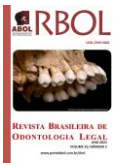
Revista Brasileira de Odontologia Legal – RBOL

ISSN 2359-3466 http://www.portalabol.com.br/rbol



Forensic anthropology

RELIABILITY OF MEASUREMENTS TAKEN FROM PANORAMIC RADIOGRAPHS FOR AUTOMATED SEX DETERMINATION.

Confiabilidade das medições obtidas em radiografias panorâmicas para determinação automatizada do sexo.

Edgard MICHEL-CROSATO[1], Adrielly Garcia ORTIZ[1], Maria Gabriela Haye Biazevic[1].

1. Departamento de Odontologia Social, Odontologia, PRPG Ciências Odontológicas. Faculdade de Odontologia da Universidade de São Paulo, São Paulo, Brasil.

Information about the manuscript:

Received: October 7, 2023
Accepted: January 3, 2024

Contact author:
Prof. PhD.Edgard Michel Crosato
Faculdade de Odontologia - Universidade de São Paulo. Av. Professor Lineu Prestes, 2227 - Cidade Universitária - São Paulo/SP - CEP: 05508-000.
E-mail: michelcrosato@usp.br.

ABSTRACT

Introduction: forensic anthropology currently serves as a potent tool in primary and secondary identification, especially in mass disasters, increasingly common in our society. Objective: Verify the reproducibility of measured measurements. Proposes methods of personal identification. Method: to verify the reliability of the method, 25 radiographs were selected, the intra-class coefficient (ICC) and the Bland-Altiman statistical technique were calculated. accuracy of the method, was used. R studio and Medcalc were used. Results: The ICC was above 0.90 for all measurements. According to the Bland-Altman analysis, the average differences observed between examiners ranged from 0.017 (m5) to 0.020 (m7). These results highlight the reliability of measurements made by two different examiners performing the same tasks. Conclusion: The variables used in this study have shown satisfactory reproducibility. Thus, the selected and evaluated data can be used as metadata in automating the sex determination process from panoramic radiographs.

KEYWORDS

Forensic dentistry; Forensic anthropology; Determination of sex.

INTRODUCTION

Forensic anthropology applied to dentistry, in practical scenarios, primarily aims to determine sex, estimate age, ancestry, stature, and other critical factors through bone structure analysis. Forensic anthropology currently serves as a potent tool in primary and secondary identification, especially in mass disasters, increasingly common in our society[1].

In the secondary identification process, gender assessment often acts as an initial element aiding an individual's identification process[2-3]. Two primary requirements of a diagnostic test are reproducibility and validity. Reproducibility, or reliability, refers to a method's ability to

yield consistent results across different examiners (inter-examiner reproducibility) conducting the test on the same patient or even the same examiner repeating the results at different times (intra-examiner reproducibility). Validity (or accuracy) pertains to the method's capability to measure what it indeed intends to measure[4].

Reproducibility can be described as the level of consistency or agreement in results when measurements or examinations are repeated under identical conditions. When measuring study variables quantitatively, the intra-class correlation coefficient (ICC) replaces the Kappa statistic, with its value ranging from 0 to 1. The calculation is based on variations between units and the sample's total variation. Populations with high variability tend to overestimate the indicator, making it not the preferred choice nowadays[4-6].

A widely adopted alternative is the Bland & Altman statistics from 1996 and 1999. This graphical tool is a method to compare two measurement techniques with quantitative variables. In this approach, differences between the two techniques are presented concerning the averages of both techniques. Alternatively, differences can be made concerning one of the two methods if that method is a reference method. Horizontal lines are drawn at the average difference and at agreement limits, defined as the average difference plus and minus 1.96 times the standard deviation of the differences. It's also feasible to use duplicate or multiple measurements per subject for each method, employing the Bland-Altman for several repetitions[4-8].

Data science techniques can be employed for sex prediction. Concepts like Big Data, Machine Learning, Data Mining are being integrated into forensic practices. Some studies showcase the best algorithms for sex prediction. Metadata, which is data about other data, is crucial when using images for sex determination. In imaging, metadata is data that describes an image. They provide information about the object, such as its creation, content, structure, ownership, and use, and others[9]. The choice of data accompanying the images plays a significant role in devising algorithms for automating this process[10-14].

In this context, this study's objective is to verify the reliability of measurements (metadata) taken using panoramic radiographs for automated sex determination for future application in Big Data, Machine Learning, Data Mining.

METHODOLOGY

Study Type

A cross-sectional study using panoramic radiographs to verify the reliability of the proposed method.

Study Population:

For the current study, 100 panoramic radiographs were randomly selected from a private collection, with 50 exams from female patients and 50 from male patients. To verify the method's accuracy, 25 panoramic radiographs were used. For analysis, 100 exams were measured, 50 from female patients and 50 from male patients, divided into 4 groups: ex1, ex2, t1, and t2.

Examiner Calibration Process

Two examiners underwent training and a calibration process to measure the measurements and indices. The intra-class coefficient (ICC), the Coefficient of Variation, and the Bland-Altman statistics were calculated.

Study Variables:

For the training definition of the algorithms used in the automation of measurements in pixels made with the ImageJ program on panoramic radiographs, 4 out of the 13 linear and angular measurements on the patients' mandible were used (Table 1 and Figure 1). These served as the basis for analyses following the Ortiz et al. methodology, 2020[15]. Measurements were taken in pixels. Reference points used for the measurements included: right and left condyles; mandible base; coronoid processes; mental foramina; gonion; mandible angle; chin, and the median line was drawn.

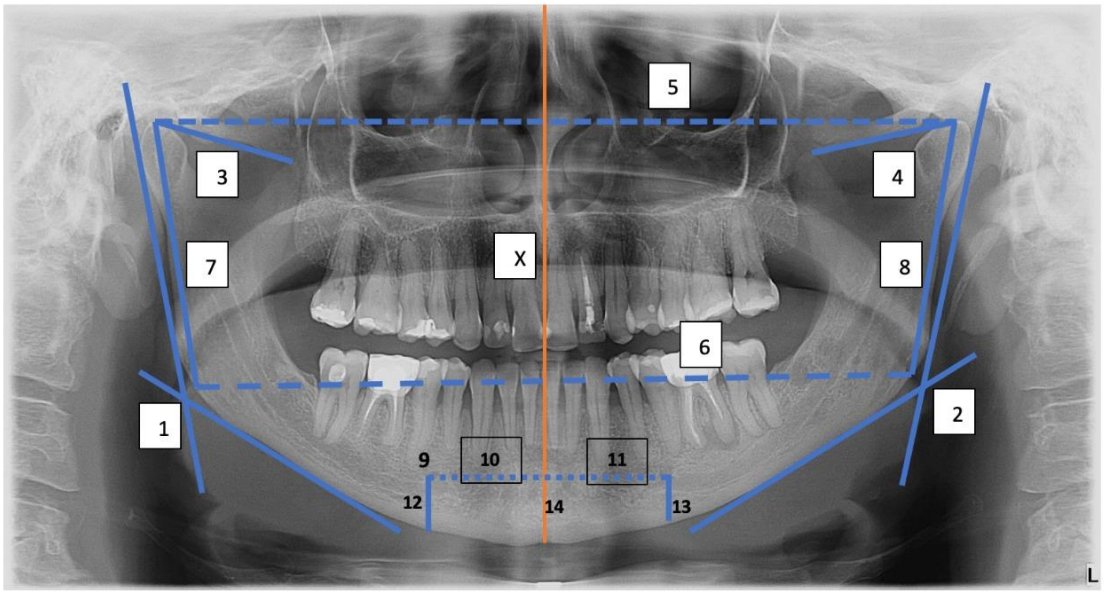


Figure 1. Bland-Altman plot for intra-rater reliability.

Table 1: Description of the measures used.

1- AMD (D)	8- C – Go (E)
2- AMD (E)	9- FM – FM: FM – PSM (D)
3- C – Co (D)	10- FM – PSM (E)
4- C – Co (E)	11- FM – FM x PSM
5- C – C	12- FM – BMD (D)
6- Go – Go	13- FM – BMD (E)
7- C – Go (D)	14- X Me

D - Right; E - Left; AMD - Right mandibular angle; C - Condyle; Co - Coronoid process; Go - Gonion; FM - Mental foramen; PSM - Median sagittal plane; Me - Menton; BMD - Base of the Mandible; X - Intersection point Me-Me - PSM-Me.

Statistical Analysis

To assess intra-rater and inter-rater reliability, we utilized the Intraclass Correlation Coefficient (ICC) and the Bland-Altman statistical technique.

We began our data analyses by checking the adherence to the normality curve using the Shapiro-Francia statistical test. In the initial phase of the analyses, all variables were deemed non-normal. To discern any potential differences in measurements between male and female subjects, we applied the Mann-Whitney test.

Standardization of the variables was required, wherein their initial values were divided by 100.

In addition to the ICC, we also calculated the Coefficient of Variation and employed the Bland-Altman statistics. For these analyses, we used the Medcalc and R Studio software, setting a significance level of 5%.

Ethical Considerations

The study was submitted to and approved by the FOUSP Research Ethics Committee, under the Opinion 79354517000000075.

RESULTS

Randomly, 50 female and 50 male panoramic radiographs were selected. In each, four linear measures were taken, and both intra-rater and inter-rater agreement were assessed. Each group used 25 radiographs.

Table 2 displays the descriptive measures and the adherence to the normality curve of variables m5, m6, m7, and m8, according to examiner 1 and 2, at times 1 and 2. Inter-rater reliability was initially assessed using the Intraclass Correlation Coefficient – Figure 2. The results in Table 3 indicate a good inter-rater reliability index. Subsequently, the Bland-Altman statistical method was applied to analyze the data distribution between examiners.

Table 2. Descriptive measures of the variables m5, m6, m7, and m8, according to examiner and time.

	N	Minimum	Maximum	Average	Median	Standard Dev.	Normal Distr.
m5_ex1	25	1,780	3,320	2,144	1,970	0,4820	<0,0001
m5_ex2	25	1,770	3,340	2,133	1,940	0,4786	<0,0001
m5_t1	25	1,740	3,180	2,137	2,000	0,4346	<0,0001
m5_t2	25	1,730	3,180	2,123	1,940	0,4309	<0,0001
m6_ex1	25	0,740	2,990	1,984	1,940	0,4198	0,0002
m6_ex2	25	0,700	2,920	1,966	1,900	0,4164	0,0001
m6_t1	25	1,800	3,370	2,096	1,940	0,4190	<0,0001
m6_t2	25	1,740	3,360	2,064	1,910	0,4288	<0,0001
m7_ex1	25	0,560	1,140	0,738	0,680	0,1612	0,0001
m7_ex2	25	0,580	1,090	0,719	0,660	0,1492	<0,0001
m7_t1	25	0,590	1,060	0,729	0,690	0,1393	0,0001
m7_t2	25	0,550	1,040	0,712	0,690	0,1291	0,0008
m8_ex1	25	0,260	0,990	0,696	0,680	0,1363	0,0190
m8_ex2	25	0,250	1,010	0,677	0,640	0,1384	0,0027
m8_t1	25	0,550	1,070	0,710	0,680	0,1315	0,0004
m8_t2	25	0,540	1,070	0,701	0,680	0,1283	0,0001

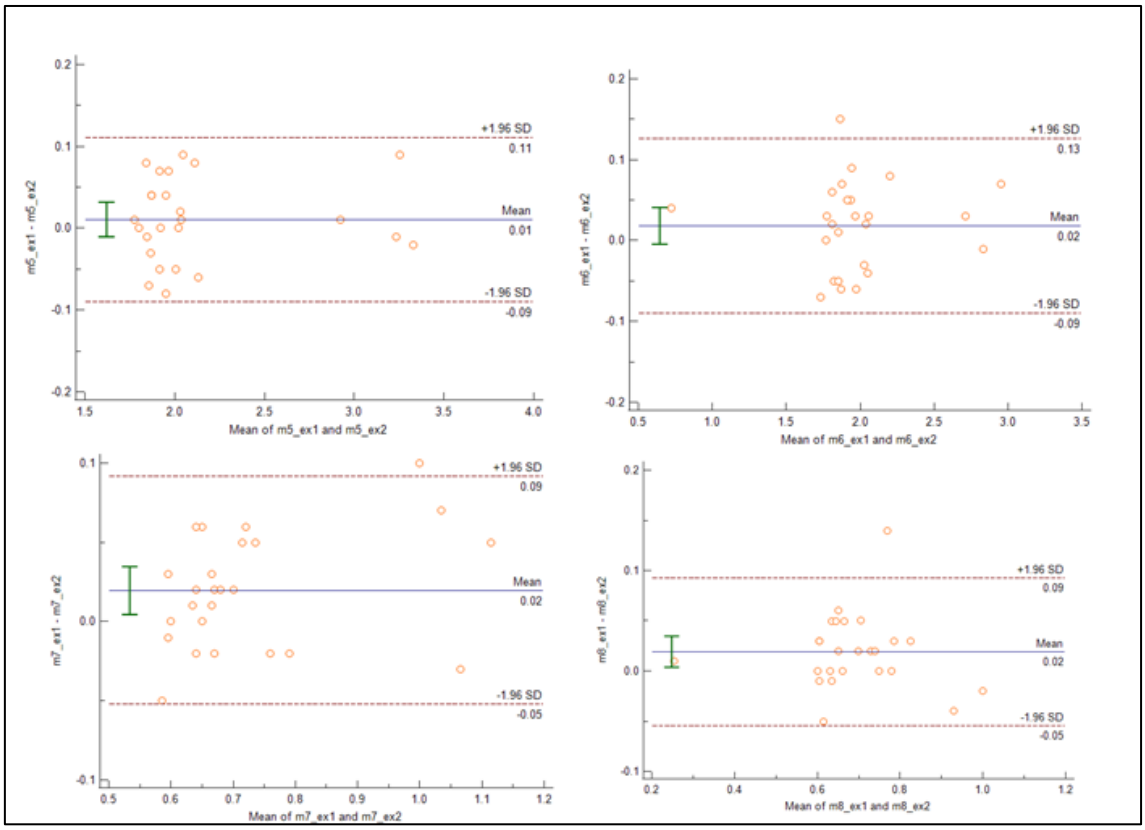


Figure 2. Bland-Altman plot for intra-rater reliability.

Table 3: Reliability of the variables m5, m6, m7, and m8.

	N	ICC	IC(95%)
Inter-rater reliability			
m5_ex1Xex2	50	0,994	0,987-0,997
M6_ex1Xex2	50	0,990	0.979-0,995
M7_ex1Xex2	50	0,965	0,924-0,984
M8_ex1Xex2	50	0,954	0,900-0,979
Intra-rater reliability			
m5_t1Xt2	50	0,994	0,987-0,997
m5_t1Xt2	50	0,988	0.973-0,994
m5_t1Xt2	50	0,952	0,896-0,978
m5_t1Xt2	50	0,973	0,940-0,988

Figure 3 illustrates the Bland-Altman plot for the inter-rater reliability. For intra-rater reliability, the initial step was the utilization of the Intraclass Correlation Coefficient. Table 3 showcases the data, with results indicating a good intra-rater reliability index.

DISCUSSION

For over seven decades, forensic studies have sought the best way to determine the gender of human remnants. Most of these estimates are based on dental analyses[13-14], while few studies explore other structures for such determination[15-17].

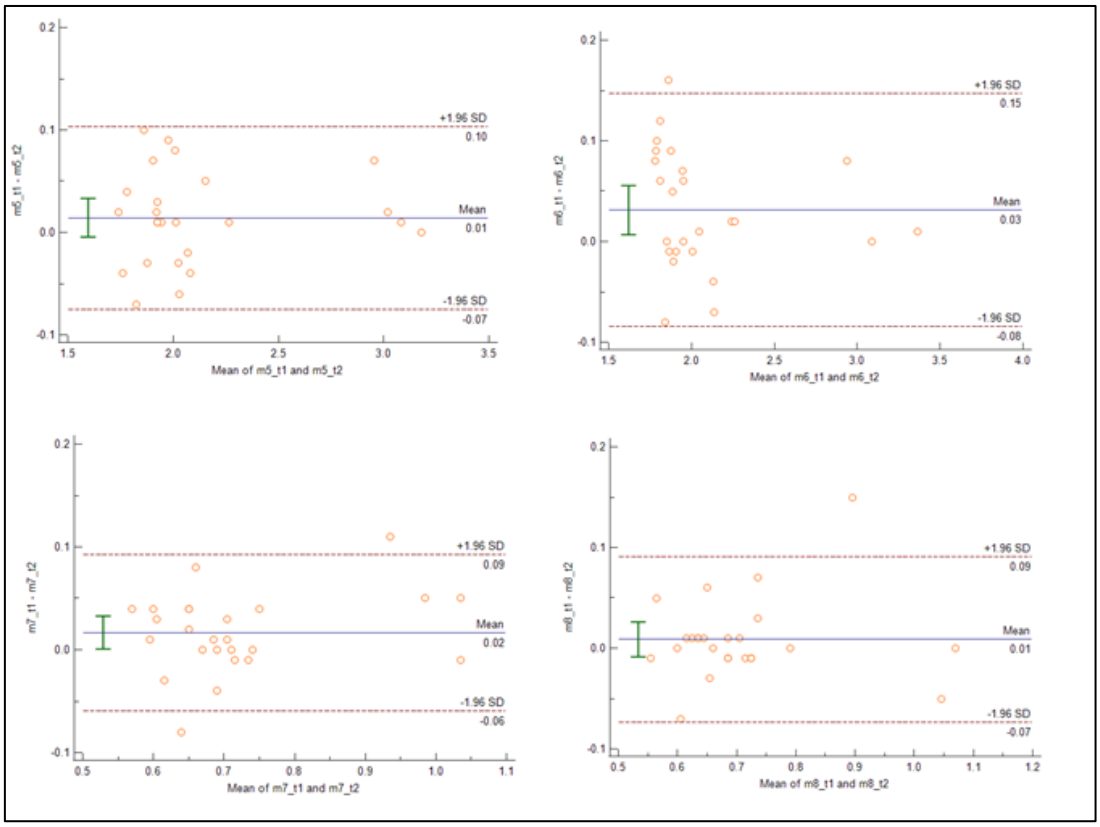


Figure 3. Bland-Altman plot for intrer-rater reliability.

The data from this study showed good reproducibility. The Bland-Altman analysis revealed that the average differences between examiners ranged between 0.017 (m5) and 0.020 (m7), underscoring the reliability of the measurements[5-8].

High reproducibility rates in forensic dental imaging exams are essential to ensure the reliability of individual identification, the credibility of the field, and the future development of AI in this important forensic area[18].

In this type of study design, it is important to determine the sample size. The sample for this purpose should be composed of individuals who represent the variability of the target population. The sample should also be large enough to provide accurate results. In general, it is recommended that the sample have at least 20 subjects or objects for each observer. However, if the expected effect size is small, or if the confidence level or acceptable error rate is low, a larger sample may be needed[19]. In the case presented, the smallest number of radiographs presented for analysis was 25 per examiner.

Thus, the tested and selected data have the potential to be used as metadata in the automation of gender determination in panoramic radiographs.

CONCLUSION

The findings of this study suggest that the proposed metadata are reliable for automating measurements in panoramic radiographs, serving as an auxiliary method in determining sex for forensic practices.

RESUMO

Introdução: A antropologia forense atualmente serve como uma ferramenta potente na identificação primária e secundária, especialmente em desastres em massa, cada vez mais comuns em nossa sociedade. Objetivo: Verificar a reprodutibilidade das medidas mensuradas. Propõe métodos de identificação pessoal. Método: para verificar a confiabilidade do método, foram selecionadas 25 radiografias, calculando-se o coeficiente intraclasse (ICC) e a técnica estatística de Bland-Altiman. Para avaliar a precisão do método, foram utilizados o R studio e o Medcalc. Resultados: O ICC foi superior a 0.90 para todas as medições. De acordo com a análise de Bland-Altman, as diferenças médias observadas entre os examinadores variaram de 0.017 (m5) a 0.020 (m7). Esses resultados destacam a confiabilidade das medições feitas por dois examinadores diferentes realizando as mesmas tarefas. Conclusão: As variáveis utilizadas neste estudo mostraram reprodutibilidade satisfatória. Assim, os dados selecionados e avaliados podem ser usados como metadados na automação do processo de determinação do sexo a partir de radiografias panorâmicas.

PALAVRAS-CHAVE

Odontologia legal; Antropologia forense; Determinação do sexo.

REFERENCES

1. Cattaneo C. Forensic anthropology: developments of a classical discipline in the new millennium. Forensic Sci Int. 2007 Jan 17;165(2-3):185-93. https://doi.org/:10.1016/j.forsciint.2006.05.018.
2. Sambhana S, Sanghvi P, Mohammed RB, Shanta PP, Thetay AA, Chaudhary VS. Assessment of sexual dimorphism using digital orthopantomographs in South Indians. J Forensic Dent Sci. 2016 Sep-Dec;8(3):180. https://doi.org/10.4103/0975-1475.195113.
3. Damera A, Mohanalakhsmi J, Yellarthi PK, Rezwana BM. Radiographic evaluation of mandibular ramus for gender estimation: Retrospective study. J Forensic Dent Sci. 2016 May-Aug;8(2):74-8. https://doi.org/10.4103/0975-1475.186369.
4. Silva AF, Velo MMAC, Pereira AC. Importância da reprodutibilidade dos métodos para diagnóstico em odontologia. RFO UPF. 2016;21(1):115-20.
5. Kimberlin CL, Winterstein AG. Validity and reliability of measurement instruments used in research. Am J Health Syst Pharm. 2008;65(23):2276-84. https://doi.org/10.2146/ajhp070364.
6. De Angelis D, Riboli F, Gibelli D, Cappella A, Cattaneo C. Palatal rugae as an individualising marker: reliability for forensic odontology and personal identification. Sci Justice. 2012 Sep;52(3):181-4. https://doi.org/10.1016/j.scijus.2011.09.002.
7. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986 Feb 8;1(8476):307-10.
8. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res. 1999 Jun;8(2):135-60. https://doi.org/10.1177/096228029900800204.
9. International Organization for Standardization. (2011). Information and documentation - Thesauri and interoperability with other vocabularies - Part 1: Thesauri for information retrieval. ISO 25964-1:2011. Geneva: ISO. Retrieved from https://www.iso.org/standard/65475.html.
10. Lefèvre T. Big data in forensic science and medicine. J Forensic Leg Med. 2018 Jul;57:1-6. https://doi.org/10.1016/j.jflm.2017.08.001.
11. DeLisi M. The big data potential of epidemiological studies for criminology and forensics. J Forensic Leg Med. 2018 Jul;57:24-27. https://doi.org/10.1016/j.jflm.2016.09.004.
12. Martins Filho IE, Lopez-Capp TT, Biazevic MG, Michel-Crosato E. Sexual dimorphism

using odontometric indexes: Analysis of three statistical techniques. J Forensic Leg Med. 2016 Nov;44:37-42. https://doi.org/10.1016/j.jflm.2016.08.010.

13. Pettenati-Soubayroux I, Signoli M, Dutour O. Sexual dimorphism in teeth: discriminatory effectiveness of permanent lower canine size observed in a XVIIIth century osteological series. Forensic Sci Int. 2002 May 23;126(3):227- 32. https://doi.org/10.1016/s0379-0738(02)00080-4.
14. Omar A, Azab S. Applicability of Determination of Gender from Odontometric Measurements of Canine Teeth in a Sample of Adult Egyptian Population, Cairo. Dent J. 2009;25(2):167-80.
15. Ortiz AG, Costa C, Silva RHA, Biazevic MGH, Michel-Crosato E. Sex estimation: Anatomical references on panoramic radiographs using Machine Learning. Forensic Imaging. 2020(20):200356.
16. Hunt EE Jr, Gleiser I. The estimation of age and sex of preadolescent children from bones and teeth. Am J Phys Anthropol. 1955 Sep; 13(3):479-87. https://doi.org/10.1002/ajpa.1330130308.
17. Capitaneanu C, Willems G, Thevissen P. A systematic review of odontological sex estimation methods. J Forensic Odontostomatol. 2017 Dec 1;35(2):1-19.
18. Vodanović M, Subašić M, Milošević DP, Galić I, Brkić H. Artificial intelligence in forensic medicine and forensic dentistry. J Forensic Odontostomatol. 2023 Aug 27;41(2):30-41.
19. Gwet KL. A practical guide to sample size determination for intra- and inter-observer agreement studies. Journal of Applied Statistics, 2020 May 35(4), 531-41. https://doi.org/10.1348/000711006X126600