Testing the agreement of medical instruments: Overestimation of bias in the Bland–Altman analysis

Rafdzah Zaki a,⁎, Awang Bulgiba a, Noor Azina Ismail b

a Julius Centre University of Malaya, Department of Social & Preventive Medicine, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, Malaysia
b Department of Applied Statistics, Faculty of Economics & Administration, University of Malaya, 50603, Kuala Lumpur, Malaysia

A R T I C L E   I N F O

Keywords:
Validation studies
Statistical data analysis
Methodology
Bland–Altman method
Limits of agreement

A B S T R A C T

Objectives. The Bland–Altman method is the most popular method used to assess the agreement of medical instruments. The main concern about this method is the presence of proportional bias. The slope of the regression line fitted to the Bland–Altman plot should be tested to exclude proportional bias. The aim of this study was to determine whether the overestimation of bias in the Bland–Altman analysis is still present even when the proportional bias has been excluded.

Methods. Data were collected from participants attending a workplace health screening program in a public university in Malaysia between 2009 and 2010. Variables collected were blood glucose level, body weight and systolic blood pressure (n=300 per variable). Readings from the original clinical dataset were compared with twenty randomly generated datasets for each variable. The Bland–Altman limits of agreement was used to determine the agreement. The presence of proportional bias was excluded for all datasets using the recommended method.

Results. The range of predicted bias was higher than the simulated bias for all datasets. The overestimation of bias increased as the range of actual bias increased.

Conclusion. Testing the slope of regression line of the Bland–Altman plot does not remove the artifactual bias in the prediction.

© 2013 Elsevier Inc. All rights reserved.

Introduction

Many important variables measured in medicine are continuous in nature such as blood pressure and oxygen levels. In any clinical situation, either at the stage of health screening, diagnosing cases, or making prognosis, accurate measurement of clinical variables is vital. Numerous new techniques have been developed with the aim of finding a cheaper and safer method to test patients. It is important to be sure that the new method of measurement is in agreement with the current or gold standard method. Agreement signifies the accuracy of that certain instrument (de Vet, 1998).

The Bland–Altman method is the most popular method used to assess the agreement of medical instruments measuring continuous variables (Zaki et al., 2012). Bland and Altman advocated the use of a graphical method (the Bland–Altman plot) and the limits of agreement (LoA) (Bland and Altman, 1987, 1995). Despite its popularity, Hopkins (2004) demonstrated that a proportional bias is produced in a Bland–Altman plot. The main concern about the proportional bias is that this will result in overestimation of prediction. The predicted bias will consist of artifactual and real bias, which cannot be differentiated by the researcher (Hopkins, 2004). Thus, the Bland–Altman plot will indicate that bias is present even when there is none. Ludbrook (2010) recommended that a linear regression line be fitted to the Bland–Altman plot (Ludbrook, 2010) to check for this bias. It was argued that, if the slope of the regression line fitted to the Bland–Altman plot is not significantly different from zero then the proportional bias is absent (Ludbrook, 2010). Thus we should not be worried about any artifactual bias.

The aim of this study was to determine whether an overestimation of bias in the Bland–Altman analysis will still exist even when the proportional bias is excluded using the linear regression analysis of the Bland–Altman plot.

Method

Data collection

Data were collected from participants attending a workplace health screening program in a large public university. This study was approved by the ethical committee of the university (MEC no: 715.23). A convenient sampling method was applied and written informed consent was obtained from all participants. A total of 300 participants (per variable) consented to be part of this study. Variables collected were blood glucose level, body weight, and systolic blood pressure (SBP). Bland (2004) recommended a minimum sample size of 100 for
the Bland–Altman analysis. Therefore, a sample size of 300 was deemed sufficient for this study.

All blood taking and analysis for the blood glucose laboratory test were performed as part of the workplace health screening program. Blood for glucose was tested in the university’s teaching hospital laboratory. Consent for blood withdrawal (venepuncture) was also done as part of the screening program. However, written consent to access the blood results was still obtained from the participant and the coordinator of the screening program. The digital weighing scale (Seca 813 Robusta High Capacity Digital Floor Scale) with a maximum capacity of 200 kg and which displays the weight in kilogram to two decimal places was used to collect measurements for body weight. The measurement of SBP was done using the Omron HEM 907XL IntelliSense Professional Digital Blood Pressure Monitor.

Data analysis

The original datasets of each variable were compared with generated comparison datasets. Twenty datasets with random distribution of specific range of bias were generated per variable using Matlab Software version 2009a. Bias generated included positive bias and a mixture of positive and negative bias in a single dataset. The Bland–Altman analysis was performed for all the simulated datasets, for each variable. The presence of proportional bias was tested by testing the slope of regression line fitted to the Bland–Altman plot. Proportional bias was excluded when the slope did not significantly differ from zero.

The ranges of predicted error were compared with the ranges of simulated (i.e. actual) error. The ranges of predicted error were determined based on the ranges between the upper and lower LoA. The ranges of prediction based on the confidence interval (CI) of the LoA were also taken into consideration (upper CI of upper LoA minus lower CI of lower LoA). Analyses were performed using the MedCalc Software version 12.1.3.

Result

Three variables (blood glucose level, body weight, and SBP) with sample sizes of 300 per variable were collected. The majority of participants were Malays (85.3%), followed by Indians (6.3%), Chinese (5.0%), and other races (3.3%). The mean age of participants was 43.3 years, and 62.6% of the participants were male.

Twenty datasets with simulated bias were generated for each variable. The scatters of differences were uniform (homoscedasticity), and the majority of participants were 43.3 years, and 62.6% of the participants were male.

Twenty datasets with simulated bias were generated for each variable. The slopes of the regression lines of the Bland–Altman analysis. Therefore, a sample size of 300 was deemed sufficient for this study.

Table 1

<table>
<thead>
<tr>
<th>Simulated random error</th>
<th>Lower limit of agreement (CI)</th>
<th>Upper limit of agreement (CI)</th>
<th>Range of simulated error</th>
<th>Range of predicted error (LoA only)</th>
<th>Range of predicted error (LoA with CIs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Error 0 to 0.1</td>
<td>−0.01 (−0.01 to 0.01)</td>
<td>0.11 (0.10 to 0.11)</td>
<td>0.1</td>
<td>0.12</td>
<td>0.12</td>
</tr>
<tr>
<td>2. Error 0 to 0.2</td>
<td>−0.01 (−0.02 to 0.01)</td>
<td>0.22 (0.20 to 0.23)</td>
<td>0.2</td>
<td>0.23</td>
<td>0.25</td>
</tr>
<tr>
<td>3. Error 0 to 0.3</td>
<td>−0.02 (−0.04 to −0.01)</td>
<td>0.32 (0.30 to 0.33)</td>
<td>0.3</td>
<td>0.34</td>
<td>0.37</td>
</tr>
<tr>
<td>4. Error 0 to 0.4</td>
<td>−0.04 (−0.06 to −0.01)</td>
<td>0.43 (0.41 to 0.46)</td>
<td>0.4</td>
<td>0.47</td>
<td>0.52</td>
</tr>
<tr>
<td>5. Error 0 to 0.5</td>
<td>−0.05 (−0.06 to −0.01)</td>
<td>0.52 (0.49 to 0.54)</td>
<td>0.5</td>
<td>0.55</td>
<td>0.6</td>
</tr>
<tr>
<td>6. Error 0 to 0.6</td>
<td>−0.05 (−0.05 to 0.01)</td>
<td>0.61 (0.58 to 0.64)</td>
<td>0.6</td>
<td>0.63</td>
<td>0.69</td>
</tr>
<tr>
<td>7. Error 0 to 0.7</td>
<td>−0.04 (−0.08 to 0)</td>
<td>0.75 (0.71 to 0.79)</td>
<td>0.7</td>
<td>0.79</td>
<td>0.87</td>
</tr>
<tr>
<td>8. Error 0 to 0.8</td>
<td>−0.05 (−0.09 to 0)</td>
<td>0.88 (0.84 to 0.93)</td>
<td>0.8</td>
<td>0.93</td>
<td>1.02</td>
</tr>
<tr>
<td>9. Error 0 to 0.9</td>
<td>−0.05 (−0.10 to 0)</td>
<td>0.96 (0.91 to 1.01)</td>
<td>0.9</td>
<td>1.01</td>
<td>1.1</td>
</tr>
<tr>
<td>10. Error 0 to 1.0</td>
<td>−0.06 (−0.12 to −0.01)</td>
<td>1.06 (1.01 to 1.12)</td>
<td>1.0</td>
<td>1.12</td>
<td>1.24</td>
</tr>
<tr>
<td>11. Error −0.1 to 0.1</td>
<td>−0.11 (−0.12 to −0.10)</td>
<td>1.12 (1.11 to 1.13)</td>
<td>1.2</td>
<td>0.23</td>
<td>0.25</td>
</tr>
<tr>
<td>12. Error −0.2 to 0.2</td>
<td>−0.22 (−0.24 to −0.20)</td>
<td>0.22 (0.20 to 0.24)</td>
<td>0.4</td>
<td>0.44</td>
<td>0.48</td>
</tr>
<tr>
<td>13. Error −0.3 to 0.3</td>
<td>−0.34 (−0.37 to −0.30)</td>
<td>0.33 (0.30 to 0.37)</td>
<td>0.6</td>
<td>0.67</td>
<td>0.74</td>
</tr>
<tr>
<td>14. Error −0.4 to 0.4</td>
<td>−0.49 (−0.53 to −0.44)</td>
<td>0.43 (0.38 to 0.47)</td>
<td>0.8</td>
<td>0.92</td>
<td>1.00</td>
</tr>
<tr>
<td>15. Error −0.5 to 0.5</td>
<td>−0.58 (−0.64 to −0.52)</td>
<td>0.58 (0.53 to 0.64)</td>
<td>1.0</td>
<td>1.16</td>
<td>1.28</td>
</tr>
<tr>
<td>16. Error −0.6 to 0.6</td>
<td>−0.61 (−0.67 to −0.54)</td>
<td>0.69 (0.62 to 0.75)</td>
<td>1.2</td>
<td>1.30</td>
<td>1.42</td>
</tr>
<tr>
<td>17. Error −0.7 to 0.7</td>
<td>−0.80 (−0.88 to −0.73)</td>
<td>0.74 (0.66 to 0.81)</td>
<td>1.4</td>
<td>1.54</td>
<td>1.69</td>
</tr>
<tr>
<td>18. Error −0.8 to 0.8</td>
<td>−0.92 (−1.01 to −0.83)</td>
<td>0.90 (0.81 to 0.99)</td>
<td>1.6</td>
<td>1.82</td>
<td>2.00</td>
</tr>
<tr>
<td>19. Error −0.9 to 0.9</td>
<td>−1.06 (−1.16 to −0.95)</td>
<td>1.05 (0.94 to 1.15)</td>
<td>1.8</td>
<td>2.11</td>
<td>2.31</td>
</tr>
<tr>
<td>20. Error −1.0 to 1.0</td>
<td>−1.06 (−1.16 to −0.94)</td>
<td>1.14 (1.03 to 1.24)</td>
<td>2.0</td>
<td>2.20</td>
<td>2.40</td>
</tr>
</tbody>
</table>

Discussion

One of the critiques of the Bland–Altman analysis is the presence of proportional bias. To check whether proportional bias is present, Ludbrook recommended that a linear regression line be fitted to the Bland–Altman plot (Ludbrook, 2010).

Our study shows that there is an overestimation in the prediction of bias in the Bland–Altman analysis even when proportional bias had been shown to be absent by testing the slope of the regression line in the Bland–Altman plot. The overestimation of bias increases when the range of actual bias increases. We found a similar pattern for all three variables which means this is not an isolated issue. The regression line analysis of the Bland–Altman plot cannot be used to exclude the proportional bias and does not remove the possibility of artifactual bias in the prediction of bias.

Although we detected an overestimation of bias in the Bland–Altman analysis, this does not suggest that this method should be abandoned. Almost all statistical methods used in method comparison studies have been criticized. The Bland–Altman method should be used with caution and should be complemented by other methods. Usage of multiple methods has the advantage of compensating for the limitations of any one single method (Laz and Szolko, 2005).

Since the Bland–Altman method is the most commonly used method to assess agreement, the issue on overestimation bias or artifactual bias in the analysis should be highlighted to medical researchers. Any discrepancy in the analysis may lead to the application of inaccurate instruments in clinical practice, thus jeopardizing the quality of care given to the patient.

The analysis in this study was limited to the three collected variables (blood glucose level, body weight and SBP), and a range of simulated bias. Further analysis with various range of variables and range of bias is required to determine the impact of this overestimation of bias on clinical practice.
Conclusion

There is an overestimation in the prediction of bias in the Bland–Altman analysis even when the proportional bias was excluded. Testing the slope of the regression line in the Bland–Altman plot does not remove the artifactual bias in the prediction. The Bland–Altman method should be used with caution and should be complemented by other methods in such cases.

Conflict of interest statement
The authors declare that there are no conflicts of interest.

Acknowledgments

This work is part of the STeMM Programme supported by the University of Malaya/Ministry of Higher Education (UM/MOHE) High Impact Research Grant (Grant number E000010-20001) and the University of Malaya student research grant (PS162/2009B).

References


Fig. 1. Relationship between the simulated and predicted error in the Bland–Altman analysis for (a) systolic BP (b) body weight, and (c) blood glucose level. Data were collected from participants attending a workplace health screening program in a public university in Malaysia between 2009 and 2010.