Introduction

The ValidSense toolbox aims to assess the agreement between two quantitative methods or devices measuring the same quantity using the **Limits of Agreement analysis** (LoA analysis), also known as the Bland-Altman analysis, using four existing variants. Moreover, a **Longitudinal analysis** is developed to assess agreement over time.

ValidSense consists of five pages, shown in the sidebar on the left. Follow these pages sequentially: Before the LoA analysis or the longitudinal analysis can be performed, the loading and preprocessing pages must be followed sequentially.

Benefits of ValidSense

- First-time right assessment of the agreement between two quantitative methods, without requiring in-depth statical knowledge or programming skills.
- Guidance on the LoA analysis.
- Assess statistical assumptions for the LoA analysis.
- **Four existing variants of the LoA analysis** are included to correct methodological problems in the data. See Variants of the LoA analysis for more information [LINK],
    - Clustering: For example, when multiple measurements within a subject are recorded.
    - Non-constant agreement over the measurement range: For example, in respiratory rate measurement, measurements are unlikely to fall below a certain threshold, and the variability increases with the mean of the respiratory rate.
- New developed **longitudinal analysis** to assess non-constant agreement over time. For example, to assess sensor drift or patient drift over time.

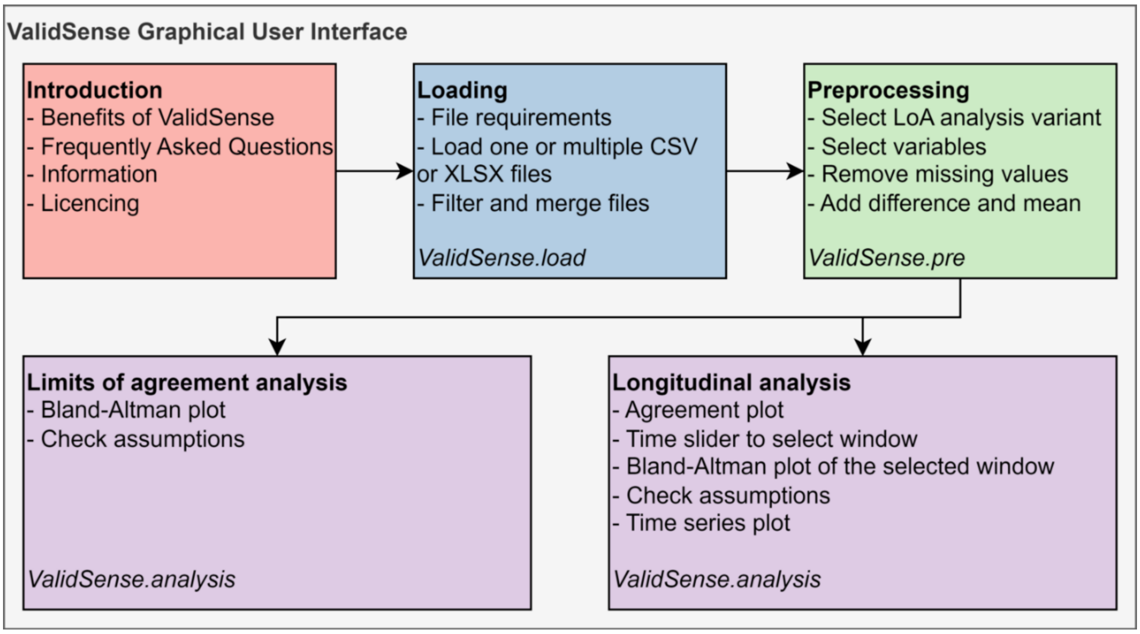

Figure 1: Overview of the five pages in the ValidSense toolbox. Sequential steps are required to perform the LoA analysis or Longitudinal analysis.

Frequently Asked Questions

- **What is the LoA analysis?**

  The **LoA analysis** compares two measurement techniques to determine their agreement. In a **Bland-Altman plot**, each datapoint represents a pair of measurements, with the horizontal axis representing the average of the two measurements and the vertical axis representing the difference between the two measurements. The Bland-Altman plot also includes a line indicating the bias (accuracy) between the two measurements and lines indicating the upper and lower LoA (precision), which define the range within which 95% of the differences between the two measurements are expected to fall. Accuracy refers to the proximity of measurements to the actual value, while precision represents the variability in repeated measurements. Figure 2 shows the relationship between the accuracy and bias, and the precision and 95% LoA. The LoA analysis computes the agreement intervals but does not evaluate the acceptability of these boundaries, which should be determined based on clinical considerations. If the two devices show sufficient agreement, they can be used interchangeably.

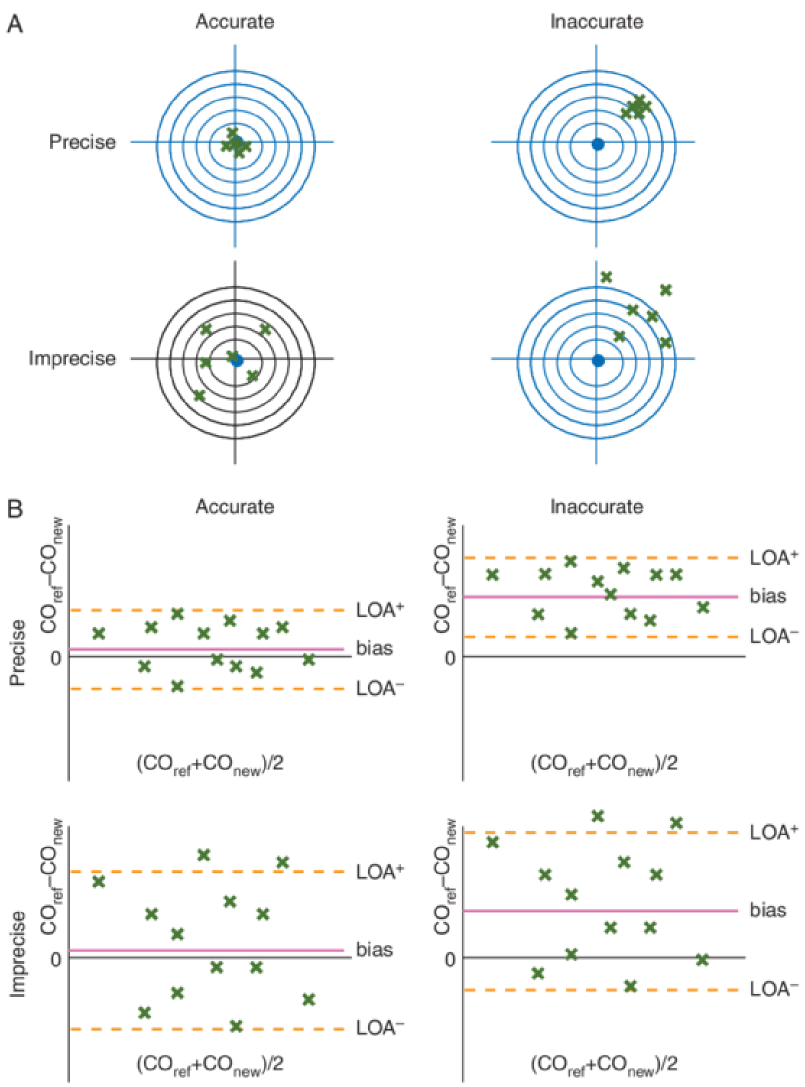

Figure 2: Bias and the limits of agreement representing the accuracy and precision between two devices. A) Accurate measurements are close to the true value, irrespective of the spread of the measurements. In contrast, precise measurements are close to each other, irrespective of their deviation from the true value. B) In Bland-Altman plots, accurate cardiac output monitors show a bias (continuous line) close to the zero line. In contrast, precise monitors show limits of agreement close to the bias (dotted lines). Figure with permission derived from [Montenij et al (2016)](https://pubmed.ncbi.nlm.nih.gov/27199309/).

- **Which variants of the existing LoA analysis are included in the ValidSense toolbox?**

  The LoA analysis was first reported in British Medical Journal in 1986, known as the classic LoA analysis. Since then, various variations of this analysis have been introduced to serve different purposes:

	LoA analysis variant	Methodological challenge	Intended use
0	Classic		Assess agreement in single pair of measurements per patient.
1	Repeated measurements	Clustering	Assess agreement in multiple measurements per patient.
2	Mixed-effects	Clustering (can correct for multiple effects)	Assess agreement based on the mixed-effects LoA analysis, allowing to correct, for example, multiple measurements per patient or systematic relationship between the difference and mean.
3	Regression of difference	Non-constant agreement over the measurement range	Assess agreement in a single measurement per patient, with a linear relationship between difference and mean for bias and/or LoA.

An example of the non-constant agreement over the measurement range is presented in figure 3.

Link to original articles:

- Classic: [Bland & Altman 1986]
- Repeated measurements: [Bland & Altman 1999 section 5.2], [Bland & Altman 2007 section 3]
- Regression of difference [Bland & Altman 1999 section 3.2]
- Mixed effect model [Parker et al. 2016]

Figure 3: Example of a Bland-Altman plot with a non-constant agreement in respiratory rate measurements. The variability increases with the mean of the respiratory rate. The regression of difference LoA analysis represents the systematic relationship between the difference and mean.

- **Why is the mixed-effect approach preferred above the repeated measurements approach?**

  We believe that the mixed-effect LoA analysis is preferred for correcting for clustering. The included subjects are a representative sample of the population of interest in scientific studies, including validation studies. Considering subjects as random effects recognises that they are a subset of a larger population in contrast to a fixed-effect approach assuming that the included subjects represent the entire population. Therefore, the regarding subjects as a random effect in the mixed-effect LoA analysis is preferred above the repeated measurements approach.

- **Why is the LoA analysis the correct method for the assessment of agreement between two quantitative methods or devices?**

  The LoA analysis is the best statistical method for assessing the agreement between two quantitative methods or devices. Correlation and regression studies are frequently proposed. However, correlation examines the magnitude and significance of the relationship between two variables, and regression predicts the best relationship between two variables by quantifying the goodness of fit. These two methods assess the **relationship's strength**, not the **agreement's quantification**. A high correlation does not automatically imply a good agreement between two variables. In other words, the correlation and regression methods evaluate the standard error rather than the standard deviation of the variables. To summarise, the appropriate approach to evaluate the agreement between two variables is to consider their differences using the LoA analysis.

- **What is the newly developed longitudinal analysis?**

  This method is developed to assess the methodological challenge of **non-constant agreement over time** when there is a drift in the accuracy over time. Drift refers to the gradual shift in baseline values of the measured physiological parameter over time. The accuracy could change over time, such as (I) sensor drift (e.g. less accurate measurements of the device after the moment of calibration) and (II) patient drift (e.g. movement, positioning, health status, or medication).

  Existing LoA analyses do not consider changes in accuracy and precision over time, which is why the **longitudinal analysis** was created. The longitudinal analysis involves breaking down a dataset into smaller parts over time and applying **existing LoA analysis** to each part. A moving time window is applied, and based on the data included in the window, the bias and 95% LoA are calculated for each time window. The classic, repeated measurements or mixed-effects LoA analysis calculates the bias and 95% LoA. A constant agreement over the measurement range is assumed.

  The **agreement plot** was developed to visualisation the outcomes of the longitudinal analysis to provide insight into the accuracy and precision over time. Figure 3 provides an example of the agreement plot. The y-axis shows the differences between the two devices (similar to the Bland-Altman plot), while the x-axis represents the start time of the window. The bias- and 95% LoA-lines indicate the accuracy and precision over the time windows. The advantage of the agreement plot is that it facilitates the identification of trends or patterns over time that may go unnoticed otherwise. Exploring the cause of changes is the subsequent step, although this falls beyond the scope of this thesis.

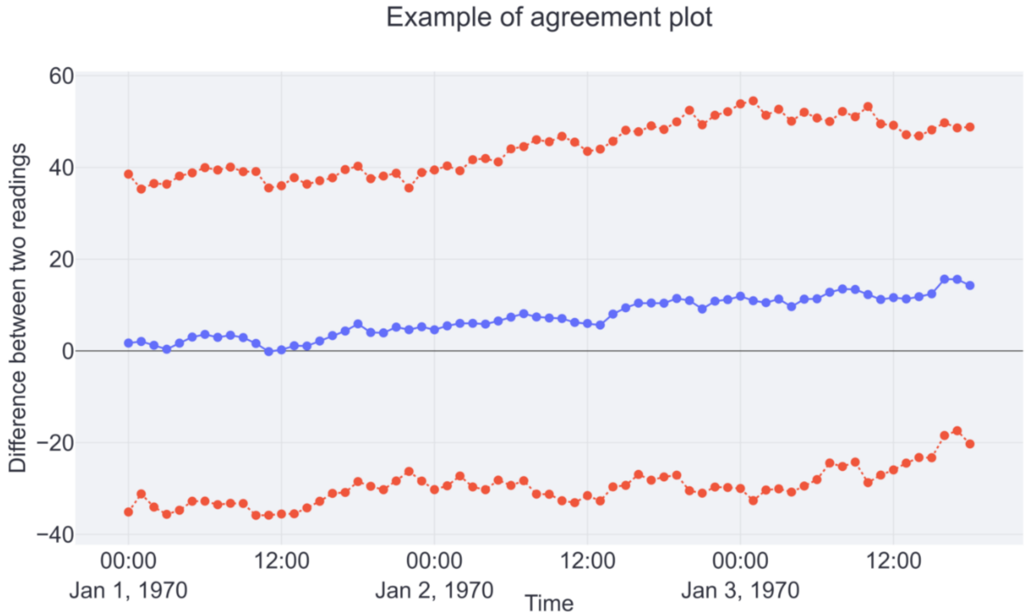

Figure 4: Example of an agreement plot indicating non-constant agreement over time. Bias (blue line) and 95% LoA (red dotted line) within the time window of six hours are presented.

- **What are time series plots?**

  Time series plots are traditional plots to help identify changes over time (example is given in below). The time series plot scatters the measurements from two devices, with the measurement value on the y-axis and the timestamp on the x-axis. The two devices are distinguished by different symbols, either a circle or a square. A moving median trendline is added to identify trends to smooth out the high variation between sequential measurements.

Figure 5: Example of time series plot, with a median moving window of 30 measurements.

- **What is the difference between the agreement plot and the time series plot?**

  The time series plot scatters all datapoints, and draws a trend line through the measurements, indicating the accuracy over time. The agreement plot gives an estimate of the accuracy and precision over time.

- **Why should the statistical assumptions be assessed?**

  Complying with the assumptions is essential for the **validity of the LoA analysis**. The toolbox guides on meeting the LoA analysis's assumptions and includes six built-in tools (histogram, Q-Q plot, scatter plot, within-cluster-SD plot, residual plot, and covariance) to verify these assumptions. The three assumptions that always must be checked are mentioned: First, the normal distribution of the differences can be checked using the histogram and Q-Q plot to ensure that 95% of paired measurements fall within the 95% LoA interval. Second, constant agreement over the measurement range to ensure that the bias and 95% LoA represent accuracy and precision. Third, independent measurements (e.g. violated in case of clustering) to ensure that the 95% LoA represent the precision between measurements. We want to emphasise that it is **essential for users to check the statistical assumptions to ensure the validity of the LoA analysis**. More information about the statistical assumptions can be found in Appendix A of this thesis (LINK TO BE ADDED).

- **Are the four LoA analyses correctly implemented in the ValidSense toolbox?**

  In the thesis (LINK TO BE ADDED) we checked if the four existing LoA analyses were correctly implemented.

- **Is the toolbox modular built?**

  We built the toolbox modular in Python to allow for adaptation and allow the use of the user's software. The source information is on GitHub.

Information

Version ValidSense: 1.0.1 (2 May 2025)

Source code: Github

Explanation: Validation of Vital Sign Monitoring Devices - University of Twente Student Theses

Licencing