Study of reliability and concordance between two VDRL tests for syphilis investigation in patients with HIV followed-up at a university hospital in Rio de Janeiro in 2017–2019

Estudo de confiabilidade e concordância entre dois testes VDRL para investigação de sífilis em pacientes com HIV em acompanhamento em um hospital universitário na cidade do Rio de Janeiro, entre os anos de 2017–2019

Isabelle de Carvalho Rangel[1], Ricardo de Souza Carvalho[1,2], Mauricéa Francisco da Silva Romero Gonzalez[1,2], Patrick Menezes Loureço[2], Beatriz Pereira de Azevedo[3], Fernando Raphael de Almeida Ferry[1]

ABSTRACT

Introduction: The laboratory diagnosis of syphilis is given by a positive treponemal test and a non-treponemal test, with VDRL (Veneral Disease Research Laboratory) being the "gold standard". **Objective:** To compare two tests commercially validated for biological fluids and analyzed by different operators, in order to assess their performance in detecting high (≥:8) and low (≤1:2) titrations, as well as to determine the agreement between results in paired serum samples from patients with syphilis and living with HIV. **Methods:** Cross-sectional study, approved by the Research Ethics Committee of the teaching hospital Gaffrée e Guinle (HUGG), under CAAE 66558117.0.0000.5258. The study population was composed by patients diagnosed with syphilis and confirmed by the positivity of one or more treponemal tests. All samples were analyzed simultaneously by two different operators, each using a kit: VDRL WAMA Diagnóstica®, São Carlos, SP, Brazil; VDRL Brás, Laborclin®, Pinhais, PR, Brazil. The SPSS statistical program was used. **Results:** 110 serum samples from patients diagnosed with syphilis treated at HUGG were analyzed. The frequency of high VDRL titrations among patients, following the VDRL criterion ≥1:8, was practically the same in both tests, with 68% in VDRL Laborclin and 69% in VDRL WAMA (p = 0.87) and VDRL ≤1:2, 80% for WAMA and 83% for Laborclin (p = 0.72). The results of VDRL were tabulated in pairs; then the Cohen's Kappa coefficient of agreement was calculated (K) 0.32 (95%CI 0.21–0.41; p<0.00001), as well as the weighted Kappa (Kw) and the intraclass correlation coefficient (ICC) 0.89 (95%CI 0.84–0.92; p<0.00001). The Bland-Altman diagram was also used. We found poor agreement between the VDRL tests when results were nominally concordant, that is, with the same titles in both tests. However, if partial agreement is considered, the interpretation of the magnitude of agreement estimators was almost complete (≥0.80). **Conclusion:** Reliability and agreement were high between the VDRL tests of both manufacturers when considering the close titrations (up to two dilutions). Further reliability and agreement studies are essential between the non-treponemal tests available and used in Brazil.

Keywords: serum tests; HIV; syphilis.

RESUMO

Introdução: O diagnóstico laboratorial da sífilis é realizado por meio da positividade de um teste treponêmico e de um teste não treponêmico, sendo o VDRL (do inglês *Veneral Disease Research Laboratory*) o "padrão ouro". **Objetivo:** Comparar dois testes comercialmente validados para fluidos biológicos e analisados por operadores diferentes, com o intuito de avaliar o desempenho dos testes em detectar titulações altas (≥ :8) e baixas (≤1:2), bem como determinar a concordância entre ambos os resultados em amostras pareadas de soro de pacientes com sífilis vivendo com HIV. **Métodos:** Estudo transversal, aprovado pelo Comitê de Ética em Pesquisa do Hospital Universitário Gaffrée e Guinle (HUGG), sob o CAAE: 66558117.0.0000.5258. A população estudada foi a de pacientes que obtiveram o diagnóstico de sífilis confirmado por meio da positividade de um ou mais testes treponêmicos. Todas as amostras foram analisadas simultaneamente por dois operadores diferentes, cada um utilizando um *kit*: VDRL WAMA Diagnóstica®, São Carlos, SP, Brasil; VDRL Brás, Laborclin®, Pinhais, PR, Brasil. Utilizou-se o programa estatístico SPSS. **Resultados:** Foram analisadas 110 amostras de soro de pacientes com diagnóstico de sífilis atendidos no HUGG. A frequência de altas titulações de VDRL entre os pacientes, seguindo o critério de VDRL ≥1:8, foi praticamente a mesma em ambos os testes, com 68% no VDRL Laborclin e 69% no VDRL WAMA (p=0,87) e para VDRL ≤1:2, 80% para WAMA e 83% para Laborclin (p=0,72). Os resultados dos títulos de VDRL foram tabulados em pares; em seguida, foram calculados o coeficiente de concordância Kappa de Cohen (K) 0,32 (IC95% 0,21–0,41; p<0,00001), o Kappa ponderado (Kw) e o coeficiente de correlação intraclasse (CCI) 0,89 (IC95% 0,84–0,92; p<0,00001), bem como utilizado o diagrama de Bland-Altman. O estudo encontrou fraca concordância entre os testes de VDRL, se considerados os resultados nominalmente concordantes, isto é, com os mesmos títulos em ambos os testes. Entretanto, se considerado a concordância parcial, a interpretação da magnitude dos estimadores de concordância passou a ser quase completa (≥0,80). **Conclusão:** A confiabilidade e a concordância foram altas entre os testes de VDRL dos dois fabricantes, quando consideradas as titulações próximas (até duas diluições). Mais estudos de confiabilidade e concordância são fundamentais entre os testes não treponêmicos disponíveis e utilizados no Brasil.

Palavras-chave: testes sorológicos; HIV; sífilis.

INTRODUCTION

Syphilis is caused by the spirochete bacteria *Treponema pallidum ssp. pallidum* (*T. pallidum*), which is sexually transmitted[1]. Syphilitic lesions increase the risk of infection by the human immunodeficiency virus

[1]Universidade Federal do Estado do Rio de Janeiro – Rio de Janeiro (RJ), Brazil.
[2]Hospital Universitário Gaffrée e Guinle – Rio de Janeiro (RJ), Brazil.
[3]Centro Universitário IBMR – Rio de Janeiro (RJ), Brazil.

(HIV); therefore, men who have sex with men (MSM) are often co-infected[2,3]. The laboratory diagnosis of syphilis is made through the positivity in the serum, plasma or blood sample of a treponemal test (TPHA, MHA-TP, TPPA, FTA-Abs, ELISA) combined with the reactivity of a non-treponemal test (VDRL, RPR, TRUST, USR, RST) [4,5]. The non-treponemal test most commonly used for its diagnosis and follow-up is the VDRL (Venereal Disease Research Laboratory), which is based on the use of antigen suspension composed of alcoholic solution containing cardiolipin, cholesterol, and purified lecithin, and which uses inactivated serum as a sample[6,7].

Currently, there are ten VDRL tests, from different manufacturers, registered by the National Health Surveillance Agency (ANVISA), marketed and widely used in Brazil. Although the tests of all manufacturers have similar specificity and sensitivity, their reactivity to the same sample can differ, since the antigen suspensions used in their composition have particularities. Thus, the same sample assessed with tests from different manufacturers or platforms show variable results in more or less in the final titration, even though it does not mean an error[8]. Likewise, tests using the same method, such as VDRL, can vary depending on reading subjectivity, that is, the eyes of the result evaluator[9]. That being said, the present study aimed to determine the agreement between two VDRL tests, from different manufacturers, used at the teaching hospital Gaffrée and Guinle (HUGG) and applied by different operators on serum samples from patients living with HIV and under investigation for syphilis being followed up at a university hospital in Rio de Janeiro.

OBJECTIVE

To compare two tests commercially validated for biological fluids and analyzed by different operators, in order to assess their performance in detecting high (≥:8) and low (≤1:2) titrations, as well as to determine the agreement between results in paired serum samples from patients with syphilis and living with HIV.

METHODS

Cross-sectional, analytical study approved by the Research Ethics Committee of HUGG/UNIRIO, under CAAE: 66558117.0.0000.5258. Serum samples were collected from patients followed up at the HUGG who had been submitted to serological investigation of syphilis and HIV and who were diagnosed with active or past syphilis, confirmed by a treponemal test (rapid treponemal test or FTA-ABS). All samples were from patients over 18 years of age, living with HIV and monitored at the HUGG Immunology Outpatient Clinic, who signed the informed consent form, from August 2017 to April 2019.

All samples were collected by venipuncture for blood collection and later centrifuged for 15 minutes at a rotation of 2,500G, to obtain serum and plasma; then, the two samples were aliquoted and analyzed on the same day by different operators, with tests from different manufacturers, both commercialized and widely used in Brazil: VDRL WAMA® (WAMA Diagnóstica, São Carlos, São Paulo, Brazil); and VDRL Brás® (Laborclin products for laboratory LTDA, Pinhais, Paraná, Brazil). For the purpose of "cut-off point" criteria, VDRL titration ≥1:8 was considered indicative of high titration and ≤1:2, of low titration for syphilis. In order to make the Bland-Altman diagram, the titrations were replaced by their equivalent in terms of dilution, thus becoming an ordinal variable, for example: VDRL 1:4 equals two dilutions (1:2=2) and 1:8 at three dilutions (1:8=3). The results were tabulated in pairs, and the agreement calculations were made in the SPSS statistical program, version 20 for MacOS, with proportion comparison calculated by the χ^2 test, with the aid of Prism 8 for MacOS.

RESULTS

Serum samples from 110 patients living with HIV were selected and their VDRL titration results were tabulated in pairs. The data were organized in a contingency table (**Table 1**) to make visualization of the agreement as a result of the titration easier. First, the Cohen's Kappa coefficient of agreement (K) was applied, with a result of 0.32 (95%CI 0.21–0.41, p<0.00001), and then the weighted Kappa coefficient of agreement (Kw) of 0.89 (95%CI 0.84–0.92, p<0.0001), which also corresponds to the intraclass correlation coefficient (ICC). The frequency of high VDRL titrations among patients, following the VDRL criterion ≥1:8, was practically the same in both tests, with 68% in the Laborclin VDRL and 69% in the VDRL WAMA (p=0.87) and for VDRL ≤1:2, 80% for WAMA and 83% for Laborclin (p=0.72). There was also no statistical difference in the proportions of VDRL ≤1:1 or ≥1:16 (p=0.20 and 0.78, respectively).

Table 1 – Frequency of titrations obtained in the VDRL WAMA and Laborclin tests, paired in exact agreement.

Cross tabulation VDRL Laborclin × VDRL WAMA													
		VDRL Laborclin											Total
		NR	1/1	1/2	1/4	1/8	1/16	1/32	1/64	1/128	1/512	1/1024	
VDRL WAMA	NR	7	5	2	1	1	0	0	0	0	0	0	16
	1/1	1	2	0	0	0	1	0	0	0	0	0	4
	1/2	1	1	0	2	2	0	0	0	0	0	0	6
	1/4	0	0	2	0	3	1	0	0	0	0	0	6
	1/8	0	0	1	1	6	2	6	0	1	0	0	17
	1/16	0	0	0	2	3	5	1	0	0	0	0	11
	1/32	0	0	0	3	1	4	5	1	1	0	0	15
	1/64	0	0	0	0	0	0	2	4	1	0	0	7
	1/128	0	0	0	0	0	0	0	1	7	0	0	8
	1/256	0	0	0	0	0	0	0	0	1	2	0	3
	1/512	0	0	0	0	0	0	0	0	1	1	0	2
	1/1024	0	0	0	0	0	0	0	0	1	1	2	4
	1/4096	0	0	0	0	0	0	0	0	0	0	1	1
Total		9	8	5	9	16	12	14	6	13	4	3	100

DISCUSSION

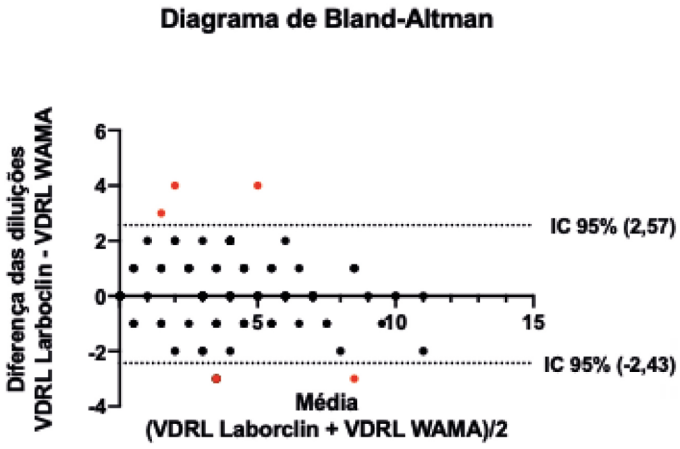
The results show that, regardless of the test used, the frequency of high or low VDRL titration was not statistically significant between tests, validating the statement that both tests had the same performance. It is noteworthy that the objective of this study was not to investigate which of the cases were active or past syphilis, since high titrations can be found in patients who have undergone treatment and thus have a falling titration; in turn, low titrations can be found in several situations, including recent infection, late syphilis, serological scar or false-reactive reaction[3].

Even though the exact agreement of titrations was shown to be weak, the weighted agreement — corrected by the proximity of the results —, was almost complete. This can be statistically explained by the result of the weighted Kappa agreement coefficient (Kw), which gives greater importance to the closer agreement to the VDRL titrations and greater weight when calculating reproducibility[10,11].

The weak agreement.between the exact titrations, that is, in the same titration result in both tests for the same sample, can be explained by the difference in degrees of reactivity in the flocculation tests, this difference being considered within the limits of technical deviation. Furthermore, the presentation of prozone reactions due to the excess of reactive serum component may seem very weak. As a result, positive titrations are totally dependent on material resources and the investigator's judgment[12]. Unfortunately, we did not find similar national studies to compare results.

Corroborating the statistical results, the agreement can be visually verified by the graphic technique proposed by Bland and Altman[13]. This technique consists of a dispersion plot, in which the differences between the paired dilutions (Laborclin – WAMA in ordinates) are plotted (or projected) against their average values ([Laborclin + WAMA]/2, in abscissa). Through this tracing, it is much easier to draw any conclusions about the magnitude of disagreement[14].

Graph 1 shows the bias (how much the differences in dilutions deviate from the zero value) and the standard error (the dispersion of the points of differences in relation to the means), demonstrating the trend of distribution for the range of points within the 95% confidence interval (95%CI -2.43–2.57) and for outliers, above and below the 95%CI.



Graph 1 – Bland-Altman diagram of the differences in VDRL dilutions between the Laborclin and WAMA tests × mean showing the trend of distribution of dilutions between the 95% confidence interval (2.57 to -2.47). The outliers, represented in red, are above and below the 95% confidence interval.

The strong point of this study is the demonstration of agreement with a considerable number of samples. However, the main limitation was the comparison with two operators, not reflecting the real situation where tests are carried out by several operators, which may not be consistent with the reality experienced by those who interpret these results.

This study was approved by the Research Ethics Committee of HUGG/UNIRIO, under CAAE: 66558117.0.0000.5258

CONCLUSION

The VDRL WAMA and Laborclin tests showed the same performance in detecting high and low titrations of non-treponemal antibodies in the paired samples. Although they did not present the same titration value, both tests obtained strong agreement considering close titrations (in up to two dilutions). This fact can be attributed to both the formulation components of each flocculation test and the investigator's judgment. Further reliability and agreement studies are essential between non-treponemal tests available and used in Brazil.

Participation of each author

Isabelle de Carvalho Rangel: main author, review of laboratory results, review of final text. Ricardo de Souza Carvalho: coauthor, attending physician, review of final text. Mauricéa Francisco da Silva Romero Gonzalez: supervision of laboratory procedures at LPC, HUGG; Patrick Menezes Loureço: operator and director of Laborclin VDRL tests; Beatriz Pereira de Azevedo: operator and conductor of the VDRL WAMA tests; Fernando Raphael de Almeida Ferry: supervising professor of authors and review of final text.

Funding

Research developed with its own funding.

Conflict of interests

The authors declare no conflicts of interest.

REFERENCES

1. Altman DG, Bland JM. Measurement in Medicine: The analysis of method comparison studies. J R Stat Soc Ser D Stat. 1983;32(3):307-17. https://doi.org/10.2307/2987937
2. Brasil. Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Vigilância, Prevenção e Controle das Doenças Sexualmente Transmissíveis, Aids e Hepatites Virais. Manual Técnico para Diagnóstico da Sífilis [Internet]. Brasília: Ministério da Saúde; 2016 [accessed on Nov 14, 2020]. Available from: http://www.aids.gov.br/pt-br/pub/2016/manual-tecnico-para-diagnostico-da-sifilis
3. Brasil. Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Doenças de Condições Crônicas e Infecções Sexualmente Transmissíveis. Protocolo Clínico e Diretrizes Terapêuticas para Atenção Integral às Pessoas com Infecções Sexualmente Transmissíveis (IST) [Internet]. Brasília: Ministério da Saúde, 2019 [accessed on Nov 14, 2020]. Available from: http://bvsms.saude.gov.br/bvs/publicacoes/protocolo_clinico_diretrizes_terapeutica_atencao_integral_pessoas_infeccoes_sexualmente_transmissiveis.pdf

4. Chow EP, Wilson DP, Zhang G. Co-infection with HIV and syphilis increases among men who have sex with men in China: a systematic review and meta-analysis. PLoS One. 2011;6(8):e22768. https://doi.org/10.1371/journal.pone.0022768
5. Conger AJ. Kappa and Rater Accuracy: Paradigms and Parameters. Educ Psychol Meas. 2017;77(6):1019-47. https://doi.org/10.1177/0013164416663277
6. Giavarina D. Understanding Bland Altman analysis. Biochem Med (Zagreb). 2015;25(2):141-51. https://doi.org/10.11613%2FBM.2015.015
7. Greenblatt RM, Lukehart SA, Plummer FA, Quinn TC, Critchlow CW, Ashley RL, et al. Genital ulceration as a risk factor for human immunodeficiency virus infection. AIDS. 1988;2(1):47-50. https://doi.org/10.1097/00002030-198802000-00008
8. Ho EL, Tantalo LC, Jones T, Sahi SK, Marra CM. Point-of-care treponemal tests for neurosyphilis diagnosis. Sex Transm Dis. 2015;42(1):48-52. https://dx.doi.org/10.1097%2FOLQ.0000000000000222
9. Larsen SA, Pope V, Johnson RE, Kennedy EJ. A manual of tests for syphilis. Washington, D.C.: APHA; 1998.
10. Dupin N. Syphilis. Rev Méd Interne. 2016;37(11):735-42. https://doi.org/10.1016/j.revmed.2016.05.010
11. Radolf JD, Deka RK, Anand A, Šmajs D, Norgard MV, Yang XF. Treponema pallidum, the syphilis spirochete: making a living as a stealth pathogen. Nat Rev Microbiol. 2016;14(12):744-59. https://doi.org/10.1038%2Fnrmicro.2016.141
12. Sanfilippo AM, Freeman K, Schmitz JL. Comparison of Manual and Fully Automated AIX1000 Rapid Plasma Reagin Assays for Laboratory Diagnosis of Syphilis. J Clin Microbiol. 2018;56(8):e00214-18. https://doi.org/10.1128/JCM.00214-18
13. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;1(8476):307-10.
14. Vogelsang TM, Haaland R. The VDRL tube flocculation test; a microflocculation test for syphilis using cardiolipin-lecithin-cholesterol antigen. Br J Vener Dis. 1951;27(2):52-8. https://doi.org/10.1136/sti.27.2.52

Correspondence address:
ISABELLE DE CARVALHO RANGEL
Laboratório de Pesquisa em Imunologia e Aids, Ambulatório de Imunologia e Aids
Rua Mariz e Barros, 775 – Maracanã
Rio de Janeiro (RJ), Brasil
CEP: 20270-901
Email: isabelle.rangel@ebserh.gov.br / isabellerangel0@gmail.com

Received on: 02.03.2020
Approved on: 14.11.2020