Meng-Jie Lu[1] / Wei-Hua Zhong[1] / Yu-Xiu Liu[1] / Hua-Zhang Miao[1] / Yong-Chang Li[1] / Mu-Huo Ji[2]

Sample Size for Assessing Agreement between Two Methods of Measurement by Bland–Altman Method

[1] Department of Medical Statistics, Jinling Hospital, Southern Medical University, 305 East Zhongshan Road, Nanjing 210002, China, E-mail: 898545969@qq.com, 1396689590@qq.com, liu_yuxiu@163.com, miaohuazhang@qq.com, lycnjzy@163.com
[2] Department of Anesthesiology, Jinling Hospital, Nanjing University, Nanjing 210002, China, E-mail: jimuhuo2009@sina.com

Abstract:
The Bland–Altman method has been widely used for assessing agreement between two methods of measurement. However, it remains unsolved about sample size estimation. We propose a new method of sample size estimation for Bland–Altman agreement assessment. According to the Bland–Altman method, the conclusion on agreement is made based on the width of the confidence interval for LOAs (limits of agreement) in comparison to predefined clinical agreement limit. Under the theory of statistical inference, the formulae of sample size estimation are derived, which depended on the pre-determined level of α, β, the mean and the standard deviation of differences between two measurements, and the predefined limits. With this new method, the sample sizes are calculated under different parameter settings which occur frequently in method comparison studies, and Monte-Carlo simulation is used to obtain the corresponding powers. The results of Monte-Carlo simulation showed that the achieved powers could coincide with the pre-determined level of powers, thus validating the correctness of the method. The method of sample size estimation can be applied in the Bland–Altman method to assess agreement between two methods of measurement.

Keywords: Bland-Altman method, limits of agreement, method comparison study, Monte-Carlo simulation, sample size estimation
DOI: 10.1515/ijb-2015-0039

1 Introduction

The original article by Bland and Altman [1] which proposed the method of agreement analysis has received more than 30,000 citations in the biomedical literature and has increased in usage in recent years. *Nature* asked Thomson Reuters, which now owns the SCI, to list the 100 most highly cited papers of all time. The third most frequently cited statistics paper (number 29) is a 1986 publication by British statisticians Martin Bland and Douglas Altman who proposed a technique – now known as the Bland–Altman plot [2]. It has been stressed that estimates of limits of agreement (LOAs) should be accompanied with confidence intervals. However, they found that "confidence intervals are seldom quoted" in reports of method comparison studies [3–5]. A reporting standards for Bland–Altman agreement analysis in laboratory research found that the CI limits of LOAs were reported in only 6 % of 50 studies published later than 2012 [6]. Many researchers forget that Bland and Altman presented the limits of agreement as a reference interval only. The LOAs do not guarantee coverage on the range of potential differences between the two measurements and cannot be used directly for statistical inference. Just as Hamilton and Stamey pointed out, the Bland–Altman limits of agreement by themselves provide only a reference interval and should never be used as the determining factor to conclude agreement between two methods [7]. They informed that future researchers should take this variability into account and always provide confidence intervals when using the LOAs approach. Like any other clinical trials, it is essential to determine the sample size for method comparison studies [8]. Although Bland has given the sample size for a study of agreement between two methods of measurement which were available from his website [9], the sample size was determined without considering the power of the statistical procedure and could not guarantee the power of test. Sample size calculations were performed in only 30 % of publications reviewed [6]. Bland and Altman hoped to find time to publish some of these, for example, on sample size estimation for measurement method agreement studies, but up to now it is still not solved satisfactorily [3].

In this paper we propose a method to calculate sample size for Bland–Altman method, and Monte-Carlo simulations are used to validate the correctness of the method. In Section 2, we introduce the assumptions

Yu-Xiu Liu is the corresponding author.

and theory of Bland–Altman method. Section 3 focuses on the derivation of the new method of sample size estimation for Bland–Altman method. In Section 4, Monte-Carlo simulation is used to obtain the corresponding powers. The results of Monte-Carlo simulation showed that the achieved powers could coincide with the pre-determined level of powers, thus validating the correctness of the method. In Section 5, we show a clinical worked example from a set of measured data of free prostate specific antigen (FPSA), which is often used to evaluate the presence of prostate cancer and other prostate disorders. We give concluding remarks in Section 6.

2 Method

2.1 Assumptions

Suppose that the measurements of two methods are made on each of n subjects drawn from some population of interest. Suppose further that the two measurements, x_i and y_i respectively and the difference, d_i for subject i (i=1,2, ..., n)

$$d_i = x_i - y_i$$

The important first step of the Bland–Altman method is to plot the data and to check its pattern and distribution. The differences for the two methods are plotted against their means and if the data are well-behaved, then construction of the various limits and interpretation of the data is simple and straightforward. The assumptions of the limits of the agreement method are that the differences values resulting from two measurements should have an approximately normal distribution, constant variance of the differences, and no proportional bias [10]. Proportional bias is present when the differences increase or decrease in proportion to the average values [11].

2 LOAs and confidence interval estimation

Suppose difference variable D is a random variable which follows a normal distribution with mean μ and variance σ^2. It is well-known that $100(1-\gamma)$ % of the population lies between $\mu \pm z_{1-\gamma/2}\sigma$. In practice, both μ and σ are unknown and have to be estimated. We take $\bar{D}$ and S_D^2 as estimators of μ and σ^2 respectively.

The $100(1-\gamma)$% LOAs can be calculated as

$$\bar{D} \pm z_{1-\gamma/2}S_D$$

where $z_{1-\gamma/2}$ is the cumulative $100(1-\gamma/2)$% percentile of a standard normal distribution, S_D is the standard deviation of the differences, $\bar{D} + z_{1-\gamma/2}S_D$ is the upper limit value of LOA, and $\bar{D} - z_{1-\gamma/2}S_D$ is the lower limit value. 95 % LOAs are the most common, which are mean minus 1.96 standard deviations and mean plus 1.96 standard deviations respectively. These limits are expected to contain 95 % of paired differences between measurements by the two methods.

It is important to consider confidence interval of LOAs [12]. The $100(1-\alpha)$% confidence interval estimation of $100(1-\gamma)$% LOAs derived by Bland and Altman can be calculated as [1]

$$Lower = \bar{D} - z_{1-\gamma/2}S_D - t_{1-\alpha/2,n-1}S_D\sqrt{\frac{1}{n} + \frac{z_{1-\gamma/2}^2}{2(n-1)}} \quad (1)$$

$$Upper = \bar{D} + z_{1-\gamma/2}S_D + t_{1-\alpha/2,n-1}S_D\sqrt{\frac{1}{n} + \frac{z_{1-\gamma/2}^2}{2(n-1)}} \quad (2)$$

We can obtain the upper confidence limit of upper limit of the LOAs and the lower confidence limit of lower limit of the LOAs, where n is the sample size. Generally, we set γ and α as 0.05. If the 95 % confidence interval for the 95 % LOAs comes within the pre-defined agreement limits which are clinically acceptable, the two methods agree sufficiently to fulfil the agreement requirements. In fact, the correspondence between confidence intervals for LOAs and hypotheses tests here is identical. Providing A is the lower limit and B is the upper limit of LOAs of population differences, we can construct the following simultaneous hypotheses: H_{01} is $A < -\delta$, H_{11} is $A \geq -\delta$ and H_{02} is $B > \delta$, H_{12} is $B \leq \delta$. When the two null hypotheses are rejected simultaneously, the two measurements would be inferred to agree. The hypotheses of Bland–Altman method are quite similar to the equivalence [13].

3 Sample size formulae

Considering the confidence interval estimation of LOAs has symmetry of μ and $-\mu$ ($\mu \geq 0$) and the sample size estimations of these two situations should be the same, we just discuss the situation when $\mu \geq 0$. According to the statistical inference principle of Bland–Altman limits of agreement, we can separate total type I error (α) into two parts which are both $\alpha/2$. Similarly, we can separate total type II error (β) into two parts. One is the first type II error (β_1) of the upper limit value of LOAs and the other is the second type II error (β_2) of the lower limit value of LOAs (Figure 1.) [13].

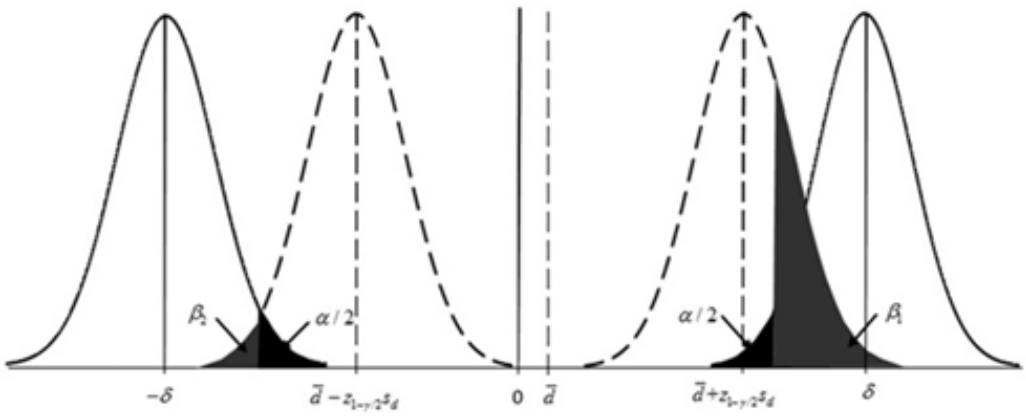

Figure 1:

We can get a direct estimate of β_1 and β_2:

$$\begin{aligned}
\overline{D} &+ z_{1-\gamma/2}S_D - z_{\beta_1}se_{LOAs} = \delta + z_{\alpha/2}se_{LOAs} \\
\overline{D} &- z_{1-\gamma/2}S_D + z_{\beta_2}se_{LOAs} = -\delta - z_{\alpha/2}se_{LOAs} \\
\beta_1 &= \Phi\left(\frac{\overline{D}+z_{1-\gamma/2}S_D-\delta}{se_{LOAs}} - z_{\alpha/2}\right) \\
\beta_2 &= \Phi\left(\frac{-\overline{D}+z_{1-\gamma/2}S_D-\delta}{se_{LOAs}} - z_{\alpha/2}\right)
\end{aligned}$$

where $\Phi(\bullet)$ is defined as the cumulative density function of standard normal distribution $N(0,1)$, se_{LOAs} is the standard error of the lower limit or upper limit of LOAs, $se_{LOAs} = S_D\sqrt{\frac{1}{n} + \frac{z^2_{1-\gamma/2}}{2(n-1)}}$, δ is the maximum allowable difference that can be accepted clinically, it needs to be defined in advance.

According to the statistical distribution theory, it is best to calculate the type II error (β) under the assumption of a non-central t-distribution [14], that is:

$$\beta_1 = 1 - probt\left[t_{1-\alpha/2,n-1}, n-1, \tau_1\right] \tag{3}$$

$$\beta_2 = 1 - probt\left[t_{1-\alpha/2,n-1}, n-1, \tau_2\right] \tag{4}$$

where $probt\left[\bullet, n-1, \tau_1\right]$ denotes the cumulative distribution function of a Student's non-central t-distribution with $n-1$ degrees of freedom and non-centrality parameter τ_1.

The non-centrality parameters τ_1 and τ_2 are non-central parameters defined as

$$\tau_1 = \frac{\delta - \bar{D} - z_{1-\gamma/2}S_D}{S_D\sqrt{\frac{1}{n} + \frac{z^2_{1-\gamma/2}}{2(n-1)}}}.$$

$$\tau_2 = \frac{\delta + \bar{D} - z_{1-\gamma/2}S_D}{S_D\sqrt{\frac{1}{n} + \frac{z^2_{1-\gamma/2}}{2(n-1)}}}.$$

We can get an estimate of the power:

$$\begin{aligned} power &= 1-\beta = 1-(\beta_1+\beta_2) \\ &= probt\left(t_{1-\alpha/2,n-1}, n-1, \tau_1\right) + probt\left(t_{1-\alpha/2,n-1}, n-1, \tau_2\right) \end{aligned} \tag{5}$$

When $\mu = 0$, the sample size calculation can be written as follows:

$$n = \frac{(2+z^2_{1-\gamma/2})[tinv(1-\beta/2, n-1, t_{1-\alpha/2,n-1})]^2 S_D^2}{2(z_{1-\gamma/2}S_D - \delta)^2} \tag{6}$$

where $tinv(1-\beta/2, n-1, t_{1-\alpha/2,n-1})$ is defined as the inverse of a Student's non-central t-distribution.

In eq. (6), $tinv(1-\beta/2, n-1, t_{1-\alpha/2,n-1})$ is related to sample size (n), we need to use iterative method to calculate sample size. Firstly we replace non-central t-distribution quantile with standard normal distribution quantile to obtain the initial value (n_0), and then iterate step-by-step until n reaches a stable value.

When $\mu > 0$, we firstly calculate by eq. (6) to obtain an initial value (n_1), then calculate by eq. (5) to achieve the power. If the estimated power is close enough to the pre-specified power then n is the sample size that we want to estimate. Otherwise we make n_1 equal to $n_1 + 1$ and judge whether the estimated power is close enough the pre-specified power. Repeat the procedure above until be closest to but greater than the pre-specified power. Table 1 summaries reasonable estimates of the sample size using eqs (5) and (6) for various standardized difference limits (μ/σ), different standardized agreement limits (δ/σ), and different type II error (β) assuming that the data are well-behaved.

Table 1 can be a reference for clinical researchers to estimate the sample size in the agreement assessment trial between two methods of measurement.

Table 1: Sample size (n) for Bland–Altman method with non-central t-distribution for different standardized difference limits (μ/σ), different standardized agreement limits (δ/σ), and different type II error (β). $(\alpha = 0.05)$.

δ/σ	μ/σ β	**0**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
2.0	0.2	19,152									
2.1	0.2	1,570	14,307								
2.2	0.2	538	1,174	14,307							
2.3	0.2	271	403	1,174	14,307						
2.4	0.2	164	206	402	1,174	14,307					
2.5	0.2	110	128	203	402	1,174	14,307				
2.6	0.2	80	89	123	203	402	1,174	14,307			
2.7	0.2	61	66	84	123	203	402	1,174	14,307		
2.8	0.2	49	51	61	83	123	203	402	1,174	14,307	
2.9	0.2	40	42	48	61	83	123	203	402	1,174	14,307
3.0	0.2	33	35	39	47	60	83	123	203	402	1,174
2.0	0.1	23,685									
2.1	0.1	1,941	19,152								
2.2	0.1	665	1,570	19,152							
2.3	0.1	334	538	1,570	19,152						
2.4	0.1	202	271	538	1,570	19,152					
2.5	0.1	136	166	271	538	1,570	19,152				
2.6	0.1	99	113	164	271	538	1,570	19,152			
2.7	0.1	75	83	110	164	271	538	1,570	19,152		
2.8	0.1	60	64	80	110	164	271	538	1,570	19,152	
2.9	0.1	49	52	62	80	110	164	271	538	1,570	19,152
3.0	0.1	41	43	49	61	80	110	164	271	538	1,570

4 Simulation

Monte-Carlo simulation studies with 10,000 replicates were performed to examine the validity and correctness of the proposed formulae for estimating sample size by calculating empirical powers. Simulation data were generated on the basis of normal distribution by considering typical situations under different parameter set-

tings. If the achieved power is very close to the pre-specified power, then it could be proved that our formulae can estimate the reasonable sample size.

As claimed above, simulating corresponding powers is easy: Firstly, we define the mean of differences (μ), the standard deviation of differences (σ) and the pre-defined clinical agreement limit (δ). And then we estimate the sample size with eqs (5) and (6), and calculate the 95 % confidence interval for the 95 % LOAs. If they lie within the pre-defined clinical agreement limits $(-\delta, \delta)$, then we draw a conclusion that the two methods agree, otherwise disagree. We repeat the procedure above 10,000 times and compute the times (t) that draw agreement conclusion. The value of $\frac{t}{10000} \times 100\%$ is the achieved power. Table 2 presents the achieved power for different parameter settings corresponding to the Table 1.

Table 2 indicates that the achieved powers are generally close to the pre-specified power of 80 % or 90 %. It shows that the formulae give reasonable estimates of the sample size using eqs (5) and (6) for various parameter settings.

Table 2: Power for Bland–Altman methods with non-central t-distribution for different standardized different limits (μ/σ), different standardized agreement limits (δ/σ), and different type II error (β). ($\alpha = 0.05$).

δ/σ	μ/σ β	**0**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
2.0	0.2	81.42									
2.1	0.2	81.33	79.34								
2.2	0.2	81.49	79.05	80.29							
2.3	0.2	81.37	79.97	79.50	79.42						
2.4	0.2	81.55	80.10	79.45	80.04	80.21					
2.5	0.2	80.42	80.59	81.20	79.88	79.40	80.03				
2.6	0.2	81.34	80.32	78.75	78.91	79.23	79.76	79.57			
2.7	0.2	81.63	80.76	80.72	80.23	80.00	79.78	80.41	79.73		
2.8	0.2	82.48	81.26	79.06	79.24	78.73	80.79	78.98	79.08	79.61	
2.9	0.2	82.82	82.30	80.36	79.20	79.37	79.81	79.62	79.86	79.89	79.35
3.0	0.2	82.38	82.12	81.16	79.71	78.83	79.69	79.58	79.66	79.46	80.23
2.0	0.1	90.28									
2.1	0.1	90.38	90.19								
2.2	0.1	89.90	89.19	90.06							
2.3	0.1	90.11	90.91	89.43	89.46						
2.4	0.1	89.99	89.27	89.28	89.58	89.84					
2.5	0.1	89.35	89.76	89.60	90.23	89.35	90.22				
2.6	0.1	88.70	89.20	88.94	89.47	90.91	90.02	89.05			
2.7	0.1	89.80	89.07	89.43	90.16	89.44	89.77	90.16	89.39		
2.8	0.1	89.83	90.63	90.44	88.68	88.99	89.05	88.92	89.72	89.69	
2.9	0.1	90.05	89.72	89.52	88.74	89.15	89.34	89.12	89.07	90.26	89.60
3.0	0.1	90.35	90.37	89.47	88.92	89.06	88.53	88.50	89.06	89.41	89.57

Bland has given the sample size for a study of agreement between two methods of measurement which were available from his website [9]. In the 1986 L*ancet* paper they gave a formula for the confidence interval for the 95 % limits of agreement. The standard error of the 95 % limit of agreement is approximately root $\left(\sqrt{3s^2/n}\right)$, where s is the standard deviation of the differences between measurements by the two methods and n is the sample size. The confidence interval is the estimate of the limit, $\bar{d}$ plus or minus $1.96s$, plus or minus 1.96 standard errors, then the sample size can be worked out.

We set $\alpha = 0.05$, $\beta = 0.20$, $\mu = -0.4\ 0.4$, $\sigma^2 = 1$, $\delta = 2.7$, and *pre-specified power* = 80 %. Figure 2 shows the sample sizes and powers of B-A method and new method under different parameter settings. With the Bland–Altman method, the sample size is calculated without considering the power of the statistical procedure, and so the probability of obtaining the required width is only 0.50. With the new method, the achieved power is generally close to the pre-specified power of 80 %.

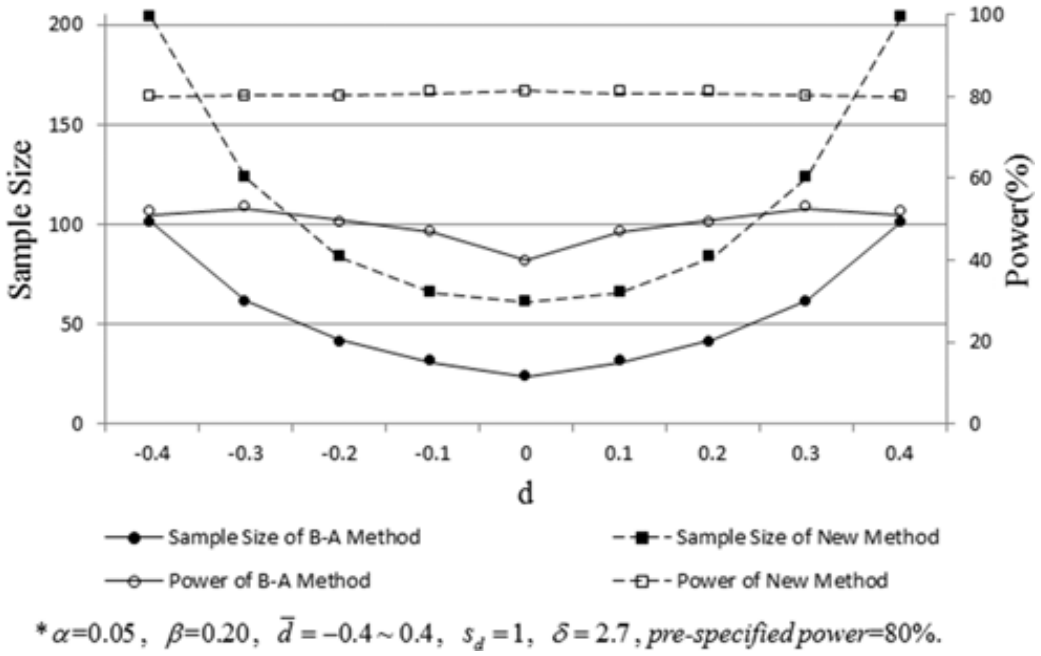

Figure 2:

The number of subjects required in the Bland–Altman method proposed by Bland is determined on the basis of the expected width of a confidence interval. It fails to explicitly consider the probability of achieving the desired interval width and may thus provide sample sizes that are too small to have enough power. However, the new method is more appropriate, because it can ensure an adequate probability of achieving the desired precision.

5 Example

We show a clinical worked example from a set of measured data of free prostate specific antigen (FPSA), which is often used to evaluate the presence of prostate cancer and other prostate disorders. AIA-1800 and I2000 methods were used to measure the FPSA. In the process of measurement, a same random sequence of sample was used in the two instruments [15]. Through a pre-experiment we get the mean and standard deviation of differences between AIA-1800 and I2000 methods are 0.001167 mmol/l and 0.001129 mmol/l respectively. Defining $\alpha = 0.05$, $\beta = 0.20$, $(-\delta, \delta) = (-0.004, 0.004)\mathrm{m}mol/l$, we can calculate that a sample size of 83 would be needed to provide 80 % power to assess agreement between two methods of measurement. Monte-Carlo simulation is used to obtain the corresponding power which is equal to 80.51 %, closely to the pre-defined power (80 %).

6 Conclusion

Based on the statistical inference principle and mathematical distribution theory, we have derived the calculating formula of sample size for Bland–Altman method under different parameter settings. For the sake of convenience, we have given a set of table which can be easily find out the sample size for different standardized difference limits (μ/σ), standardized agreement limits (δ/σ), and type error (β) under two-sided $\alpha = 0.05$. Both α and β should be considered to have sample size large enough to ensure that the half width of a $100(1-\alpha)\%$ confidence interval is no larger than a pre-specified width with a pre-specified assurance probability $100(1-\beta)\%$. We carried out Monte-Caro simulation studies to validate the correctness of the proposed method. The simulation results reveal that the achieved powers could coincide with the pre-determined level of powers, thus validating the correctness of the formulae.

It is important to be aware of the pre-specified clinically acceptable agreement limits. As with equivalence or non-inferiority clinical trials, the clinical agreement limits need to be determined in advance by clinical researchers and biostatistician. Defining these agreement limits may be a difficult aspect in designing the measurement comparison studies, because they depend upon not only the clinical scenario but also other variables.

Nevertheless, an attempt must be made to define them; a Delphi survey (opinion from experts) may be used to design the study. This survey is a group facilitation technique, which is an iterative multistage process designed to transform an opinion into group consensus [16].

Lin et al. [17] had discussed some issues about sample size using the tolerance interval; however, there are some deficiencies. First of all, in Lin's study, the hypothesis of the sample size calculating method is provided just under $\mu = 0$, which is not considered for $\mu \neq 0$. In fact, the two measurements may not be perfectly consistent $(\mu \neq 0)$, but we still believe their consistency as the population difference within a certain acceptable range (δ). In Lin's paper, it can be seen that the simulated power is consistently less than the pre-specified power for all design specifications, so the sample sizes are much under-estimated.

Hahn and Meeker [18] defined a tolerance interval that is an interval that one can claim to contain at least a specified proportion, p, of the population with a specified degree of confidence, $100(1–\alpha)\%$ and provided the sample size estimation about the tolerance interval. We can see that their sample size estimation is based on the desired precision without considering type II error (β) or power, and just addresses the frequently asked question "How large a sample do I need to obtain a confidence interval?" Although our confidence interval of LOAs is similar to their tolerance interval, the theories and the procedures of sample size estimation are totally different. Our method of sample size estimation is derived not only on the pre-determined level of α but also on the β.

Recently, studies of the agreement between two instruments or clinical tests have proliferated in ophthalmic literature. McAlinden et al. used a method of sample size calculation for agreement studies on the basis of method proposed by Bland [19]. The sample size was calculated without considering the power of the statistical procedure, so the probability of obtaining the required width was only 0.50 [20]. During the study design stage, considering the power in sample size calculations could lead to expected conclusions under the predetermined power level. Cesana et al. provided another sample size estimation required for demonstrating a Pearson correlation coefficient between the differences and the means of the measures [20], and we think this method is unreasonable. Actually, the correlation coefficient given by Cesana reflected the proportional bias. As we know, one of the assumptions about the application of the Bland–Altman method is no proportional bias. Without fulfillment of the assumption, this method would not be applicable.

There are some limitations to this study. Our sample size formulae are just appropriate for the data which are well-behaved. If the data behavior is not very well, such as non-normality or non-constant variance of the differences (heteroscedasticity) and proportional bias, the formulae are not suitable to solve the problem of estimating sample size.

Acknowledgement

We are grateful for the constructive comments of Dr. Jian-Jun Yang. We also thank the editors and the anonymous reviewers for valuable comments that have helped us significantly improved our manuscript.

Funding

This study was funded by a grant from the National Natural Science Foundation of China (No.81473066).

References

[1] Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;i:307–310.
[2] Noorden RV, Mahen B, Nuzzo R. The top 100 papers. Nature. 2014;514:550–553.
[3] Bland JM, Altman DG. Agreed statistics measurement method comparison. Anesthesiology. 2012;116:182–185.
[4] Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res. 1999;8:135–160.
[5] Bland JM, Altman DG. Appling the right statistics: analyses of measurement studies. Ultrasound Obst Gyn. 2003;22:85–93.
[6] Chhapola V, Kanwal SK, Brar R. Reporting standards for Bland–Altman agreement analysis in laboratory research: a cross-sectional survey of current practice. Ann Clin Biochem. 2015;52:382–386.
[7] Hamilton C, Stamey J. Using Bland—Altman to assess agreement between two medical devices-don't forget the confidence intervals!. J Clin Monit Comput. 2007;21:331–333.
[8] Bella ML, Teixeira-Pintoc A, McKenzied JE, Oliviere J. A myriad of methods: calculated sample size for two proportions was dependent on the choice of sample size formula and software. J Clin Epidemiol. 2014;67:601–605.

[9] Bland JM. How can I decide the sample size for a study of agreement between two methods of measurement? Available at: http://www-users.york.ac.uk/ mb55/meas/sizemeth.htmAccessed: 15 Aug 2015.
[10] Woodman RJ. Bland—Altman beyond the basics: creating confidence with badly behaved data. Clin Exp Pharmacol Physiol. 2010;37:141–142.
[11] Ludbrook J. Confidence in Altman-Bland plots: a critical review of the method of differences. Clin Exp Pharmacol Physiol. 2010;37:143–149.
[12] Stockl D, Cabaleiro DR, Uytfanghe KV, Thienpont LM. Interpreting method comparison studies by use of the Bland—Altman plot: reflecting the importance of sample size by incorporating confidence limits and predefined error limits in the graphic. Clin Chem. 2004;50:2216–2218.
[13] Julious SA. Sample size for clinical trials with Normal data. Stat Med. 2004;23:1921–1986.
[14] Forbes C, Evans M, Hastings N, Peacock B. Statistical distributions, 4th ed Hoboken: John Wiley & Sons, 2011:187–188.
[15] Zhou YH, Zang JJ, Wu MJ, Xu JF, He J. Allowable total error and limits for erroneous results (ATE/LER) zones for agreement measurement. J Clin Lab Anal. 2011;25:83–89.
[16] Hasson F, Keeney S, Mckennna H. Research guidelines for the Delphi survey technique. J Adv Nurs. 2000;32:1008–1015.
[17] Lin SC, Whipple DM, Ho CS. Evaluation of statistical equivalence using limits of agreement and associated sample size calculation. Commun Stat Theor Methods. 1998;27:1419–1432.
[18] Hahn GJ, Meeker WQ. Statistical intervals – a guide for practitioners. New York: John Wiley & Sons, 1991:150–167.
[19] McAlinden C, Khadka J, Pesudovs K. Statistical methods for conducting agreement (comparison of clinical tests) and precision (repeatability or reproducibility) studies in optometry and ophthalmology. Ophthalmic Physiol Opt. 2011;31:330–338.
[20] Cesana BM, Antonelli P. Agreement analysis: further statistical insights. Ophthalmic Physiol Opt. 2012;32:436–440.