

SEÇÃO DE BIOESTATÍSTICA

ANÁLISE DE CONCORDÂNCIA ENTRE MÉTODOS DE BLAND-ALTMAN

BLAND-ALTMAN ANALYSIS OF AGREEMENT BETWEEN METHODS

Vânia Naomi Hirakata¹, Suzi Alves Camey²

RESUMO

Quando se pretende avaliar a concordância entre dois métodos que deveriam medir a mesma quantidade, são utilizadas análises que nem sempre estão corretas. É importante que seja evitado o uso da correlação nestas situações e que a metodologia seja utilizada de forma adequada, incluindo os limites de concordância e seus intervalos de confiança, além de comentar se os limites encontrados são diferenças aceitáveis do ponto de vista clínico. A proposta do presente artigo é apresentar um método bastante simples que já é utilizado há bastante tempo, que é a análise de concordância entre métodos de Bland-Altman, salientando alguns problemas detectados na sua utilização. Para isto, foi feita uma simulação de três diferentes situações, com comentários e soluções. O programa R, por ser livre e ter incorporado comandos para a análise de Bland-Altman, foi utilizado para a análise dos resultados.

Palavras-chave: *Bland-Altman; metodologia; concordância*

ABSTRACT

When the intention is to evaluate the agreement between two methods that would measure the same quantity, the analyses used are not always correct. It is important to avoid the use of correlation in these situations and to properly use the methodology, including the analysis of limits of agreement and confidence intervals, and to comment on whether the limits are acceptable differences from a clinical point of view. The purpose of this paper is to present a simple method which has been already in use for some time, the Bland-Altman analysis of agreement between methods, pointing out some problems encountered with its use. For this, a simulation of three different situations was performed, with comments and solutions. The R program was used to analyze the results, because it's free and has available commands of the Bland-Altman analysis.

Keywords: *Bland-Altman; methodology; agreement*

Rev HCPA 2009;29(3):261-268

Um tipo de problema bastante comum que motiva pesquisas na área biomédica é o estudo de diferentes formas de avaliar alguma medida clínica que se apresenta de forma quantitativa, como por exemplo, a pressão arterial, a taxa de filtração glomerular ou a idade gestacional. O que se busca são métodos menos invasivos, com menos efeitos adversos, mais simples ou mais baratos do que os considerados "ideais". O objetivo é descobrir se as duas medidas são equivalentes e se uma poderia substituir a outra, ou seja, se existe ou não concordância. O problema é que não existe um único teste estatístico que avalie a variabilidade das diferenças entre os valores obtidos pelos dois métodos (erro) e a própria diferença entre os valores (viés) ao mesmo tempo (1). Os dois métodos deveriam apresentar um viés aproximadamente igual a zero e um erro que não tivesse um impacto clínico significativo para que fossem substituíveis.

Bland e Altman (1) publicaram em 1983 o primeiro artigo com uma metodologia alternativa ao cálculo do coeficiente de correlação, metodologia empregada até então. O coeficiente de correlação não avalia concordância e sim associação, coisas bem diferentes. Por exemplo, se um aparelho fornecesse os valores 80, 95, 96 e 100 mmHg para as medidas de pressão arterial sistólica de quatro indivíduos e outro aparelho fornecesse (para os mesmos indivíduos) 90,

105, 106 e 110 mmHg, respectivamente, a correlação seria igual a 1 ($r=1$, perfeita!), enquanto a discordância entre os valores é óbvia. A metodologia proposta inicialmente por Bland e Altman para avaliar a concordância entre duas variáveis (X e Y) parte de uma visualização gráfica a partir de um gráfico de dispersão entre a diferença das duas variáveis (X - Y) e a média das duas (X + Y)/2. Neste gráfico é possível visualizar o viés (o quanto as diferenças se afastam do valor zero), o erro (a dispersão dos pontos das diferenças ao redor da média), além de *outliers* e tendências. Logo em seguida, por sugestão de colegas da área clínica, escreveram um artigo em uma linguagem mais apropriada para leitores não estatísticos (2). Em 1995, por perceberem que, mesmo após uma publicação voltada ao público de interesse (clínicos), em uma revista vista por eles e utilizando uma linguagem simples, o coeficiente de correlação continuava a ser utilizado, e apesar de terem alcançado um bom impacto, tentaram uma abordagem em uma revista voltada a epidemiologistas (3). Finalmente, em 1999, publicaram o mais completo e longo artigo explicando detalhadamente o método, enfatizando que a repetibilidade de cada medida também deve ser avaliada e detalhando uma forma de cálculo dos limites de concordância quando existirem replicações de cada método (4). Atualmente, o método encontra-se bem divulgado, exemplo disso

1. Unidade de Bioestatística, Grupo de Pesquisa e Pós-Graduação, Hospital de Clínicas de Porto Alegre.

2. Departamento de Estatística, Instituto de Matemática, Programa de Pós-Graduação em Epidemiologia, UFRGS.

Contato: Vânia Hirakata. E-mail: vania.hirakata@gmail.com (Porto Alegre, RS, Brasil).

é que numa busca simples no Pubmed com o termo "Bland-Altman" no título, foram encontrados 10 artigos em 2008 e 2009, em revistas de diversas áreas. Além disso, a metodologia já encontra-se implementada em diversos pacotes estatísticos. No R, um software livre específico de análise estatística, existem rotinas prontas desenvolvidas por usuários que fornecem os valores e o gráfico com os limites de concordância. No Medcalc e Analyse-it, softwares comerciais, também são possíveis a visualização gráfica com o viés e os intervalos de concordância com apenas um comando.

A proposta do presente artigo é mostrar a utilização do método num exemplo aplicado, simulando três diferentes situações, bem como apresentar formas de se realizar estas análises no pacote R.

MÉTODOS

Inicialmente, a sugestão é que se faça um gráfico de dispersão entre as duas medidas, sem a reta dos mínimos quadrados, mas sim com a reta da diagonal de concordância. O próximo passo seria fazer o gráfico de Bland-Altman, um gráfico de dispersão relacionando as médias dos dois métodos $(X + Y)/2$, no eixo do X, com o viés (diferença entre eles), $X - Y$, no eixo do Y. Este gráfico permite avaliar a relação das discordâncias com as medidas avaliadas.

Deve ser avaliado se as diferenças entre as variáveis dependem ou não do tamanho da medida. Isto pode ser feito através de uma correlação entre as diferenças e as médias, que deve ser nula. A hipótese do viés ser ou não igual a zero pode ser testada por um teste t para amostras emparelhadas. A partir do cálculo do viés (\bar{d}) e do seu desvio-padrão (s_d) é possível chegar aos limites de concordância: $\bar{d} \pm 1,96s_d$, que devem ser calculados e incluídos no gráfico. Se o viés apresenta distribuição normal, estes limites representam a região em que se encontram 95% das diferenças nos casos estudados. Nas situações em que o viés não apresenta distribuição normal, é recomendada uma abordagem não-paramétrica (4).

No caso de haver relação entre os valores das diferenças e das médias, existem diversas sugestões de tratamento, todas para tentar homogeneizar a variação das diferenças em relação aos tamanhos das medidas. Uma sugestão é aplicar o logaritmo nos valores originais e em

tão aplicar o método. Outras, talvez mais simples, são de utilizar a razão entre X e Y (X/Y) no lugar das diferenças (4) ou o percentual das diferenças (5) e avaliar a relação entre estes valores e as médias.

Quanto à precisão dos valores de viés e dos limites de concordância, podemos calcular o erro-padrão e os intervalos de confiança, uma vez que o viés apresenta distribuição normal.

Para o viés (\bar{d}), o erro padrão (EP_d) é igual a s_d/\sqrt{n} , onde n é o tamanho da amostra, portanto o intervalo de confiança pode ser calculado assim: $\bar{d} \pm t \times EP_d$, com t sendo o valor tabelado da distribuição t para $n-1$ graus de liberdade. O erro padrão para os limites de concordância ($\bar{d} \pm 1,96 \times s_d$) pode ser obtido por $1,71 \times EP_d$, portanto o limite inferior teria seu intervalo de confiança calculado por $(\bar{d} - 1,96 \times s_d) \pm t \times 1,71 \times EP_d$ e o limite superior por $(\bar{d} + 1,96 \times s_d) \pm t \times 1,71 \times EP_d$ (4).

No cálculo do tamanho da amostra para empregar-se a análise de concordância, Bland, em sua página pessoal (6), sugere que seja calculado considerando a precisão do viés e dos limites de concordância, pois, como a decisão clínica deve ser baseada nos limites de concordância, é importante que os mesmos sejam precisos. De uma forma geral, ele mostra que, com uma amostra de 100 é possível estimar os valores (viés e limites) com um intervalo de confiança de 95% aproximado de $\pm 0,34s_d$ e, com 200, um intervalo de $\pm 0,24s_d$.

Bland-Altman no R

O R (7) é um programa livre que pode ser obtido em <http://www.r-project.org/>. Para informações sobre como instalar o R e suas bibliotecas veja os seguintes materiais:

- Curso sobre R aplicado a Epidemiologia: [Curso R Epi2008](#)
- Excelente tutorial sobre R: [R tutorial](#)

Após instalar e executar o R clique em >Arquivo>Novo Script. Isso fará abrir uma janela em branco onde deverão ser escritos os comandos (sintaxe) a serem executados pelo R (Imagem 1).

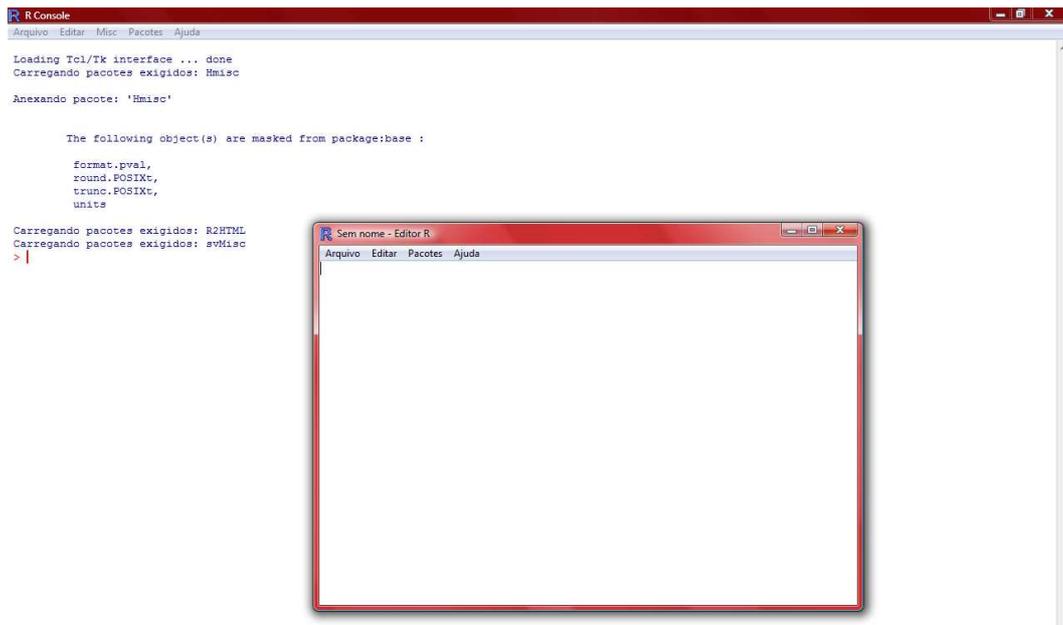


Imagem 1 - Console do R após iniciá-lo e janela de sintaxe.

Na hora de escrever a sintaxe temos que lembrar que o R diferencia letras maiúsculas de minúsculas. No mesmo diretório onde está armazenado o conjunto de dados deve ser colocado o arquivo [bland_altman.r](#) (clique no link para acessá-lo). Na janela da sintaxe colar os comandos que estão no Quadro 1. As linhas que começam com o símbolo # são apenas comen-

tários sobre o comando que está na linha abaixo. Para executar os comandos basta colocar o cursor na linha que se deseja executar, ou selecionar um bloco de comandos, e digitar CTRL+R. Os gráficos gerados na análise podem ser copiados e colados em documentos do Word ou PowerPoint.

Quadro 1 - Comandos necessários para fazer as análises e os gráficos.

```
#comando para carregar a biblioteca que permite ao R ler um arquivo do SPSS
library(foreign)
#comando para especificar o caminho onde se encontra o arquivo do SPSS. No exemplo a seguir o
arquivo encontra-se na pasta Meus Documentos
setwd("C:\\Meus Documentos")
# comando para ler o arquivo do SPSS, considerando os missing's. Neste exemplo o nome do ar-
quivo é nome_do_arquivo.sav
dados<-read.spss("nome_do_arquivo.sav",to.data.frame = T, use.missings = T)
# comando para exibir os nomes das variáveis do SPSS
names(dados)
# comando para carregar o banco de dados
attach(dados)
# carrega a função para fazer o gráfico do Bland-Altman
source("bland_altman.r")
# os nomes das variáveis utilizadas nesses comandos são explicadas no exemplo a seguir.
# define n como tamanho da amostra, no nosso exemplo 50
n <- 50
# comando para realizar o teste t para amostras pareadas entre as variáveis dum e eco_1
t.test(dum, eco_1, paired = T)
# comando para calcular o coeficiente de correlação entre as variáveis dum e eco_1
correl <- cor.test(dum, eco_1)
correl
# as 3 linhas a seguir servem para fazer o diagrama de dispersão, com o coeficiente de corre-
lação
plot(dum,eco_1,xlim=c(min(eco_1,dum),max(eco_1,dum)),ylim=c(min(eco_1,dum),max(eco_1,dum)))
abline(coef=c(0,1))
text(max(eco_1,dum),min(eco_1,dum),pos=2,paste("r=",round(correl$estimate[[1]],2),"",
IC95%=["",round(correl$conf.int[[1]],2),"",round(correl$conf.int[[2]],2),"",sep=""))
# comando para abrir uma nova janela de gráfico
windows()
#comando para fazer o gráfico de Bland-Altman das variáveis dum e eco_1
B_A("dum", "eco_1")
```

Exemplo

Para ilustrar o método vamos utilizar um exemplo hipotético onde o tempo de gestação de 50 mulheres foi estimado através da data da última menstruação (DUM) e pelo ultrassom.

A pergunta natural é: Os métodos concordam?

Aqui como em muitos casos não adianta termos uma boa correlação entre as medidas, é necessário que haja concordância entre elas.

Com objetivo de mostrar as diferentes conclusões que podem ser obtidas através da metodologia de Bland-Altman, foram simuladas 3 situações diferentes:

- Medidas correlacionadas e concordantes;
- Medidas correlacionadas e não-concordantes com viés constante;

- Medidas correlacionadas e não-concordantes com viés correlacionado com média das medidas.

A Tabela 1 apresenta o banco de dados com as seguintes variáveis:

sujeito: número de identificação da mulher;

dum: número de semanas de gestação calculado a partir da DUM;

eco_1: número de semanas de gestação calculado a partir do ultrassom obtido após a 20ª semana de gestação;

eco_2: número de semanas de gestação calculado a partir do ultrassom obtido antes da 20ª semana de gestação;

eco_3: as primeiras 25 mulheres tiveram o número de semanas de gestação calculado a partir do ultrassom obtido após a 20ª semana de gestação e as restantes antes da 20ª semana de gestação.

Tabela 1 - Banco de dados.

sujeito	dum	eco_1	eco_2	eco_3
1	36,7	37,0	37,8	37,0
2	37,6	37,8	38,6	37,8
3	39,1	39,5	40,9	39,5
4	39,8	39,4	41,1	39,4
5	38,3	38,4	40,3	38,4
6	37,0	36,6	37,5	36,6
7	39,0	38,7	39,5	38,7
8	39,1	39,3	40,2	39,3
9	38,6	38,6	39,2	38,6
10	37,5	37,5	38,7	37,5
11	39,3	39,4	41,1	39,4
12	38,2	37,8	39,7	37,8
13	36,3	36,1	38,1	36,1
14	38,8	38,5	39,1	38,5
15	38,1	38,5	39,5	38,5
16	38,0	38,0	39,6	38,0
17	38,0	38,1	38,1	38,1
18	37,8	37,6	38,9	37,6
19	36,6	36,4	37,6	36,4
20	36,4	36,8	36,7	36,8
21	37,9	37,6	38,0	37,6
22	39,4	39,2	40,4	39,2
23	37,3	37,5	38,8	37,5
24	37,0	36,7	37,7	36,7
25	38,1	38,2	38,5	38,2
26	36,4	36,5	37,9	37,9
27	38,0	38,3	38,4	38,4
28	38,8	38,6	39,9	39,9
29	37,2	37,1	37,7	37,7
30	37,9	38,3	39,7	39,7
31	37,7	37,4	38,9	38,9
32	38,2	38,0	38,4	38,4
33	36,1	36,4	37,4	37,4
34	39,0	39,3	40,9	40,9
35	37,6	37,9	38,4	38,4
36	38,1	38,0	39,5	39,5
37	38,9	39,1	40,7	40,7
38	38,1	38,3	39,0	39,0
39	38,0	37,8	38,0	38,0
40	38,2	37,9	38,5	38,5
41	38,6	38,7	39,7	39,7
42	38,9	39,1	40,1	40,1
43	37,8	37,7	38,0	38,0
44	38,4	38,3	38,8	38,8
45	37,2	36,9	37,5	37,5
46	37,6	37,3	38,0	38,0
47	38,1	38,1	38,2	38,2
48	36,7	36,5	36,9	36,9
49	37,8	37,6	39,5	39,5
50	38,2	38,4	39,2	39,2

A seguir vamos interpretar os resultados das comparações em cada uma das situações colocadas anteriormente.

Medidas correlacionadas e concordantes

Quando comparamos o número de semanas de gestação calculado a partir da DUM com o número de semanas de gestação calculado a partir do ultrassom obtido após a 20ª semana de

gestação (eco_1) verificamos que as medidas têm uma forte correlação, pois o coeficiente de correlação é 0,96 (Figura 1a). Elas também apresentam uma boa concordância, uma vez que pelo gráfico da metodologia de Bland-Altman (Figura 1b) podemos perceber que o viés é próximo de zero e não estatisticamente significativo. Os limites de concordância apontam que as diferenças entre as duas medidas parecem ser menores do que 4 dias (0,52 x 7 dias).

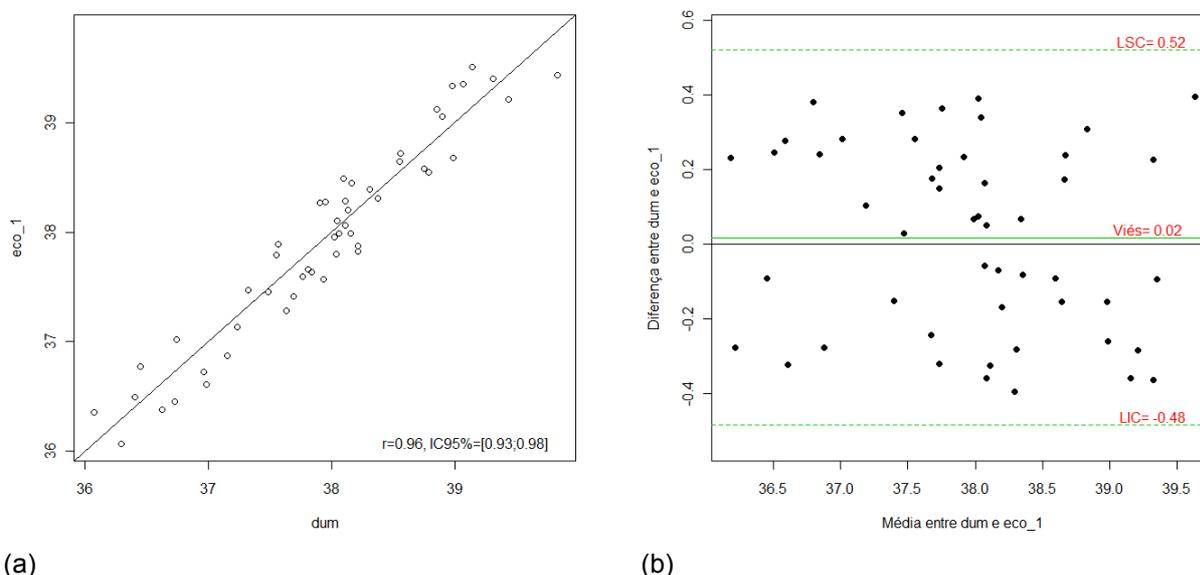


Figura 1 - Gráfico de dispersão para as variáveis dum e eco_1 (a) e para a diferença e média entre dum e eco_1 (b). LSC: limite superior de concordância; LIC: limite inferior de concordância.

Medidas correlacionadas e não-concordantes com viés constante

Na comparação do número de semanas de gestação calculado a partir da DUM com o número de semanas de gestação calculado a partir do ultrassom obtido antes da 20ª semana de gestação (eco_2) notamos novamente que as medidas também são fortemente correlacionadas ($r=0,86$; Figura 2a). Entretanto, a Figura 2b aponta que existe um viés estatisticamente significativo. Com a distribuição espacial dos pon-

tos da Figura 2b é homogênea, no sentido de não indicar uma relação entre as diferenças e as médias das medidas, podemos concluir que esse viés é sistemático. Portanto, apesar das medidas não serem concordantes poderíamos usar o número de semanas de gestação calculado a partir da DUM mais o viés (0,96 semana, aproximadamente 1 semana) como uma estimativa do número de semanas de gestação calculado a partir do ultrassom obtido antes da 20ª semana de gestação.

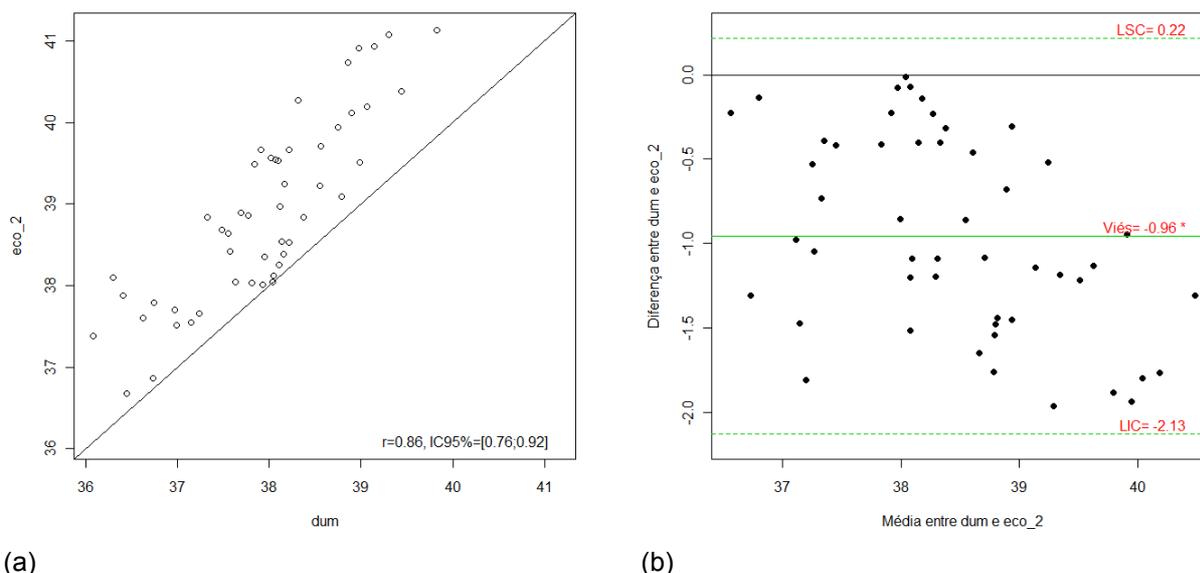


Figura 2 - Gráfico de dispersão para as variáveis dum e eco_2 (a) e para a diferença e média entre dum e eco_2 (b). LSC:limite superior de concordância, LIC: limite inferior de concordância.

Medidas correlacionadas e não-concordantes com viés correlacionado com média das medidas

Na Figura 3, temos um exemplo no qual novamente as medidas têm uma forte correla-

ção, mas além das medidas não concordarem, existe uma relação entre o viés (diferença entre as medidas) e a média da DUM e da eco_3. Percebemos neste caso que a o viés parece ser maior para valores de médias maiores que 38 semanas de gestação.

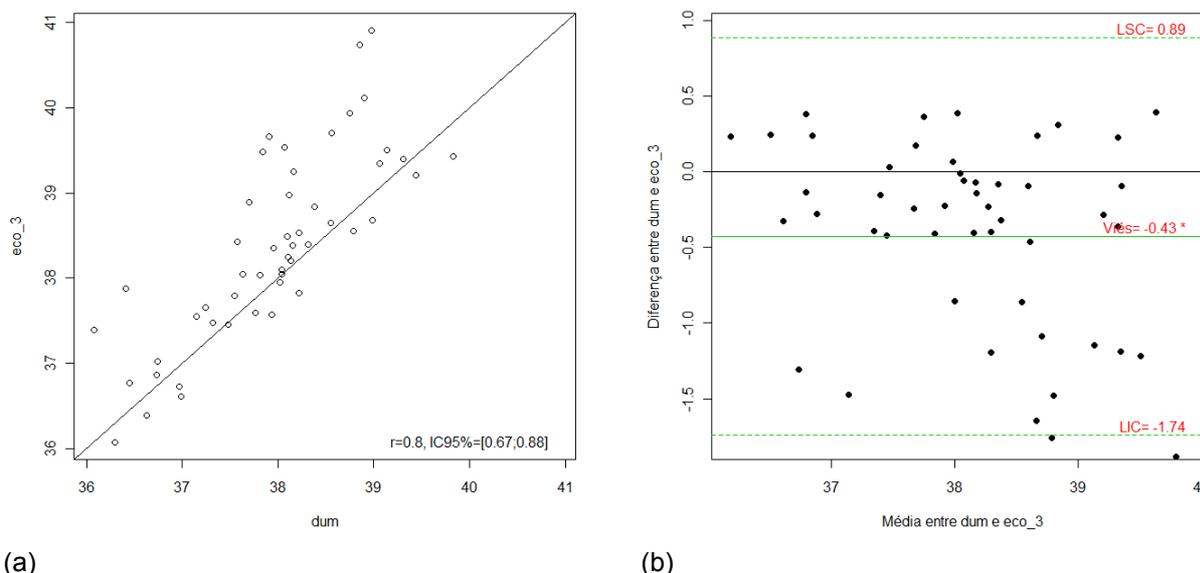


Figura 3 - Gráfico de dispersão para as variáveis dum e eco_3 (a) e para a diferença e média entre dum e eco_3 (b). LSC: limite superior de concordância; LIC: limite inferior de concordância.

Quando selecionamos no banco de dados as primeiras 25 mulheres, ou seja, aquelas nas quais o número de semanas de gestação foi calculado a partir do ultrassom obtido após a 20ª semana de gestação percebemos que o viés

é não estatisticamente significativo e praticamente igual a zero (Figura 4b). Também não há mais evidências de que o viés esteja relacionado com as médias.

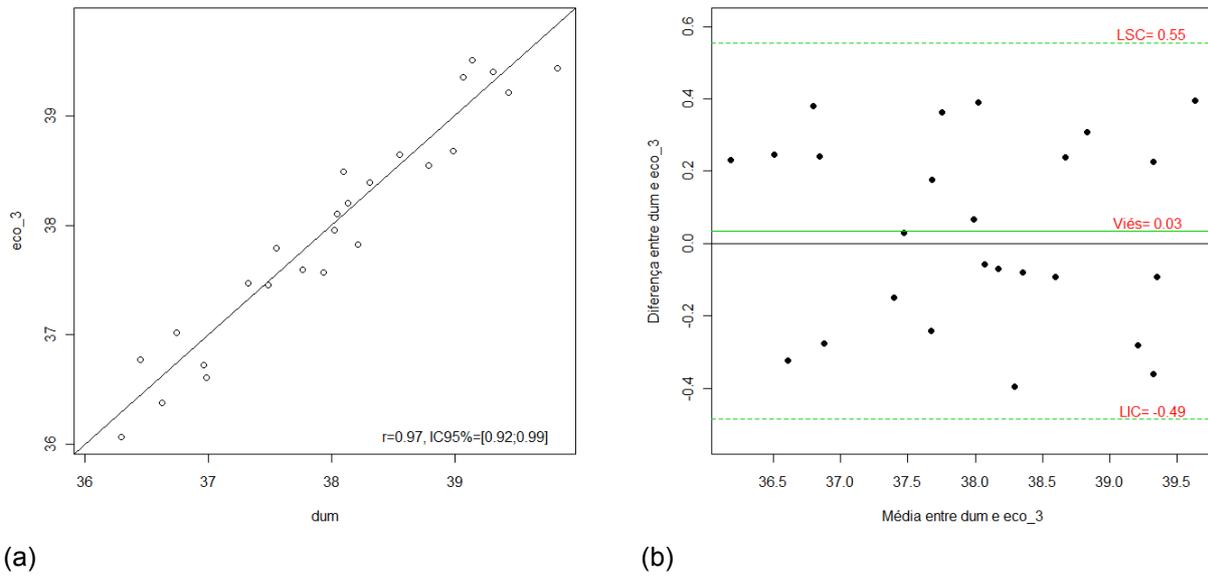


Figura 4 - Gráfico de dispersão para as variáveis dum e eco_3 (a) e para a diferença e média entre dum e eco_3 (b) apenas das primeiras 25 mulheres (ultrassom obtido após a 20ª semana de gestação). LSC: limite superior de concordância; LIC: limite inferior de concordância.

Na Figura 5, temos a situação em que as medidas não concordam, pois o viés é estatisticamente significativo, e que o viés aumenta à medida que a média entre dum e eco_3 aumenta. Isso impossibilita o uso do viés como um fator de correção entre as medidas.

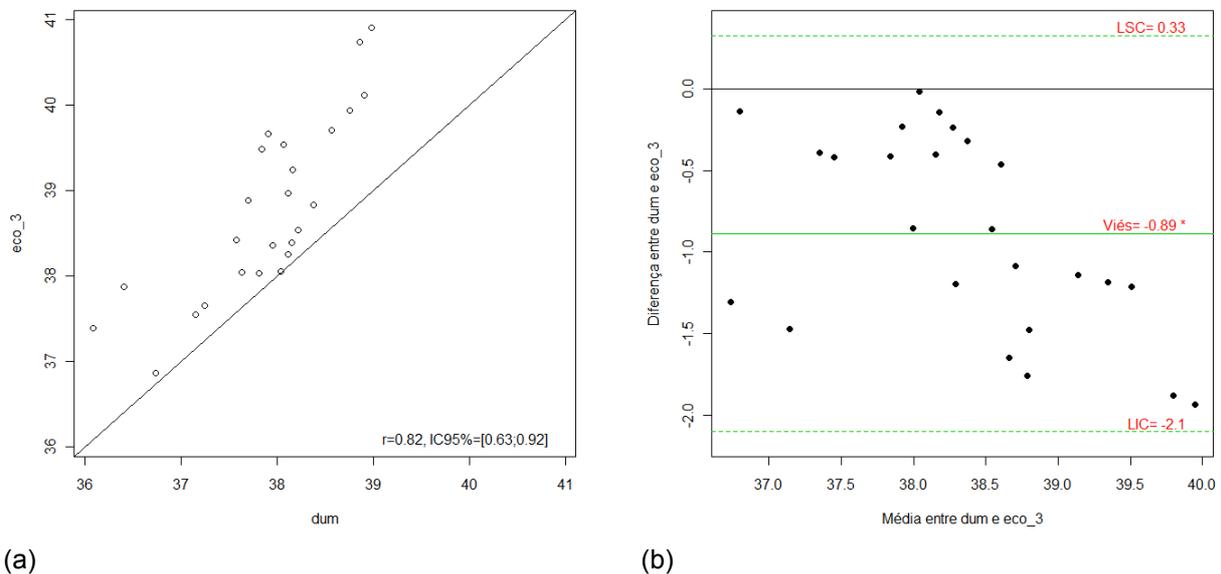


Figura 5 - Gráfico de dispersão para as variáveis dum e eco_3 (a) e para a diferença e média entre dum e eco_3 (b) apenas das últimas 25 mulheres.

CONSIDERAÇÕES FINAIS

Apesar de amplamente divulgado, o método ainda não é devidamente utilizado. Segundo alguns autores que fizeram pesquisas em revistas de diferentes áreas (8, 9), ainda é comum a utilização do método de forma incompleta, por exemplo, sem incluir os limites de concordância

ou ainda sem calcular os limites de confiança dos limites de concordância. Outro erro bastante citado é o de se utilizar os valores de algum dos métodos no lugar da média, artifício este que acaba forçando uma correlação que, na realidade, pode não ser verdadeira (9).

Os limites de concordância devem ser avaliados do ponto de vista clínico, ou seja, se

aquelas diferenças dadas pelos limites podem ser consideradas aceitáveis do ponto de vista clínico (4). Alguns autores, inclusive, apontam que esta diferença aceitável deve ser estabelecida a priori, antes do estudo ser realizado (8).

O coeficiente de correlação é uma medida de associação, não sendo apropriado para avaliar concordância entre métodos (1,2). Portanto, quando o objetivo do trabalho for avaliar a concordância entre dois métodos, a metodologia de Bland-Altman deve ser utilizada de forma completa, incluindo os limites de concordância, e seus intervalos de confiança. Uma possível associação entre o viés e a magnitude das medidas também deve ser avaliada e, quando necessário, fazer alguma correção nos dados para uma correta conclusão.

REFERÊNCIAS

1. Altman DG, Bland JM. Measurement in Medicine: The analysis of method comparison studies. *Journal of the Royal Statistical Society. Series D (The Statistician)*. 1983;32(3):307-17.
2. Bland JM, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;327(8476):307-10.
3. Bland JM, Altman D. Comparing two methods of clinical measurement: a personal history. *Int J Epidemiol*. 1995;24(suppl_1):S7-14.
4. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*. 1999;8(2):135-60.
5. Linnet K, Bruunshuus I. HPLC with enzymatic detection as a candidate reference method for serum creatinine. *Clin Chem*. 1991;37(10):1669-75.
6. Sample size for a study of agreement between two methods of measurement [Internet]. [citado 2010 Jan 7]; Available from: <http://www-users.york.ac.uk/~mb55/meas/sizemeth.htm>
7. R Development Core Team. R: A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; Available from: <http://www.R-project.org>
8. Mantha S, Roizen MF, Fleisher LA, Thisted R, Foss J. Comparing methods of clinical measurement: reporting standards for Bland and Altman analysis. *Anesth Analg*. 2000;90(3):593-602.
9. Dewitte K, Fierens C, Stockl D, Thienpont LM. Application of the Bland-Altman plot for interpretation of method-comparison studies: a critical investigation of its practice. *Clin Chem*. 2002;48(5):799-801.

Recebido: 07/10/09

Aceito: 09/11/09