California State University, Monterey Bay

Digital Commons @ CSUMB

An Alternative Choice for the Critical Value of Limits of Agreement and Simulation-Based Sample Size Calculation in Bland Altman Analysis

Steven B. Kim
California State University, Monterey Bay, stkim@csumb.edu

Follow this and additional works at: https://digitalcommons.csumb.edu/math_fac

AN ALTERNATIVE CHOICE FOR THE CRITICAL VALUE OF LIMITS OF AGREEMENT AND SIMULATION-BASED SAMPLE SIZE CALCULATION IN BLAND ALTMAN ANALYSIS

STEVEN B. KIM

Department of Mathematics and Statistics
California State University, Monterey Bay
100 Campus Center
Seaside, CA 93955
USA
e-mail: stkim@csumb.edu

Abstract

Bland Altman analysis is a statistical method for assessing the degree of agreement between two methods of measurement. In medical and health sciences, it is a popular method because of its simple calculation and visualization. Under normality assumption, the calculation is based on two sufficient statistics $\bar{d}$ and s, where $\bar{d}$ is the sample mean of differences and s is the sample standard deviation of the differences. The interval $\bar{d} \pm 1.96s$ is referred to as 95% limits of agreement (LOA) in literature. In a seminar paper, Bland and Altman [2] interpreted LOA as "If the differences are normally distributed, 95% of differences will lie between these limits". This interpretation seems to be widely accepted, but there is a caveat because the coverage probability of LOA is a random variable. In this article, we demonstrate the sampling distribution of its coverage probability by simulation, and we discuss

an alternative choice for the critical value. In addition, using simulation, we perform sample size calculation which satisfies a specified condition for the sampling distribution of coverage probability.

1. Introduction

Researchers and practitioners in medical and health sciences are interested in accurate and precise measurement methods. They make important decisions based on measurement readings, and there is often a gold standard method of measurement. When a new method of measurement is developed, researchers would like to compare the validity and reliability of the new method. Bland Altman analysis is a method of assessing the degree of agreement between the gold standard method and the new measurement method (Altman and Bland [1, 2]). It gained popularity due to its simple calculation and visualization, it has been understood by many researchers and practitioners, and it has been further developed to account for more complex situations (Hopkins [3]; Giavarina [4]; Schluter [5]; and Taffe [6]).

In a seminar paper of Bland and Altman [2], under the normality assumption for differences between measurement by one method and measurement by another method, the degree of agreement between two measurements is assessed by an interval referred as limits of agreement (LOA). It is an approximate 95% prediction interval (PI) for differences with the critical value of 1.96 or 2 calculated from the standard normal distribution. An exact 95% PI can be used with the critical value which can be obtained from a T distribution (Geisser [7]). Whether it is approximate or exact, in this article, we challenge the common interpretation of 95% LOA. The coverage probability of 95% LOA can converge to 0.95 as the sample size n increase, but it cannot be exactly 0.95 for finite n. In fact, under- or over-coverage can be severe for small n. Therefore, it is worth discussing the sampling distribution of the coverage probability of LOA. In this article, we demonstrate the sampling distribution using simulation.

For conservative researchers, the sampling distribution of coverage probability by a 95% PI can be worrisome for small n. To address this concern, severe under-coverage in particular, a 95% tolerance interval (TI) is considered for 95% LOA instead of a 95% PI. A TI has a different critical value which is devised to guarantee the coverage probability of 0.95 or higher for a fixed confidence level (Howe [8]; Guenther [9]; Vardeman [10]; and Hahn et al. [11]). To avoid the opposite concern of being too conservative, a reasonably large sample size is needed. In this regard, we can find a sample size such that the coverage probability of 95% LOA is about 0.95 within a given range at a fixed confidence level. In this article, a simulation-based method of sample size calculation is demonstrated, and a sample code in statistical software R is given in the Appendix for practitioners.

The sample size calculation discussed in this article is to control the coverage probability of LOA (PI or TI) which is different perspective from the sample size calculation to perform hypothesis testing for a fixed significance level, statistical power and effect size discussed by Lu et al. [12]. The sample size calculation discussed in this article helps more accurate interpretation of 95% LOA.

2. Choosing the Critical Value in Bland Altman Analysis

2.1. Limits of agreement (LOA)

Let (x_i, y_i) be paired measurements obtained from the i-th subject for $i = 1, 2, \ldots, n$, where x_i is observed from a current standard method and y_i is from a new method. Let $d_i = y_i - x_i$, the difference between two observed outcomes for $i = 1, \ldots, n$. Assume $(d_1, \ldots, d_n)$ is a simple random sample from a normal distribution with unknown mean μ and unknown standard deviation σ. Let $\bar{d} = \frac{1}{n} \sum_{i=1}^{n} d_i$ and $s = \frac{1}{n-1} \sum_{i=1}^{n} (d_i - \bar{d})^2$ be the sample mean and standard deviation of the

differences. Let $l = \overline{d} - \gamma s$ and $u = \overline{d} + \gamma s$ be a lower limit and an upper limit, for some critical value γ devised to capture future values d_{n+1}, $d_{n+2}, \ldots$ with a target coverage probability of $\tau = 0.95$. Then the interval (l, u) is referred to as 95% limits of agreement (LOA).

2.2. Challenging interpretation of LOA

In a seminar paper of Bland and Altman [2], γ was approximated by $\gamma_1 = z_{0.975} = 1.96$, the 97.5th percentile of the standard normal distribution, and resulting LOA was interpreted as "If the differences are normally distributed, 95% of differences will lie between these limits". It is not an accurate interpretation particularly when n is not large. It is meant to cover 95% of differences to be observed from the population, but the coverage probability of resulting LOA must be less or greater than 0.95. Since l and u are random from sample to sample, and a future value d_{n+1} is also random, the true coverage probability p is not a fixed number from sample to sample. In fact

$$p = \int_l^u \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt, \tag{1}$$

is random from sample to sample, so it has a sampling distribution which depends on n.

Given n, the sampling distribution of p can be simulated as follow. From Equation (1), we have

$$p = \int_{l^*}^{u^*} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt,$$

where

$$l^* = \frac{\overline{d} - \mu}{\sigma} - \gamma \frac{s}{\sigma},$$

$$u^* = \frac{\overline{d} - \mu}{\sigma} + \gamma \frac{s}{\sigma}.$$

Note that $(\bar{d} - \mu)/\sigma$ is from $N(0, n^{-1})$, and $(n-1)(s/\sigma)^2$ is from χ^2_{n-1}, and $\bar{d}$ and s are independent. To this end, we can simulate z from $N(0, 1)$ and w from χ^2_{n-1} independently, then let $z^* = z/\sqrt{n}$ and $w^* = \sqrt{w/(n-1)}$ to simulate $l^* = z^* - \gamma w^*$ and $u^* = z^* + \gamma w^*$. Then the coverage probability is $\Phi(u^*) - \Phi(l^*)$, where Φ is the cumulative distribution function of $N(0, 1)$. Figure 1 shows simulated sampling distributions based on 100,000 replicates for $n = 10, 20, 30, 50$, and it demonstrates why providing an accurate interpretation of observed LOA is challenging. The approximated probability of $p \geq 0.95$ is 0.369, 0.403, 0.419, 0.437 for $n = 10, 20, 30, 50$, respectively.
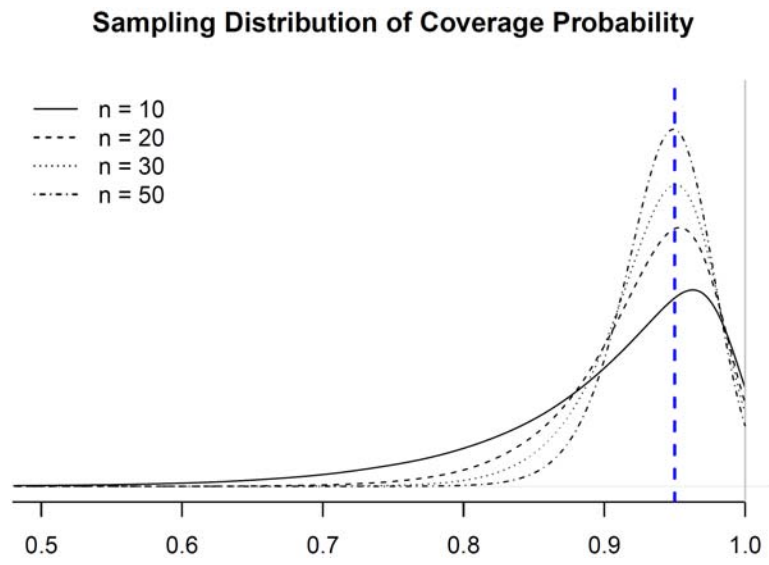
Sampling Distribution of Coverage Probability

Figure 1. Sampling distribution of coverage probability p by LOA when $\gamma = \gamma_1 = 1.96$.

2.3. Prediction interval

If the author correctly understood the original motivation of Bland and Altman [2], the LOA with $\gamma_1 = 1.96$ is meant to approximate a 95% prediction interval (PI) because the exact choice of γ for a 95% PI is

$$\gamma_2 = t_{0.975, n-1} \sqrt{1 + \frac{1}{n}}, \tag{2}$$

where $t_{0.975, n-1}$ is the 97.5th percentile of T distribution with $n-1$ degrees of freedom (Geisser [7]). (Note that $\gamma_2 \to \gamma_1 = 1.96$ as $n \to \infty$.) Using the critical value γ_2, simulated sampling distributions of coverage probability are shown in Figure 2 for n = 10, 20, 30, 50. In the simulation, the approximated probability of $p \geq 0.95$ is 0.668, 0.614, 0.593, 0.572 for n = 10, 20, 30, 50, respectively.
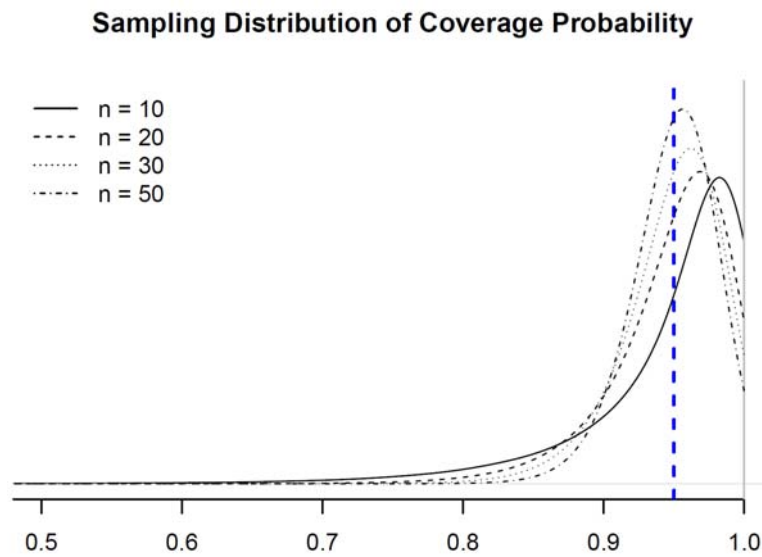

Figure 2. Sampling distribution of coverage probability p by LOA when $\gamma = \gamma_2$.

2.4. Tolerance interval

When a practitioner desires $p \geq \tau = 0.95$ with a probability κ before analyzing data, κ is referred to as the confidence level. A 95% tolerance interval (TI) at confidence level κ is devised for this purpose. In this case, the choice of the critical value γ is

$$\gamma_3 = z_{0.975}\sqrt{\frac{n^2 - 1}{nc_{1-\kappa,\, n-1}}},$$

where $c_{1-\kappa,\, n-1}$ is the $100(1-\kappa)$-th percentile of chi-square distribution with $n-1$ degrees of freedom (Howe [8]). At $\kappa = 0.9$, using simulation of 100,000 replicates, the approximated probability of $p \geq 0.95$ is 0.898, 0.899, 0.900, 0.899 for $n = 10, 20, 30, 50$, respectively, as shown in Figure 3.
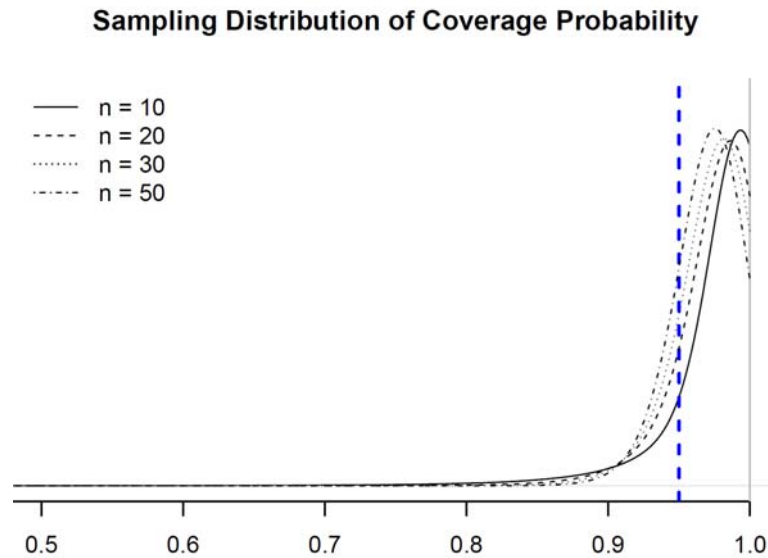
Sampling Distribution of Coverage Probability

Figure 3. Sampling distribution of coverage probability p by LOA when $\gamma = \gamma_3$.

There are alternative choices of γ for a TI. For example, an alternative choice is $\gamma_4 = \alpha\gamma_3$ with a correction factor

$$\alpha = \sqrt{1 + \frac{n - 3 - c_{1-\kappa, n-1}}{2(n + 1)^2}}$$

(Guenther [9]). When γ_4 is used for a TI, the sampling distribution of p is very close to Figure 3, and the approximated probability of $p \geq 0.95$ is 0.902, 0.901, 0.902, 0.900 for $n = 10, 20, 30, 50$, respectively, using simulation of 100,000 replicates.

2.5. Trade-off between confidence and precision (length)

Bland and Altman [2] provided an example of comparing two methods (a large meter and a mini meter) of measuring peak expiratory flow rate (l/min). The LOA was calculated by sufficient statistics $\bar{d} = -2.1$ and $s = 38.8$ from a sample of size $n = 17$. Using the approximate critical value $\gamma = 1.96$, the resulting 95% LOA was $d \pm 1.96s = (-78.1, +73.9)$. Using $\gamma = \gamma_2 = 2.181$, the exact critical value of a 95% PI, resulting 95% LOA would be $(-86.7, +82.5)$. The resulting LOA was interpreted as "95% of differences will lie between these limits". By carefully observing Figure 2 for $n = 10$ and $n = 20$, however, it is likely that the true coverage probability is far below 0.95. Using simulation, for a sample of size $n = 17$, we can approximate that the coverage probability is below 0.9 with a probability 0.25 when $\gamma = 1.96$ and 0.12 when $\gamma = 2.181$.

Using $\gamma = \gamma_3 = 2.644$, the critical value of a 95% TI at confidence level $\kappa = 0.9$, the 95% LOA would be $(-104.7, +100.5)$ which is 1.2 times longer than $(-86.7, +82.5)$. At the cost of increasing the length by 1.2 times, we used the procedure which guarantees "at least 95% of differences will lie between these limits with a probability $\kappa = 0.9$". For conservative researchers, this operating characteristic may be worth the cost.

For a given confidence level $\kappa,$ we define the relative length of a 95% TI to a 95% PI as

$$\rho = \frac{\gamma_3}{\gamma_2} = \frac{t_{0.975, n-1}\sqrt{1 + \frac{1}{n}}}{z_{0.975}\sqrt{\frac{n^2 - 1}{nc_{1-\kappa, n-1}}}}.$$

Figure 4 shows ρ with respect to n at $\kappa = 0.9, 0.95, 0.99.$ The relative length ρ is between 1.09 and 1.27 for $10 \leq n \leq 100$ for $\kappa = 0.9,$ between 1.12 and 1.43 for $\kappa = 0.95,$ and between 1.18 and 1.80 for $\kappa = 0.99.$
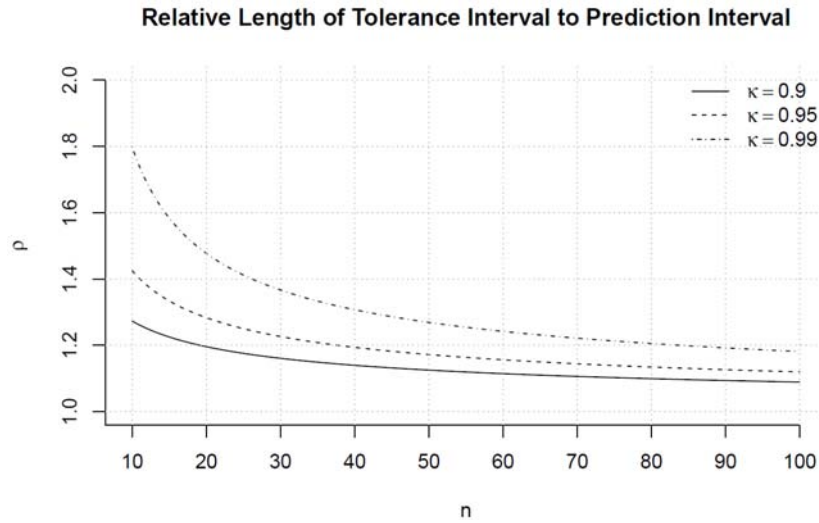
Relative Length of Tolerance Interval to Prediction Interval

Figure 4. Relative length of a 95% TI to a 95% PI at $\kappa = 0.9, 0.95, 0.99.$

3. Sample Size Calculation

3.1. Simulation-based sample size calculation for desired rate of coverage probability

In a small-sample study, 95% LOA with γ_2 suffers from under-coverage (i.e., likely to see $p \ll 0.95$), and 95% LOA with γ_3 for $\kappa \geq 0.9$ can be too conservative (i.e., likely to see $p \gg 0.95,$ close to one). In

practice, we may desire $|p - 0.95| \leq \epsilon$ with for a fixed probability κ and a fixed small $\epsilon > 0$. Using the simulation, we can determine n which satisfies

$$\Pr(|p - 0.95| \leq \epsilon) = \kappa. \tag{3}$$

By letting $\gamma = \gamma_2$ of Equation (2), the exact critical value for a 95% PI, we can search for n such that Equation (3) is satisfied for given κ and ϵ.

A more general procedure for $(\tau \times 100)\%$ LOA follows, where τ is a target coverage probability. For a $(\tau \times 100)\%$ PI, we can search for n such that

$$\Pr(|p - \tau| \leq \epsilon) = \Pr(\tau - \epsilon \leq p \leq \tau + \epsilon) = \kappa$$

by choosing γ to be $\gamma_2 = t_{1-(1-\tau)/2, n-1}$. Similarly, for a $(\tau \times 100)\%$ TI, we can search for n such that

$$\Pr(\tau \leq p \leq \tau + 2\epsilon) = \kappa - \delta$$

by choosing γ to be

$$\gamma_3 = z_{1-(1-\tau)/2} \sqrt{\frac{n^2 - 1}{nc_{1-\kappa, n-1}}}.$$

Distinction between κ and δ is needed for the TI approach because a $(\tau \times 100)\%$ TI requires $P(p \geq \tau) = \kappa$ at confidence level κ by definition.

3.2. Numerical example of sample size calculation

For a PI with $\tau = 0.95$, $\epsilon = 0.01$, $\kappa = 0.9$, and $\gamma_2 = t_{0.975, n-1}$, our goal is to find the sample size n such that $\Pr(|p - 0.95| \leq 0.01) = 0.9$. By simulation, we can search from low n to high n until the condition is satisfied. As shown in Figure 5, we find $n = 710$ by simulation of 100,000 replicates for each candidate n. Instead of the naive search from small n to high n, the bisection method can be more efficient computationally. A sample code of bisection method is given in the Appendix.

Table 1 shows resulting n for γ_2 which satisfies $\Pr(|p - \tau| \le \epsilon) = \kappa$ for $\tau = 0.9, 0.95, 0.99$; $\epsilon = 0.01, \ldots, 1 - \tau$; $\kappa = 0.9, 0.95, 0.99$. Table 2 shows resulting n for γ_3 which satisfies $\Pr(\tau \le p \le \tau + 2\epsilon) = \kappa - \delta$ for comparable values of τ, ϵ, κ, and δ. It appears that the use of γ_3 (i.e., TI approach targeting above τ) requires a smaller sample size than the use of γ_2 (i.e., PI approach targeting around τ) for similar conditions. For example, we need $n = 1562$ to satisfy $\Pr(|p - 0.9| \le 0.01) = 0.9$ by a 90% PI ($\gamma_2 = 1.646$), and we need $n = 1400$ to satisfy $\Pr(0.9 \le p \le 0.92) = 0.9$ by a 90% TI ($\gamma_3 = 1.698$ at $\kappa = 0.95$ and $\delta = 0.05$). For another example, we need $n = 173$ to satisfy $\Pr(|p - 0.95| \le 0.02) = 0.9$ by a 95% PI ($\gamma_2 = 1.974$), and we need $n = 76$ to satisfy $\Pr(0.95 < p < 0.99) = 0.9$ by a 95% TI ($\gamma_3 = 2.282$ at $\kappa = 0.95$ and $\delta = 0.05$).
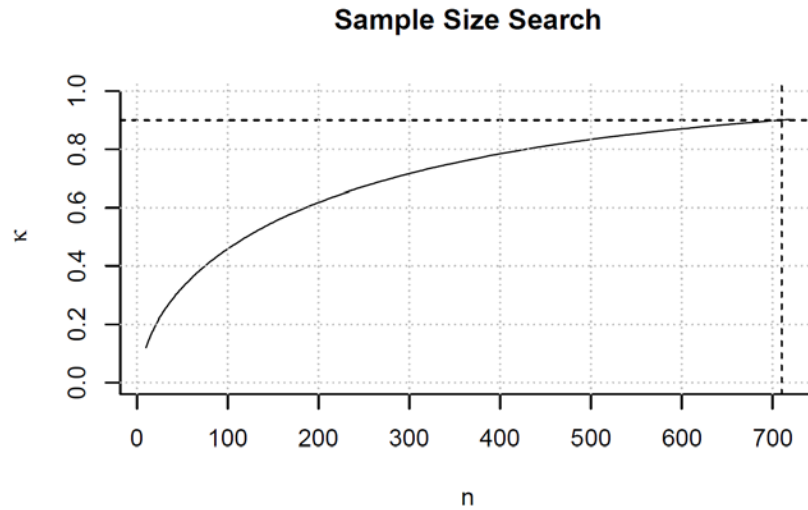

Figure 5. Searching for sample size for $\kappa = 0.9$ and $\epsilon = 0.01$.

Table 1. Searching for sample size for using γ_2 (PI approach)

τ	ϵ	$\kappa = 0.9$	$\kappa = 0.95$	$\kappa = 0.99$
0.90	0.01	**1562**	2210	3822
0.90	0.02	389	553	958
0.90	0.03	173	246	430
0.90	0.04	96	138	246
0.90	0.05	61	88	162
0.90	0.06	42	61	116
0.90	0.07	30	45	88
0.90	0.08	22	35	70
0.90	0.09	17	28	58
0.90	0.10	13	24	48
0.95	0.01	710	1004	1745
0.95	0.02	**173**	250	453
0.95	0.03	73	109	215
0.95	0.04	38	62	130
0.95	0.05	22	42	90
0.98	0.01	195	289	552
0.98	0.02	37	76	170
0.99	0.01	53	110	248

Table 2. Searching for sample size for using γ_3 (TI approach)

τ	ϵ	$(\kappa, \delta) = (0.95, 0.05)$	$(\kappa, \delta) = (0.99, 0.04)$	$(\kappa, \delta) = (0.995, 0.005)$
0.90	0.005	5927	9092	14455
0.90	0.010	**1400**	2156	3410
0.90	0.015	585	900	1418
0.90	0.020	306	472	742
0.90	0.025	180	279	434
0.90	0.030	114	175	273
0.90	0.035	74	114	175
0.90	0.040	48	74	114
0.90	0.045	30	46	70
0.95	0.005	2488	3834	6048
0.95	0.010	526	814	1278
0.95	0.015	188	291	454
0.95	0.020	**76**	118	181
0.98	0.005	526	811	1278

4. Discussion

In Bland Altman analysis, particularly in a small-sample experiment, inaccurate interpretation of LOA is often overlooked in literature. As demonstrated by the sampling distribution of p, under-coverage can be severe when n is small, but the interpretation of LOA does not account for this fact. For conservative researchers, it is recommended to use a 95% TI for 95% LOA instead of using an approximated 95% PI. The critical value of a 95% TI guarantees that the coverage probability is at least 0.95 at a fixed confidence level κ. This known operating characteristic may be worth the cost of wider limits. If a researcher wants to determine a sample size for LOA based on the operating characteristic of PI or of TI, the simulation method can be used as demonstrated in Section 3 with the code in the Appendix.

Under the normality assumption, the 95% LOA, which is calculated by $\bar{d} \pm 1.96s$, can serve as the point estimation for the population parameters, namely the 2.5th and 97.5th percentiles of the population distribution of the differences between measurements provided by two devices. The original interpretation of 95% LOA (i.e., capturing 95% of random differences observed from the population) is valid when resulting 95% LOA is equal to the true values of $\mu \pm 1.96\sigma$, but it is unlikely or impossible. When $\bar{d} \pm 1.96s$ is devised from the perspective of capturing random differences, the coverage probability is a random variable, so a special care is needed for its interpretation.

Some authors mentioned and addressed proportional bias and homoscedasticity of variance of measurement errors (Taffe [6, 14]; Nawarathna and Choudhary [15, 16]; and Carstensen [17]). In this complicated case, simultaneous confidence bands based on a regression model would be more appropriate than a pointwise interval. If such a case is predictable before data collection, the simulation design for sample size calculation will be more complicated than discussed in this article.

Lastly, sample size procedures for precise interval estimation of normal percentiles, $\mu \pm 1.96\sigma$, were discussed (Shieh [13]). Researchers shall distinguish that the objective of CIs for $\mu \pm 1.96\sigma$ and the objective of LOA are different. The CIs are devised to capture unknown $\mu \pm 1.96\sigma$ which are fixed numbers (i.e., parameters), and the LOA is devised to capture future outcomes which are random numbers (i.e., random variables).

References

[1]   D. G. Altman and J. M. Bland, Measurement in medicine: The analysis of method comparison studies, The Statistician 32 (1983), 307-317.

[2]   J. M. Bland and D. G. Altman, Statistical method for assessing agreement between two methods of clinical measurement, Lancet 327 (1986), 307-310.

[3] W. G. Hopkins, Measures of reliability in sports medicine and science, Sports Medicine 30(1) (2000), 1-15.

DOI: https://doi.org/10.2165/00007256-200030010-00001

[4] D. Giavarina, Understanding Bland Altman analysis, Biochemia Medica 25(2) (2015), 141-151.

DOI: https://doi.org/10.11613/BM.2015.015

[5] P. J. Schluter, A multivariate hierarchical Bayesian approach to measuring agreement in repeated measurement method comparison studies, BMC Medical Research Methodology 9 (2009); Article 6.

DOI: https://doi.org/10.1186/1471-2288-9-6

[6] P. Taffe, Effective plots to assess bias and precision in method comparison studies, Statistical Methods in Medical Research 27(6) (2018), 1650-1660.

DOI: https://doi.org/10.1177/0962280216666667

[7] S. Geisser, Predictive Inference: An Introduction, Chapman & Hall, New York, 1993.

[8] W. G. Howe, Two-sided tolerance limits for normal populations-some improvements, Journal of the American Statistical Association 64(326) (1969), 610-620.

DOI: https://doi.org/10.1080/01621459.1969.10500999

[9] W. C. Guenther, Sampling Inspection in Statistical Quality Control, William C. Guenther and C. Griffin, London, 1977.

[10] S. B. Vardeman, What about the other intervals? The American Statistician 46(3) (1992), 193-197.

[11] G. J. Hahn, W. Q. Meeker and L. A. Escobar, Statistical Intervals: A Guide for Practitioners (2nd Edition), Wiley Publishing, 2014.

[12] M. J. Lu, W. H. Zhong, Y. X. Liu, H. Z. Miao, Y. C. Li and M. H. Ji, Sample size for assessing agreement between two method of measurement by Bland-Altman method, International Journal of Biostatistics 12(2) 2016.

DOI: https://doi.org/10.1515/ijb-2015-0039

[13] G. Shieh, The appropriateness of Bland-Altman's approximate confidence intervals for limits of agreement, BMC Medical Research Methodology 18 (2018); Article 45.

DOI: https://doi.org/10.1186/s12874-018-0505-y

[14] P. Taffe, Assessing bias, precision, and agreement in method comparison studies, Statistical Methods in Medical Research (2019) (In press).

[15] L. S. Nawarathna and P. K. Choudhary, A heteroscedastic measurement error model for method comparison data with replicate measurements, Statistics in Medicine 34(7) (2015), 1242-1258.

DOI: https://doi.org/10.1002/sim.6424

[16]   L. S. Nawarathna and P. K. Choudhary, Measuring agreement in method comparison studies with heteroscedastic measurements, Statistics in Medicine 32(29) (2013), 5156-5171.

DOI: https://doi.org/10.1002/sim.5955

[17]   B. Carstensen, Comparing methods of measurement: Extending the LoA by regression, Statistics in Medicine 29(3) (2010), 401-410.

DOI: https://doi.org/10.1002/sim.3769

∎

Appendix

```
### SAMPLE SIZE CALCULATION USING PREDICTION INTERVAL


n.sim = 100000 # # # SIMULATION SIZE

tau = 0.9 # # # TARGET COVERAGE PROBABILITY

alpha = 1 – tau

kappa = 0.9 # # # CONFIDENCE LEVEL

epsilon = 0.01 # # # ROOM FOR ERROR

n0 = 2 # # # LOWER END FOR SEARCH

n1 = 10000 # # # UPPER END FOR SEARCH


for ( i in 1:1000 ) {

    m = round ( 0.5 * ( n0 + n1 ) )

    gamma = qt( 1 – 0.5*alpha , m –1 ) * sqrt ( 1 + 1/m )

    set.seed(123); z.sim = rnorm (n.sim ) * sqrt (1/m)

    set.seed(123); w.sim = rchisq ( n.sim , m –1 )

    q.sim = sqrt ( w.sim / ( m – 1 ) )

    u.sim = z.sim + gamma * q.sim

    l.sim = z.sim – gamma * q.sim

    p.sim = pnorm (u.sim) – pnorm (l.sim)

    p.m = mean ( abs( p.sim – tau ) < epsilon )

    if ( p.m > kappa ) n1 = m
```

```
    if ( p.m < kappa ) n0 = m

    if ( abs( n1 – n0 ) == 1 ) break

    }

print (n1)


# # # SAMPLE SIZE CALCULATION USING TOLERANCE INTERVAL


n.sim = 100000 # # # SIMULATION SIZE

tau = 0.9 # # # TARGET COVERAGE PROBABILITY

alpha = 1 – tau

kappa = 0.95 # # # CONFIDENCE LEVEL

delta = 0.05

epsilon = 0.01 # # # ROOM FOR ERROR

n0 = 2 # # # LOWER END FOR SEARCH

n1 = 10000 # # # UPPER END FOR SEARCH


for ( i in 1:10000 ) {

    m = round ( 0.5 * ( n0 + n1 ) )

    temp1 = qnorm ( 1 – 0.5* alpha )

    temp2 = sqrt ( ( m^2 – 1 ) / ( m * qchisq ( 1 – kappa , m – 1 ) ) )

    gamma = temp1 * temp2

    set.seed(123); z.sim = rnorm (n.sim ) * sqrt (1/m)

    set.seed(123); w.sim = rchisq ( n.sim , m – 1 )
```

```
    q.sim = sqrt ( w.sim / ( m − 1 ) )

    u.sim = z.sim + gamma ∗ q.sim

    l.sim = z.sim − gamma ∗ q.sim

    p.sim = pnorm (u.sim) − pnorm (l.sim)

    p.m = mean ( tau < p.sim & p.sim < tau + 2 ∗ epsilon )

    if ( p.m > kappa − delta ) n1 = m
    if ( p.m < kappa − delta ) n0 = m
    if ( abs( n1 − n0 ) <= 2 ) break
    }
print (n1)
```