20, 40, 80, 160, 320, 640, and 1280 pg/mL.
- QC samples, which include:
  - Stanford reference sample measurements per RIA (*harmonization samples*, mostly measured in quadruplicate or, in a few cases, in duplicate). These reference samples have a known concentration of 280 pg/mL (16 RIAs) or 329 pg/mL (64 RIAs) and are used to correct for variation between RIAs. Two samples with different concentrations were used because after performing RIA 16, the 280 pg/mL reference sample was used up, and a new sample with a different concentration (329 pg/mL) was taken into use.
  - In line with updated quality requirements, *internal control sample* measurements per RIA (measured in duplicate) were introduced to monitor long-time stability of the assay. These QC samples were created by the Department of Clinical Chemistry of the LUMC by pooling individual samples, resulting in an internal control sample with an intermediate hypocretin-1 concentration, which is used to internally verify the reliability of RIAs over a longer period. Internal control sample measurements of 49 RIAs were available.

  - *Kit control sample* measurements per RIA (measured in duplicate); an extra QC sample was provided by the supplier of the kit. It should be noted that unlike the Stanford harmonization and internal control samples, the kit control samples differed for each RIA. As such, they cannot be used to determine the inter-assay CV. Kit control duplicate measurements of 57 RIAs were available.

- Individuals samples with a hypocretin-1 concentration below 200 pg/mL. These were added to determine the LLOQ and the CV in the lower range.

Outliers due to analytical errors were excluded. If the CV of a CPM duplicate was ≥20%, the sample was not collected, since this is likely due to an error during the RIA's execution.

Sample collection, storage, and preparation

The CSF was obtained by lumbar puncture during regular clinical practice of local centers. Centers centrifuged the samples at 4000 rpm for 5 min, as part of the general practice to remove cell debris, before transport and sent the supernatant liquid. Internal control CSF was stored at –70°C, while regular patient samples were stored at –20°C. CSF hypocretin-1 levels were determined using RIA every 3 months. The total process of preparation and measurement takes 3 days. On the third day, the RIA buffer is added to the samples (in borosilicate glass tubes) and the samples are centrifuged at 3600 rpm for 25 min. Then, the specific radioactive activity is measured. The RIAs were performed under the conditions recommended by the supplier of the kit.

Analysis

Intra-assay CV. The intra-assay CV was determined for each RIA and the average intra-assay CV for all RIAs. The measurements of all three QC samples (Stanford harmonization, internal control, and kit control) were used for this calculation when available (the internal and kit control samples were not measured for every RIA and two RIAs lacked QC sample measurements). The following method was used to estimate the intra-assay CV (see Figure 1 for an overview of this method). The calibration curve for each RIA was plotted and used to convert the CPM measurements of the QC samples to concentrations. Using these concentrations, 3 sub-intra-assay CVs could be estimated per RIA with the following formula: (SD quadruplicates or duplicates)/(mean quadruplicates or duplicates) × 100%. The sub-intra-assay CVs were then used to estimate the total intra-assay CV for each RIA with another formula: $\sqrt{[\Sigma CV_i^2/N]}$ [21]. If only one sub-CV was available, that one was used as the total intra-assay CV. Subsequently, the average of all the total intra-assays was determined. Seventy-seven of 80 RIAs could be used in this calculation: the Stanford harmonization, internal control, and kit sample measurements of two RIAs were missing, while the results for the calibration curve were missing for a third RIA.

Inter-assay CV. The Stanford harmonization and internal control samples were used to determine the inter-assay CV. As with the intra-assay CV, the converted hypocretin-1 concentrations were used rather than the direct quadruplicate or duplicate CPM measurements. The variability between RIAs was estimated twice, using Stanford converted concentrations, and using unconverted concentrations. This way, we assessed the necessity and impact of the current process of harmonizing RIA outcomes using reference samples from Stanford.

Stanford harmonization measurements (those with the same known concentration of 329 pg/mL) were only used in the calculation of the unconverted inter-assay CV. The internal control measurements were used for the unconverted and Stanford converted inter-assay CV, giving a total of three inter-assay CVs:

1. The inter-assay CV using the new Stanford harmonization measurements not converted to Stanford values was determined as follows: (SD mean Stanford harmonization sample concentrations)/(average mean Stanford harmonization sample concentration) × 100%. 58 RIAs were available for this calculation. The RIAs that could not be used consisted of four RIAs with Stanford harmonization outliers, one RIA with missing Stanford harmonization measurements, and one RIA lacking the data for the calibration curve.
2. The inter-assay CV estimated using the internal control samples, converted and unconverted for Stanford, was obtained using the following formula: (SD mean Stanford (un)converted internal control sample concentration/average mean Stanford (un)converted internal control sample concentration) × 100%. 40 out of 74 RIAs were available for this calculation. Excluded were: two RIAs with internal control sample outliers, two RIAs for which it was not possible to use internal control sample concentrations due to faulty measurements, one RIA missing a calibration curve, and 29 RIAs lacking internal control sample measurements.
3. To correct the internal control duplicate, the conversion factor was estimated for each RIA by dividing 329 pg/mL by the mean of the measured Stanford harmonization concentrations derived from the calibration curves. Afterwards, the conversion factor was applied to the measured internal control concentrations, resulting in Stanford converted internal control concentrations.

Determining the lower limit of quantification. The LLOQ was set as the concentration with a CV of ≤25%, following international guidelines [19, 20].

To determine the LLOQ, we used the raw data of all samples with an average concentration below 200 pg/mL (unconverted as well as converted to Stanford values). Duplicate measurements were excluded when the CV of CPM measurements was more than 20%, as this suggests an analytical error as opposed to normal variation. Using these data, the intra-assay CVs of the concentration measurements were estimated. By plotting

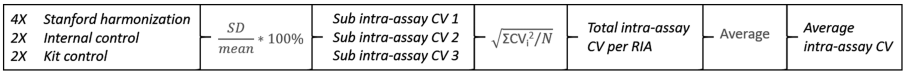

Figure 1. Overview of the method used to determine the intra-assay CV for each RIA and the average intra-assay CV for all RIAs (n = 77).

samples with the CV per sample on the y-axis and average concentration per sample on the x-axis, logarithmic modeling was applied and the LLOQ was interpolated.

To estimate the CV at a 110 pg/mL concentration, we used the equations based on logarithmic modeling using the above-mentioned duplicate individual sample measurements.

Determining the impact of 125I radioactivity decay on RIA reliability. To assess the impact of shelf life and amount of specific activity on RIA reliability, we performed an additional analysis in which we divided the RIAs into two separate groups: a group with <3000 CPM and a group with ≥3000 CPM. Due to radioactive decay, we assumed that the RIAs with <3000 CPM were nearer the expiration date. Due to non-normal distribution, we used the Mann-Whitney *U* test to examine whether the intra-assay CVs differed significantly between groups. We also assessed whether radioactive decay affects the LLOQ, we measured the LLOQ in the group with lower and in the group with higher specific activity.

Data were analyzed using Microsoft Excel 2016 and SPSS version 25.0. For all statistical analyses, a statistical significance level of $\alpha = .05$ (two-tailed) was used. Figures were created using the GraphPad Prism 8 software.

Results

Calibration curves

Seventy-nine sigmoidal calibration curves were obtained from which concentrations were estimated (see Figure 2 for two examples). Calibration curve standards were missing for one RIA, excluding this RIA from the analyses. The CPM range of the calibration curves is plotted in Figure 3.

Variability within RIAs: the intra-assay CV

The mean intra-assay CV was 4.73% (SD ± 2.82%) when using all available RIAs. The mean intra-assay CV was 5.04% (SD ± 2.85%)

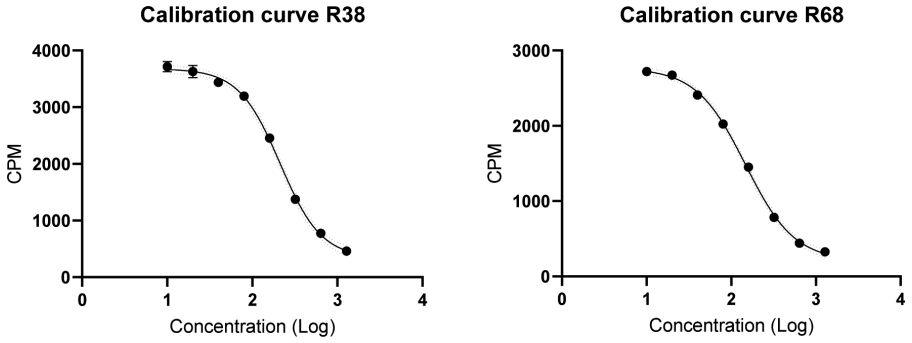

Figure 2. Two examples of RIA calibration curves. RIA R38 left (A), R68 right (B) on the x-axis is the concentration transformed to a log scale and on the y-axis the CPM.

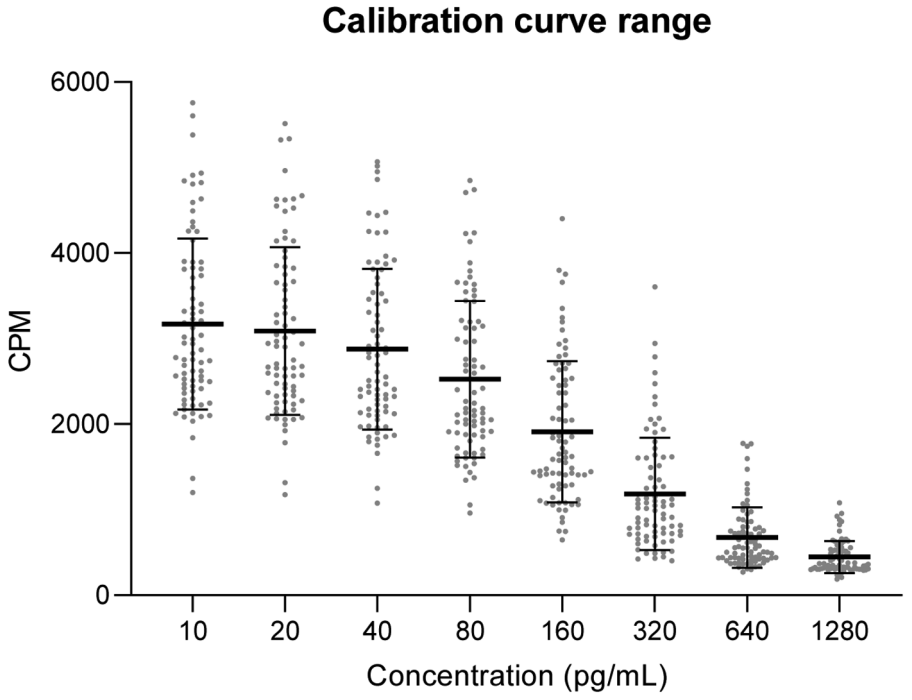

Figure 3. RIA calibration curve range comparison, the bars represent the means and SDs of each standard concentration. Shown here are all standards from 79 RIAs plotted against their CPMs, visualizing the spread of the CPM range across the RIAs. As the concentration increases, the CPM spread decreases.

when excluding the 29 RIAs that missed one or two (of three) sub-intra-assay CVs (those estimated with either the Stanford harmonization sample alone or with the kit or internal control sample).

Variability between RIAs: the inter-assay CV

The inter-assay CV was computed three times: twice using the internal controls (unconverted and as converted for Stanford) and once using the new Stanford harmonization samples (those with the known concentration of 329 pg/mL).

Using the internal controls, the inter-assay CV was 28.28% (n = 48) when not converted using the Stanford harmonization sample (18.49% without the two outliers, n = 46) and 7.46% when converted (7.49% without the two outliers). The average of the applied conversion factors was 1.23, causing the interpolated semilog line to shift to the right (as can be seen in Figure 4).

The inter-assay CV estimated using the new Stanford harmonization samples (n = 62) was 33.17% when outliers were included, and 23.14% when four outliers were excluded (n = 58).

Lower limit of quantification

The LLOQ graphs were based on 809 unconverted samples and 635 Stanford converted samples. The resulting graphs gave the following LLOQs (with a CV of <25%): 22.9 pg/mL for the unconverted <200 pg/mL samples and 27.9 pg/mL for the Stanford converted <200 pg/mL samples, see Figure 4.

The concentration with a CV of 20% was 34.5 pg/mL when unconverted and 40.7 pg/mL when converted.

Reliability around 110 pg/mL

The CV with a unconverted concentration of 110 pg/mL was 5.89% and 6.79% with a converted concentration.

Impact of ^{125}I radioactivity decay on RIA reliability

The intra-assay CV of RIAs with lower specific activity (CPM range <3000, n = 41) was significantly higher than the intra-assay CV of RIAs with higher specific activity (CPM range ≥3000, n = 36), with a median of 5.58% (min 1.6%, max 17.0%) for the lower activity and 3.24% (min 0.4%, max 11.6%, p = .013) for the higher activity.

When assessing whether radioactive decay impacts the LLOQ determined using converted hypocretin-1 concentration measurements, a lower LLOQ was found in the group with lower specific activity (22.6 pg/mL, $y = -12.08 \times \ln(x) + 62.683$) than in the group with higher specific activity (35.1 pg/mL, $y = -14.22 \times \ln(x) + 75.603$). When assessing the impact on unconverted concentration measurements, however, the LLOQ was almost the same (20.2 pg/mL, $y = -11.39 \times \ln(x) + 59.218$ and 24.5 pg/mL, $y = -12.55 \times \ln(x) + 65.149$.

Discussion

We evaluated technical aspects of CSF hypocretin-1 measurements using the RIA kit from Phoenix Pharmaceuticals Inc. which is currently used worldwide. We found (1) a high intra-assay reliability, including around the cutoff point of 110 pg/mL; (2) a low inter-assay reliability without conversion to Stanford values, which improved substantially after conversion; (3) an LLOQ (with a CV ≤25%) obtained from the converted samples of 27.9 pg/mL; and (4) that the CPM range differences between RIAs varies widely. These results have important implications for the clinical use of RIAs for hypocretin-1 measurements in CSF for the diagnosis of NT1.

Intra- and inter-assay CV

Specifications of inter-individual biologic variation and what is considered an acceptable level of imprecision for laboratory testing are listed for many analytes [22, 23], but this is not the case for hypocretin-1. While guidelines generally recommend an acceptance criterion of a ≤15% deviation within and between assays, when it comes to ligand-binding assays such as the RIA, a higher deviation of ≤20% is accepted [19, 20, 24, 25].

The intra-assay CV we estimated using our data is acceptable, with even the intra-assay variability around the cutoff point of 110 pg/mL being far below the accepted deviation of 20%, and similar to intra-assay variations found in earlier studies [3, 26–28]. Specific methods on how these variations were estimated in

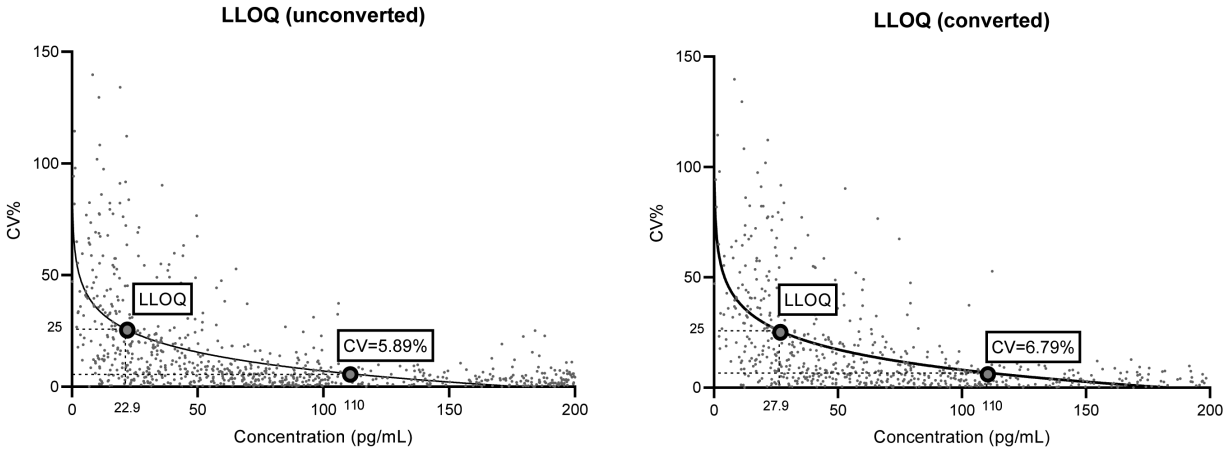

Figure 4. Lower limit of quantification (LLOQ). Shown here are the graphs from which the LLOQs were determined (at a concentration CV of 25%). For the 809 unconverted <200 pg/mL samples (A), the LLOQ is 34.5 pg/mL (formula: $y = -12.17 \times \ln(x) + 63.091$). For the 635 converted samples (B), the LLOQ is 40.7 pg/mL (formula: $y = -13.29 \times \ln(x) + 69.255$).

these other studies were not provided. To illustrate the importance of the method used, Keating et al. [12] found an average intra-assay hypocretin-1 variation of 9.4% (higher than the present study and other studies). In order to obtain an intra-assay variation, the authors divided the absolute value of the difference in a duplicate measurement by the higher of the two values, which tends to give a higher intra-assay variation than with the CV calculation used in this study. More than one method is available for the calculation of the intra- and inter-assay CV. This makes it all the more important that the methods used are clearly reported, to facilitate the comparison between outcomes of past and present studies. It is known that the hypocretin-1 protein, among other proteins, is very hydrophobic [29]. Not using the proper collection or storage tubes can result in loss of hypocretin-1 due to absorption to the tube wall. No studies have been performed to determine the type of tube best suited to transport and store CSF samples for hypocretin-1 measurements. However, based on studies [30–33] regarding hydrophobic proteins in CSF, polypropylene tubes are recommended for storage and collection. For the analysis borosilicate glass tubes are used [34].

We performed two assessments of inter-assay variability without conversion for Stanford, based on the Stanford harmonization and internal control measurements, and found high inter-assay CVs. Our results show a low inter-assay reliability when measurements are not corrected using a harmonization sample. Previous studies have also shown large inter-assay variations although, as with the intra-assay variations, methods of calculation are not mentioned [27, 35]. In contrast, certain other applications of the RIA tend to yield lower inter-assay variations [13, 36–39]. One reason for this is the difference in the unit of measurement used in different applications of the RIA. If the sample contains a large quantity of the substance measured, the concentrations of interfering compounds are relatively lower, reducing the impact of cross-reactivity. Another possible explanation may be that the antibodies in these other RIAs cross-react less with matrix proteins. For example, Sakai et al. [28] found that inter-assay CVs between 2.5% and 6.4% in their hypocretin-1 RIAs, using self-created polyclonal anti-hypocretin-1 antibodies, which were about 20 times more sensitive than the antibodies provided by Phoenix Pharmaceuticals.

After converting the mean internal control concentrations using the Stanford conversion factor, the internal control inter-assay CV was reduced to a more acceptable level. The exact concentration of the internal control sample was unknown in this study, but the average internal control mean after conversion (314 pg/mL) was quite close to the known concentration of the new Stanford harmonization samples (329 pg/mL), which were used in calculating the conversion factors that corrected the internal control sample values. It may be more interesting to see what would happen to the inter-assay CV when converting sample concentrations that are substantially different from the Stanford harmonization sample concentrations used. No repeated sample data were available to conduct such an assessment during this retrospective study.

The effect of outliers was also illustrated by our results, in that two or four outliers can increase the inter-assay CV enormously. The consistent deviation of these RIAs suggests that a systematic human error may have taken place in these cases.

Lower limit of quantification

The LLOQ before Stanford conversion of the hypocretin-1 concentrations was lower than the LLOQ after conversion, which is as expected since most of the samples used were corrected with a conversion factor above 1.0 (average of 1.23). This increases the concentration of most samples, causing the interpolated semilog line to shift to the right.

Previous studies mostly used the limit of detection (LOD) instead of the LLOQ [3, 26, 27, 40]. Our study used the LLOQ, which as mentioned previously, shows the point from which a concentration can be reliably determined. This is clinically more relevant than a detection limit, as the LOD only indicates the lowest concentration distinguishable from background noise. Thus, the presence of an analyte can reliably be detected, but the reported concentration is not necessarily reliable [17]. Past studies have reported hypocretin-1 concentrations below 27.9 pg/mL without the caveat that these concentrations may not be accurate. We suggest that hypocretin-1 concentrations measured using RIA should be reported as undetectable when below 27.9 pg/mL.

Variation in RIA calibration curve range and the impact on RIA reliability

The calibration curves used in our study varied widely (see Figure 3), a variation that can be explained by the fact that, as the RIA kit gets older, the radioactivity of the I^{131} labeled hypocretin-1 decreases. This causes narrowing and lowering of the CPM range, thus flattening its calibration curve.

The effects of this flattening were noticeable in the comparison between the intra-assay CVs from RIAs with an average CPM range above 3000 and RIAs with an average CPM range below 3000, with the latter assumed to be RIAs nearing the supplier-defined expiration date of the RIA. The intra-assay CV was substantially and significantly higher for the older RIAs (CPM range <3000) than for the newer RIAs (CPM range ≥3000). As mentioned above, this is likely due to differences in calibration curve spread. When the calibration curve spread decreases small differences in CPM measurements will have more impact when converting the measured CPM to a hypocretin-1 concentration, causing more variation and less intra-reliability. This effect is also seen when we look at the difference in the LLOQ measured using raw data from RIAs with high and RIAs with low specific activity. While the difference is minimal when using unconverted concentrations, it becomes more evident when using converted concentrations, with the RIAs with higher specific activity having a higher LLOQ, indicating more variability.

Keating et al. [12] also found that RIAs with a longer shelf time have higher intra-assay variations when determining hypocretin-1 concentrations. While presently, per operating instructions, RIAs need to be used within a period of 6 weeks, these results further suggest that expeditious use of RIAs is desirable. Even more so as we found that the radioactivity of Phoenix Pharmaceuticals RIA kits has, on average, been lower in recent years (see Supplementary Figure S1).

Limitations

Some limitations should be mentioned. First of all, for multiple calculations, we could only use the samples that were measured with consecutive RIAs as part of the standard protocol (Stanford harmonization, internal control, and kit samples). Thus while assessment of the accuracy of RIAs at low, intermediate, and high concentration levels is recommended by the EMA (and other) guidelines for the validation of analytical methods [6], this was not possible in this retrospective study. The Stanford

harmonization, internal control and kit samples we used to determine variability all had intermediate concentration levels and, given the sigmoidal shape of calibration curves, it is also of interest to assess the inter-assay variability of low and high concentration samples. Furthermore, these measurements were performed in a single laboratory, which potentially decreases the generalizability of the results. Prospective inter-laboratory research adhering to the recommendations of the FDA and EMA (or other relevant) guidelines would be preferred. We are aware that recently more sophisticated mass spectrometry-based methods for the determination of hypocretin-1 were published, which is a very promising evolution [10, 11, 41]. At the moment, the field of quantitative clinical chemistry proteomics is developing fast [42], and the measurement of CSF hypocretin-1 by this technique can overcome some of the limitations of the RIAs. Still, the RIA remains the only test that is sufficiently validated to assess hypocretin deficiency to diagnose narcolepsy. The enzyme-linked immunosorbent assay (ELISA) cannot be used as a reliable alternative method to measure hypocretin-1 in CSF [41, 43].

Conclusion

In conclusion, the intra-assay variation when measuring hypocretin-1 using the commonly used RIA kit from Phoenix Pharmaceuticals is of an acceptable level. The low intra-assay variability around the current cutoff point of 110 pg/mL is particularly of interest in clinical practice. In contrast, the inter-assay variation, and thus the inter-assay reliability, is too high without conversion using a harmonization sample. After conversion, the inter-assay variation becomes acceptable, though this correction should be tested on samples with different concentrations in the future, preferably including concentrations on more extreme ends of the calibration curve. In addition, the shelf life of the RIA kits should be kept in mind, as the amount of specific activity has a significant impact on intra-assay reliability. A hypocretin-1 concentration below 28 pg/mL should be reported as undetectable. In addition, we find that concentrations around the currently used cutoff value to diagnose NT1 can be determined reliably. Lastly, we recommend standardization of RIAs following our findings. This is achieved by using polypropylene plastic tubes for collection and storage of the CSF, using RIA kits immediately after receiving them from Phoenix Pharmaceuticals preventing a lower specific activity of the RIA, and by applying a harmonization sample. This way, the results of RIAs for hypocretin-1 measurements will be more reliable and comparable across studies.

Supplementary material

Supplementary material is available at *SLEEP* online.

Disclosure Statement

Financial disclosure: S.O. consulted for Bioprojet, Jazz Pharmaceuticals, UCB Pharma, and Takeda, all paid to the institution, and not related to the present work. R.F. consulted for and/or received lecture fees from Bioprojet, Takeda, Lundbeck, TEVA, Lilly, Novartis, and Allergan, and grant support from Jazz Pharmaceuticals and Bioprojet, all not related to the present work. D.B. received a reimbursement for a course from Bioprojet, not related to the present work. G.J.L. consulted for Bioprojet, Jazz Pharmaceuticals, UCB Pharma, NLS, and Takeda, all paid to the institution, and not related to the present work. He also served as a member of advisory boards on narcolepsy: Bioprojet, Jazz Pharmaceuticals, UCB Pharma, and NLS. The other authors have indicated no potential financial conflicts of interest.

Non-financial disclosure: The authors declare that there is no non-financial conflict of interest.

Data Availability

The data are available from the corresponding author upon reasonable request.

References

1. Longstreth WT, Jr., *et al*. The epidemiology of narcolepsy. *Sleep*. 2007;**30**(1):13–26. doi:10.1093/sleep/30.1.13
2. Liblau RS, *et al*. Hypocretin (orexin) biology and the pathophysiology of narcolepsy with cataplexy. *Lancet Neurol*. 2015;**14**(3):318–328.
3. Ripley B, *et al*. CSF hypocretin/orexin levels in narcolepsy and other neurological conditions. *Neurology*. 2001;**57**(12):2253–2258.
4. Nishino S, *et al*. CSF hypocretin levels in Guillain-Barré syndrome and other inflammatory neuropathies. *Neurology*. 2003;**61**(6):823–825.
5. American Academy of Sleep Medicine. In: *International Classification of Sleep Disorders. Diagnostic and Coding Manual*. 3rd edn. Westchester, IL: American Academy of Sleep Medicine. 2014.
6. Mignot E, *et al*. The role of cerebrospinal fluid hypocretin measurement in the diagnosis of narcolepsy and other hypersomnias. *Arch Neurol*. 2002;**59**(10):1553–1562.
7. Postiglione E, *et al*. Narcolepsy with intermediate cerebrospinal level of hypocretin-1. *Sleep*. 2022;**45**(2). doi:10.1093/sleep/zsab285
8. Goldsmith SJ. Radioimmunoassay: review of basic principles. *Semin Nucl Med*. 1975;**5**(2):125–152.
9. Groch MW. Radioactive decay. *Radiographics*. 1998;**18**(5):1247–1256.
10. Bardsen K, *et al*. Considerably lower levels of hypocretin-1 in cerebrospinal fluid is revealed by a novel mass spectrometry method compared with standard radioimmunoassay. *Anal Chem*. 2019;**91**(14):9323–9329.
11. Hirtz C, *et al*. From radioimmunoassay to mass spectrometry: a new method to quantify orexin-A (hypocretin-1) in cerebrospinal fluid. *Sci Rep*. 2016;**6**:25162.
12. Keating G, *et al*. Hypocretin measurement: shelf age of radioimmunoassay kit, but not freezer time, influences assay variability. *Scand J Clin Lab Invest*. 2017;**77**(5):390–393.
13. Hanneman SK, *et al*. Estimating intra- and inter-assay variability in salivary cortisol. *Biol Res Nurs*. 2011;**13**(3):243–250.
14. Rodbard D. Statistical quality control and routine data processing for radioimmunoassays and immunoradiometric assays. *Clin Chem*. 1974;**20**(10):1255–1270.
15. Sudhansu C, Luigi F-S. *Oxford Textbook of Sleep Disorders*. Oxford, UK: Oxford University Press; 2017.

16. Nishino S, *et al*. Low cerebrospinal fluid hypocretin (orexin) and altered energy homeostasis in human narcolepsy. *Ann Neurol*. 2001;**50**(3):381–388.
17. Armbruster DA, *et al*. Limit of blank, limit of detection and limit of quantitation. *Clin Biochem Rev*. 2008;**29**(Suppl 1):S49–S52.
18. Wolfinger RD, *et al*. Two approaches for estimating the lower limit of quantitation (LLOQ) of microRNA levels assayed as exploratory biomarkers by RT-qPCR. *BMC Biotechnol*. 2018;**18**(1):6.
19. Han F, *et al*. HLA DQB1*06:02 negative narcolepsy with hypocretin/orexin deficiency. *Sleep*. 2014;**37**(10):1601–1608. doi:10.5665/sleep.4066
20. European Medicines Agency. Guideline on bioanalytical method validation. 2011. www.ema.europa.eu/docs/en GB/document library/Scientific guideline/2011/08/WC500109686.pdf. Accessed January 24, 2021.
21. Influentialpoints. Coefficient of variation. https://influentialpoints.com/Training/coefficient_of_variation-principles-properties-assumptions.htm. Accessed June 17, 2020.
22. Luraschi P, *et al*. Updating biological variation databases and selecting reasonably correct published values. *Scand J Clin Lab Invest*. 2000;**60**(8):733–735.
23. Fraser CG. Biological variation in clinical chemistry. An update: collated data, 1988–1991. *Arch Pathol Lab Med*. 1992;**116**(9):916–923.
24. Administration, Food. USFDA. *Guidance for Industry: Bioanalytical Method Validation*. Silver Spring, Maryland: US Department of Health and Human Services, Food and Drug Administration. 2001; 66.
25. Viswanathan CT, *et al*. Quantitative bioanalytical methods validation and implementation: best practices for chromatographic and ligand binding assays. *Pharm Res*. 2007;**24**(10):1962–1973.
26. Nishino S, *et al*. Hypocretin (orexin) deficiency in human narcolepsy. *Lancet*. 2000;**355**(9197):39–40.
27. Heier MS, *et al*. CSF hypocretin-1 levels and clinical profiles in narcolepsy and idiopathic CNS hypersomnia in Norway. *Sleep*. 2007;**30**(8):969–973. doi:10.1093/sleep/30.8.969
28. Sakai N, *et al*. HPLC analysis of CSF hypocretin-1 in type 1 and 2 narcolepsy. *Sci Rep*. 2019;**9**(1):477.
29. Hungs M, *et al*. Hypocretin/orexin, sleep and narcolepsy. *Bioessays*. 2001;**23**(5):397–408.
30. Hansson O, *et al*. The impact of preanalytical variables on measuring cerebrospinal fluid biomarkers for Alzheimer's disease diagnosis: a review. *Alzheimers Dement*. 2018;**14**(10):1313–1333.
31. Toombs J, *et al*. Amyloid β peptides are differentially vulnerable to preanalytical surface exposure, an effect incompletely mitigated by the use of ratios. *Alzheimers Dement (Amst)*. 2018;**10**:311–321.
32. Vanderstichele HM, *et al*. Optimized standard operating procedures for the analysis of cerebrospinal fluid Aβ42 and the ratios of Aβ isoforms using low protein binding tubes. *J Alzheimers Dis*. 2016;**53**(3):1121–1132.
33. Simonsen AH, *et al*. Pre-analytical factors influencing the stability of cerebrospinal fluid proteins. *J Neurosci Methods*. 2013;**215**(2):234–240.
34. Goebel-Stengel M, *et al*. The importance of using the optimal plasticware and glassware in studies involving peptides. *Anal Biochem*. 2011;**414**(1):38–46.
35. Lopez R, *et al*. Temporal changes in the cerebrospinal fluid level of hypocretin-1 and histamine in narcolepsy. *Sleep*. 2017;**40**(1). doi:10.1093/sleep/zsw010
36. Heilmann RM, *et al*. Development and analytical validation of a radioimmunoassay for the measurement of alpha$_1$-proteinase inhibitor concentrations in feces from healthy puppies and adult dogs. *J Vet Diagn Invest*. 2011;**23**(3):476–485.
37. Buono A, *et al*. Development, analytical validation, and initial clinical evaluation of a radioimmunoassay for the measurement of soluble CD25 concentrations in canine serum. *Vet Immunol Immunopathol*. 2019;**215**:109904.
38. Benabdelaziz A, *et al*. Development of in-house RIA kit for progesterone in cow's skim milk. *J Immunoass Immunochem*. 2020;**41**(2):195–207.
39. Deberg M, *et al*. Highly specific radioimmunoassay for human insulin based on immune exclusion of all insulin precursors. *Clin Chem*. 1998;**44**(7):1504–1513.
40. Knudsen S, *et al*. Validation of the ICSD-2 criteria for CSF hypocretin-1 measurements in the diagnosis of narcolepsy in the Danish population. *Sleep*. 2010;**33**(2):169–176. doi:10.1093/sleep/33.2.169
41. Lindström M, *et al*. Orexin-A measurement in narcolepsy: a stability study and a comparison of LC-MS/MS and immunoassays. *Clin Biochem*. 2021;**90**:34–39.
42. Smit NPM, *et al*. The time has come for quantitative protein mass spectrometry tests that target unmet clinical needs. *J Am Soc Mass Spectrom*. 2021;**32**(3):636–647.
43. Ono T, *et al*. Measurement of cerebrospinal fluid orexin-A (hypocretin-1) by enzyme-linked immunosorbent assay: a comparison with radioimmunoassay. *Psychiatry Clin Neurosci*. 2018;**72**(11):849–850.