2018

Usefulness of the Captia Syphilis IgG EIA test method and reverse algorithm for detection of syphilis infection in a public health setting

Patricia Armour
Virginia Commonwealth University

Usefulness of the Captia™ Syphilis IgG EIA test method and reverse algorithm for detection of syphilis infection in a public health setting

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

by

Patricia A. Armour, MPA, MT(ASCP)
Bachelor of Science, University of Wisconsin-Superior, 1973
Masters of Public Administration, University of Nevada-Las Vegas, 2011

Director: Teresa Nadder, PhD, MLS(ASCP)^{CM}
Chairman and Associate Professor
Department of Clinical Laboratory Science

Virginia Commonwealth University
Richmond, Virginia
April, 2018

Acknowledgements

I would like to express my great appreciation to my dissertation committee, Dr. Teresa Nadder, Dr. William Korzun, Dr. Jenica Harrison, and Dr. Diane Dodd-McCue for their enthusiastic encouragement and useful critiques of my dissertation work. The advice and guidance given by Dr. Teresa Nadder, my advisor and committee chair, during this five-year journey is gratefully acknowledged.

I would also like to extend my appreciation to the VCU School of Allied Health Professions Doctoral Program administration and faculty. The hybrid distance learning program provided an accessible, high quality learning environment which laid an excellent foundation for completing my dissertation. Special thanks go to Dr. Paula Kupstas, Jeff Lodge, Jerry Clarke, and Monica White for keeping me connected and on task with my degree requirements; to Dr. George Canavos, Dr. Ernie Steidle, and Dr. Henry Carretta for providing a comprehensive statistical methods and data analysis background; to Dr. James Cotter for introducing me to the Donabedian Framework; and to Dr. Diane Dodd-McCue for her positive guidance and for sharing her extensive knowledge of research design, implementation ,and interpretation.

Words cannot express my appreciation for my 2013 cohort colleagues, Jason, Marianne, Allison, and Mercedes. The multiple online study sessions helped me get through the didactic work and I could not have asked for a more supportive cohort to accompany me on this journey.

I gratefully acknowledge my public health colleagues who willingly shared their expertise and knowledge of syphilis infection along with the dedicated public health laboratory staff,

Sharon, Paul, Angela, and Joseph, who diligently entered data into the internal laboratory syphilis QA spreadsheet which eventually became the basis for my dissertation.

I also cannot express enough gratitude for my family who have been so supportive during the past five years. The encouragement of my son Justin, daughter Jennifer, and son-in-law Rick, helped get me through long study hours and the occasional writing block. My grandsons, Brayden and Connor, were also supportive and understanding when I was unable to make it to a game or event because of a course deadline. I hope they gain an understanding that we all should and can be lifelong learners. My sister, Dr. Barbara King, was an inspiration for me to keep going and finish no matter what the obstacle. I would not have been able to accomplish completing this dissertation without my supportive, encouraging, and patient husband, Charlie. He was there when I needed a pat on the back or a push to move forward. I am glad to be finished with this part of the journey and now on to the next challenge.

Table of Contents

List of Tables ……..……...viii

List of Figures…….....xi

List of Abbreviations…….……xii

Abstract……..xv

Chapter 1: Introduction……. ……..……1

Syphilis Diagnosis…….. .5

Research Question and Hypotheses…….....9

Study Design……..…....11

Summary…….....13

Organization of Remaining Chapters……...…..13

Chapter 2: Literature Review……..…….....15

History of Syphilis…….....15

Biology of Syphilis Infection……18

Pathophysiology of Syphilis Infection…….…….....23

History of Syphilis Diagnostic Testing……..…24

Current Syphilis Diagnostic Laboratory Testing……28

Direct detection methods…….…...28

Serological test methods…….….....30

Syphilis Test Interpretation……..…...32

Treponemal Immunoassay Index Values……..….46

Laboratory Test Usefulness……...….50

Contingency table……...53

Bayes' nomogram……..57

Receiver operating characteristic (ROC) plots……..58

Diagnostic test evaluation……..60

Cost analysis……...64

Usefulness measure summary……..……..80

Public Health and Syphilis…….....80

Summary…….…...89

Chapter 3: Methodology…….....92

Theoretical Framework……..92

Research Question and Hypotheses……....96

Research Design…….…....97

Design Validity…….….99

Statistical conclusion validity……99

Internal validity……113

Construct validity……114

External validity……...122

Measurement Reliability and Validity……..123

Example of interdependency of measurement validity and reliability……..124

Method verification……..128

Measurement reliability……...129

Lot to lot variability……..132

EVOLIS instrument reliability…….141

RPR titer reliability……..146

Equivalence……..148

Measurement validity……...…154

Reliability and validity summary…….156

Population and Sample Size…….157

Statistical Power Analysis……..……..162

Variables……..167

Data Collection Procedures……..180

Human Subjects…….……..182

Data Analysis……..…….183

Summary…….…….187

Chapter 4: Results……...…….192

Descriptive Variables……...192

Observed Variables……..196

First null hypothesis…….196

Second null hypothesis…….202

Third null hypothesis……209

Summary……..213

Chapter 5: Discussion……...216

Results……..217

Public health population……..……217

Usefulness measures……222

Limitations……...234

Conclusions and Future Studies……...237

List of References……244

Vita...……260

List of Tables

1. Reported Syphilis Cases (Total and by Stage), United States, 1990-2016……3

2. History of Syphilis Test Development……...25

3. Syphilis Test Characteristics……..……29

4. Sensitivity and Specificity of Selected Syphilis Test Methods by Disease Stage……35

5. Likelihood Ratio (LR) Interpretation……...…..53

6. Diagnostic Test Measures: Calculation of Sensitivity, Specificity, Predictive Values, Likelihood Ratios, Accuracy, and Prevalence……...54

7. Impact on Predictive Value and Likelihood Ratio at 1% Disease Prevalence for a Test with 99% Sensitivity and 95% Specificity……...56

8. Impact on Predictive Value and Likelihood Ratio at 10% Disease Prevalence for a Test with 99% Sensitivity and 95% Specificity……..57

9. Attributes for Determining Preferable Statistical Measure of a Diagnostic Test……63

10. Microcosting: MacroVue RPR (Manual Method)…….....74

11. Microcosting: Captia™ Syphilis IgG (EVOLIS Analyzer)……75

12. Microcosting: Fujirebio Serodia TP-PA (Manual Method)……76

13. Comparison of MacroVue RPR and Captia™ Syphilis IgG Costs, Efficiency, Time…...78

14. CDC Syphilis Case Definition by Stage (2012)……...115

15. Statistical Measures for Four Creatinine Assays Performed on the Same Instrument….125

16. Captia™ Syphilis IgG Quality Control Results January 2012 to December 2013……...130

17. Serodia TP-PA Quality Control Results December 2011 to January 2014……..130

18. MacroVue RPR Quality Control Results December 2011 to January 2014…….131

19. Captia™ Syphilis IgG Lot to Lot External Control OD Variation Assessment……...138

20. EVOLIS Analyzer Service Record January 2012 to December 2013……..143

21. Virotrol RPR Panel Control Lot-to-Lot Variation Assessment…….149

22. Nonmatching Study Variables Identified During Electronic Chart Review……....153

23. Study Exclusion Category, Number and Rationale……...162

24. Study Variables, Definition, Type, and Operationalized Outcome…….....169

25. Equivocal Syphilis IgG Test Result Frequency and Diagnosis/Stage Frequency..……..170

26. Invalid Syphilis IgG Test Result Analysis Dates and Clinical Data……..172

27. Inconclusive TP-PA Test Result Analysis Dates and Clinical Data……..174

28. Traditional Algorithm Interpretation Contingency Table Measurement Outcomes…….176

29. Reverse Algorithm Interpretation Contingency Table Measurement Outcomes……..177

30. Syphilis IgG and TP-PA Contingency Table Measurement Outcomes……180

31. Study Hypotheses, Theoretical Domains, Variables, and Analyses……..185

32. Gender Frequency Count and Percentage…….....192

33. Age Group Frequency Count and Percentage…….…….193

34. Reason for Clinic Visit Frequency Count and Percentage……194

35. Infection DiaTanle 39gnosis Frequency Count and Percentage……...195

36. Syphilis Stage Frequency Count and Percentage…….…….196

37. Captia™ Syphilis IgG Contingency Table (Equivocal Counted as Reactive)……..197

38. Captia™ Syphilis IgG Contingency Table (Equivocal Counted as Nonreactive)…….197

39. McNemar Two-sided Hypothesis Test of the Difference Between Captia™ Syphilis IgG Equivocal Counted as Reactive or Nonreactive……..198

40. Fujirebio Serodia TP-PA Contingency Table……..….199

41. Captia™ Syphilis IgG and Serodia TP-PA Sensitivity, Specificity, 95% CI……199

42. Captia™ Syphilis IgG and Serodia TP-PA PPV, NPV, 95% CI, Adjusted PPV and NPV……...200

43. Captia™ Syphilis IgG and Serodia TP-PA Likelihood Ratios, 95% CI, Accuracy, and Study Prevalence……..……201

44. McNemar Test Captia™ Syphilis IgG and Serodia TP-PA Frequency Table…….…201

45. McNemar Two-sided Hypothesis Test of the Difference Between Captia™ Syphilis IgG and Serodia TP-PA…….....202

46. Hypothesis Test for AUC Comparison of the Ability of Syphilis IgG S/CO Value and RPR Titer Value to Detect Syphilis Infection……203

47. Hypothesis Test Comparing Two AUCs (Empirical Estimation) for Syphilis IgG S/CO Value and RPR Titer value……..204

48. Captia™ Syphilis IgG S/CO Cutoff Values, TPR, TNR, PPV, Accuracy, Youden Index, and Distance to Corner…….207

49. RPR Titer Cutoff Values, TPR, TNR, PPV, Accuracy, Youden Index, and Distance to Corner……....208

50. Traditional Algorithm Contingency Table……...209

51. Reverse Algorithm Contingency Table……....210

52. Traditional and Reverse Algorithm Sensitivity, Specificity, 95% CI……..….210

53. Traditional and Reverse Algorithm PPV, NPV, 95% CI, Adjusted PPV and NPV…….210

54. Traditional and Reverse Algorithm Likelihood Ratios, 95% CI, Accuracy, Study Prevalence……..212

55. McNemar Test Traditional and Reverse Algorithm Frequency Table……..…212

56. McNemar Two-sided Hypothesis Test of the Difference Between Traditional and Reverse Algorithms……..213

57. Summary Null Hypothesis Statistical Tests……..214

58. Summary Usefulness Statistics for Syphilis IgG, TP-PA, and Algorithms……..214

59. Syphilis IgG S/CO Frequency Counts, FN, FP, Discordant, and TP-PA Confirmation…227

60. Captia™ Syphilis IgG S/CO 0.8 Cutoff False Negative Diagnostic Characteristics…...229

List of Figures

1. Reported Cases of Syphilis (Total and by Stage), United States, 1990-2016……1

2. Schematic Representation of the Course of Untreated Syphilis in an Immunocompetent Person…….....19

3. Syphilis Staging and Antibody Response…….…..34

4. Traditional Syphilis Test Algorithm and Laboratory Interpretation……...……36

5. Reverse Sequence Syphilis Test Algorithm and Laboratory Interpretation……..….37

6. Example Bayes' Nomogram with 30% Pretest Probability and +LR 19.8…….…59

7. Schematic Model for the Appropriate Selection and Use of Laboratory Tests……65

8. Comparison of Donabedian Quality Framework Model and Clinical Laboratory Path of Workflow……...94

9. Western Region Public Health Laboratory Captia™ Syphilis IgG Repeat Algorithm…..105

10. Western Region Public Health Laboratory RPR QC Documentation Form……..108

11. Western Region Public Health Laboratory TP-PA Day of Use Documentation Form…111

12. Graph of Validity and Reliability Results of Four Creatinine Assays Performed on the Same Instrument…….126

13. Western Region Public Health Laboratory Qualitative RPR Day of Use Worksheet……150

14. Western Region Public Health Laboratory Quantitative RPR Day of Use Worksheet…..151

15. Graphical Comparison of Captia™ Syphilis IgG and RPR Titer ROC Curves……205

List of Abbreviations

AACC: American Association of Clinical Chemists
ACA: Affordable Care Act
AI: Antibody Index
AIDS: Acquired immune deficiency syndrome
APHL: Association of Public Health Laboratories
ASCLS: American Society for Clinical Laboratory Science
ASCP: American Society of Clinical Pathologists
AUC: area under the curve
CAP: College of American Pathologists
CD4: Cluster of differentiation 4
CDC: Centers for Disease Control and Prevention
CFR: Code of Federal Regulations
CI: confidence interval
CIA: chemiluminescent immunoassay
CLIA: Clinical Laboratory Improvement Amendments
CLSI: Clinical and Laboratory Standards Institute
CMS: Centers for Medicare and Medicaid Services
CNS: Central Nervous System
CPRR: cost per reportable result
CS: congenital syphilis
CSF: cerebrospinal fluid
CSTE: Council of State and Territorial Epidemiologists
CV: coefficient of variation
DAT: direct access testing
DFA-TP: direct fluorescent antibody-*T. pallidum*
DNA: deoxyribonucleic acid
ECDC: European Centers for Disease Prevention and Control
EIA: enzyme immunoassay
ELISA: enzyme linked immunosorbent assay
EQ: equivocal
FDA: Food and Drug Administration
FN: false negative
FP: false positive
FTA: fluorescent treponemal antibody
FTA-ABS: fluorescent treponemal antibody absorption
FTC: Federal Trade Commission
GUD: genital ulcer disease

HAART: highly active antiretroviral therapy
HHS: Health and Human Services
HIPAA: Health Insurance Portability and Accountability Act
HIV: Human immunodeficiency virus
HMAs: High Morbidity Areas
HPV: human papilloma virus
HTLV: human T-lymphotrophic virus
HTR: high titre reactive
Ig: immunoglobulin
IOM: Institute of Medicine
IRB: Institutional Review Board
IT: information technology
IVD: in vitro diagnostic device
JICA: Japan International Cooperation Agency
LGBT: Lesbian, Gay, Bisexual, and Transgender
LHD: local health department
LIMS: Laboratory Information Management System
LR: likelihood ratio
LTR: low titre reactive
MBIA: microbead immunoassay
MHA-TP: micro-hemagglutination assay – *T. pallidum*
MLO: Medical Laboratory Observer
MSM: men who have sex with men
MSW: men who have sex with women
NAAT: nucleic acid amplification test
NASEM: National Academies of Sciences, Engineering and Medicine
NCSS: Number Cruncher Statistical Software
NGU: nongonococcal urethritis
NPV: negative predictive value
NQF: National Quality Forum
N: nonreactive
NR: nonreactive
OB/GYN: obstetrical/gynecological
OD: optical density
OR: odds ratio
PASS: Power Analysis and Sample Size
P&S: primary and secondary
PCR: polymerase chain reaction
PHI: protected health information
PHL: public health laboratory
PHLT: Public Health Laboratory - Toronto
PM: preventive maintenance
POC: Point of Care
PPV: positive predictive value
QA: quality assessment
QC: quality control

QMS: Quality Management System
R: reactive
ROC: receiver operating characteristic
RPR: rapid plasma reagin
RR: relative risk
S/CO: signal to cutoff
SD: standard deviation
SEE: Syphilis Elimination Effort
SLE: systemic lupus erythromatosus
SOP: standard operating procedure
SPSS: Statistical Package for Social Science
SSuN: Sexually transmitted Surveillance Network
STD: sexually transmitted disease
STI: sexually transmitted infection
TN: true negative
TNR: true negative rate
TP: true positive
TPHA: *T. pallidum* hemagglutination assay
TPI: *T. pallidum* immobilization
TP-PA: *T. pallidum* passive particle agglutination
TPR: true positive rate
TRUST: toluidine red unheated serum test
US: United States
USR: unheated serum reagin
VCU: Virginia Commonwealth University
VDRL: Venereal Disease Research Laboratory
WSLH: Wisconsin State Laboratory of Hygiene
WSW: women who have sex with women

Abstract

USEFULNESS OF THE CAPTIA™ SYPHILIS IGG EIA TEST METHOD AND REVERSE ALGORITHM FOR DETECTION OF SYPHILIS INFECTION IN A PUBLIC HEALTH SETTING

By Patricia A. Armour, MPA, MT(ASCP)

A dissertation submitted in partial fulfillment of the requirements for a degree of Doctor of Philosophy at Virginia Commonwealth University

Virginia Commonwealth University, 2018

Major Director: Teresa Nadder, PhD, MLS(ASCP)^{CM} Chairman and Associate Professor, Department of Clinical Laboratory Science

Syphilis, a systemic sexually transmitted disease, is on the rise in the US, with infection rates the highest recorded since 1994 according to the CDC. Useful laboratory testing is an important diagnostic tool for determining individual syphilis infection and preventing community-wide disease spread.

The purpose of this study was to determine the usefulness of a specific automated treponemal test method, the Captia™ Syphilis IgG EIA, and the syphilis reverse algorithm interpretation for detecting syphilis infection among patients seeking care in a public health clinic. The study employed a retrospective, nonexperimental descriptive correlational design with data collected between 2012-2013 from 4,077 public health clinic patients with 21% of the patients diagnosed with syphilis infection.

There was a statistically significant difference between the Captia™ Syphilis IgG and the Fujirebio Serodia TP-PA test results; between the Captia™ Syphilis IgG Signal to Cutoff (S/CO) and the MacroVue RPR titer continuous variables; and between the reverse and traditional syphilis interpretation algorithms. The reverse algorithm using the Captia™ Syphilis IgG test method provided more useful performance measures with a sensitivity of 82%; specificity of 99%; accuracy of 95%; positive likelihood ratio of 63.06 and negative likelihood of 0.18 than the traditional algorithm using the MacroVue RPR test method. Statistical comparison of the area under the curve (AUC) for the continuous variables, Captia™ Syphilis IgG S/CO and RPR titer, concluded that the Syphilis IgG AUC (0.9500) was higher than the RPR titer (0.8155) indicating greater accuracy for detecting syphilis infection.

This was the first study to determine that the Captia™ Syphilis IgG, the S/CO value, and reverse algorithm are useful diagnostic predictors of syphilis infection among public health clinic patients. The data from this study can be utilized by future researchers and scientists who are developing or improving syphilis detection methods.

Chapter 1: Introduction

Syphilis, a systemic sexually transmitted disease (STD), is on the rise in the United States (US). According to the Centers for Disease Control and Prevention (CDC) National Overview of Sexually Transmitted Diseases report (2017), the total case counts and rates of syphilis infection are the highest recorded since 1994. Reported syphilis cases declined significantly from the early 1990s through 2005, when cases began steadily climbing (Figure 1).

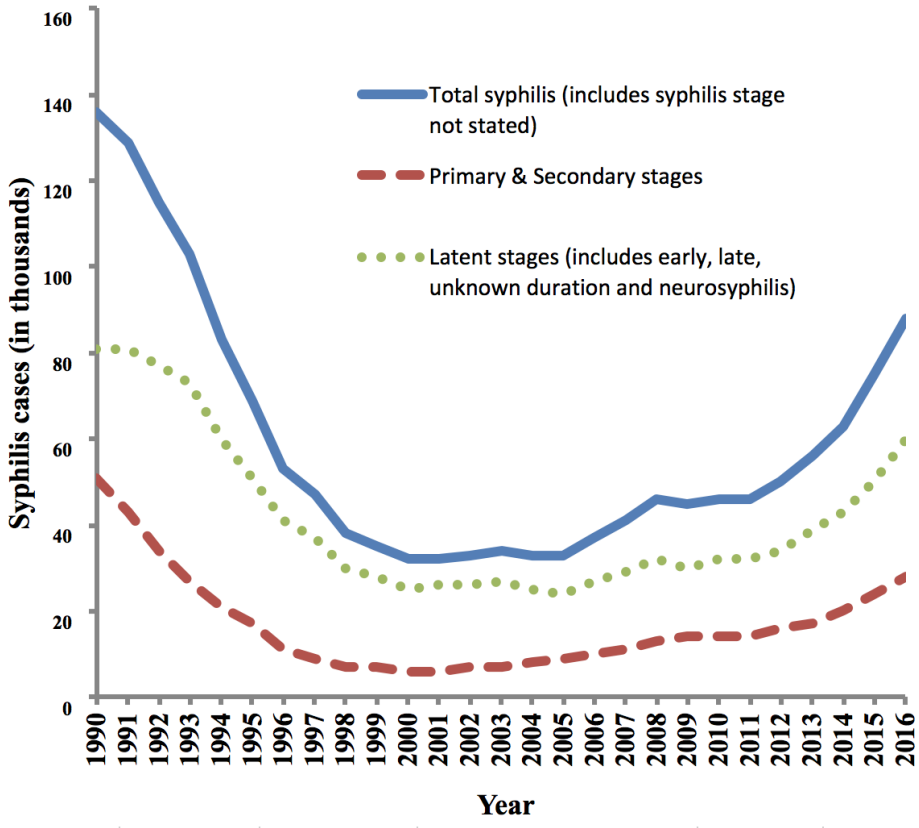

Figure 1. Reported Cases of Syphilis (Total and by Stage), United States, 1990-2016. Adapted from "Sexually Transmitted Disease Surveillance 2016", p 86, by Centers for Disease Control and Prevention, 2017.

The total number of syphilis cases reported in 2016 (88,042) represents a 17.8% increase from 2015 (74,702). The rate of primary and secondary (P&S) syphilis, the most infectious stages, has also been rising since 2000 with over a four-fold increase from 5,979 cases in 2000 to 27,814 cases in 2016 (Table 1). Increased P&S cases have been identified among gay, bisexual, and other men who have sex with men (collectively referred to as MSM), which is of public health significance due to the increased likelihood of acquiring and transmitting human immunodeficiency virus (HIV) if the person is infected with syphilis (Patton, Xu, Nelson & Weinstock, 2014). The highest P&S syphilis rates in 2016 were among men aged 25-29 years (48.5/100,000 population), among men in the Western region (12.6/100,000 population), and among black men (23.1/100,000 population (CDC, 2017).

Syphilis is a multi-system disease caused by the bacterium, *Treponema pallidum* subsp. *pallidum*, herein after referred to as *T. pallidum*. The disease progresses in overlapping stages; and if left untreated, can be associated with significant complications. The initial infection begins with exposure to an infectious syphilitic lesion, which may be in a genital (e.g. cervix, penis) or extragenital (e.g., rectum, oral cavity) location (LaFond & Lukehart, 2006). Once infected, the individual progresses through primary, secondary, and early latent syphilis stages, which are collectively referred to as early syphilis. Symptoms that develop during these stages are nonspecific and typically resolve spontaneously without antibiotic treatment. However, without treatment, the *T. pallidum* organisms will migrate throughout the body and infect other organs (e.g., brain, heart, or eye). An individual is considered to be infectious during the early syphilis stages (infection of one year or less). Once the disease progresses to the late syphilis stage (infection greater than 1 year), the individual is no longer considered to be infectious; however, the long-term complications of untreated syphilis infection may not appear for 10-20

Table 1

Reported Syphilis Cases (Total and by Stage), United States, 1990-2016

Year	Total syphilis[a]	Primary and Secondary	Latent[b]
1990	135,590	50,578	81,147
1991	128,719	42,950	81,345
1992	114,730	34,009	76,654
1993	102,612	26,527	72,665
1994	82,713	20,641	59,620
1995	69,359	16,543	50,953
1996	53,240	11,405	40,553
1997	46,716	8,556	37,078
1998	38,289	7,007	30,439
1999	35,386	6,617	28,189
2000	31,618	5,979	25,059
2001	32,286	6,103	25,677
2002	32,919	6,862	25,597
2003	34,289	7,177	26,680
2004	33,423	7,980	25,068
2005	33,288	8,724	24,225
2006	36,958	9,756	26,830
2007	40,925	11,466	29,024
2008	46,292	13,500	32,346
2009	44,832	13,997	30,404
2010	45,844	13,774	31,683
2011	46,040	13,970	31,712
2012	49,915	15,667	33,914
2013	56,482	17,375	38,748
2014	63,450	19,999	42,993
2015	74,702	23,872	50,343
2016	88,042	27,814	59,600

[a]Total includes stage of syphilis not stated. [b]Latent includes early, late, unknown duration and neurosyphilis stages. Note: Adapted from "Sexually Transmitted Disease Surveillance 2016", p 86, by Centers for Disease Control and Prevention, 2017.

years from the initial infection. These complications can involve the heart (cardiovascular syphilis) and central nervous system (neurosyphilis) (Larsen, Pope, Johnson & Kennedy,1998). More recently, Woolston et al. (2015) reported a cluster of ocular syphilis cases, mainly in

MSM, occurring in Seattle and San Francisco within a short time period from December 2014 to March 2015. This unusual occurrence led CDC to release a clinical advisory notifying healthcare providers of an increase in ocular syphilis, especially among the MSM population. Ocular syphilis can lead to blindness or could be a symptom of disease progression to the neurosyphilis stage (CDC, 2015).

An infected woman can transmit *T. pallidum* organisms to the fetus at any stage (early or late) of infection, thus leading to congenital syphilis (CS). Miscarriage, stillbirth and early infant death can result from the overwhelming fetal infection. If the child survives, there are other malformations of teeth, joints, and legs that may occur along with deafness (Genc & Ledger, 2000; Larsen et al., 1998). According to CDC surveillance data (2017), the CS rate in the US had been declining with a low of 8.4 cases per 100,000 live births in 2012. However, there was a rapid increase (86.9%) between 2012-2016 with a rate of 15.7 cases per 100,000 live births reported nationally in 2016. The 2016 syphilis rate represents a 27.6% increase relative to 2015 (12.3 cases/100,000 live births) and included 41 cases of syphilitic stillbirths. The highest reported CS cases in 2016 were in the West (25.6 cases/100,000 live births). The increase in CS infections parallels the increase in syphilis infection in women. Syphilis is a treatable disease and CDC considers a case of CS to be a sentinel public health event that reflects a breakdown in prevention activities within public health and health care systems. The major opportunities to prevent syphilis transmission during pregnancy are primary prevention of syphilis infection among reproductive age women and men who have sex with women (MSW), and prevention of fetal transmission among pregnant women who already have syphilis infection. Accurate laboratory testing for syphilis infection is a critical component of these prevention efforts (Bowen, Su, Torrone, Kidd & Weinstock, 2015).

Syphilis Diagnosis

Syphilis is a multi-system disease that progresses in overlapping stages, and the infected person may be asymptomatic during the later stages. Diagnosis requires use of multiple laboratory tests along with an interpretation algorithm combined with clinical evaluation to determine if an individual is currently infected or was previously infected.

Infection with *T. pallidum* induces an antibody response with Immunoglobulin (Ig) M detectable within two weeks and IgG within four weeks of infection. IgM antibodies persist for approximately 18 months after successful treatment, and IgG antibodies usually persist for life. The antibodies do not provide immunity, as a person can be re-infected with syphilis even after successful antibiotic treatment (Cortez & Greenwald, 2014).

Diagnostic laboratory tests for syphilis include either direct identification of *T. pallidum* or serologic tests for antibodies. Darkfield microscopy or molecular testing of the exudate from the suspected lesion can be used to perform direct identification of *T. pallidum*. Serologic tests include treponemal and nontreponemal antibody assays. Treponemal serologic tests detect antibodies (IgG and/or IgM) to *T. pallidum*. Because IgG antibodies may persist for life even after successful treatment, treponemal diagnostic tests cannot distinguish between active or previously treated syphilis infections. A second nontreponemal test must be performed to differentiate an active syphilis infection. Treponemal tests are highly sensitive and specific for syphilis infection. Tests available in the US include enzyme immunoassays (EIA), chemiluminescent immunoassays (CIA), microbead immunoassays (MBIA), *T. pallidum* passive particle agglutination (TP-PA) and fluorescent treponemal antibody absorption (FTA-ABS). Testing can be performed either manually or using an automated analyzer. Nontreponemal tests detect antibodies directed against lipoidal antigens that are released when cells are infected with

T. pallidum, an indirect measurement of syphilis infection. Nontreponemal antibodies usually disappear within a year or two after successful antibiotic treatment. Nontreponemal tests are highly sensitive but have lower specificity for syphilis infection (more false positives); therefore, a treponemal test must be performed to confirm the syphilis infection. Nontreponemal tests available in the US include Venereal Disease Research Laboratory (VDRL), rapid plasma reagin (RPR), and toluidine red unheated serum test (TRUST). These tests are usually performed manually and are labor intensive (Association of Public Health Laboratories [APHL], 2009; CDC, 2011; Larsen et al., 1998).

There are a number of limitations with the syphilis serologic assays. False positive results can occur with either treponemal or nontreponemal tests. Nontreponemal tests detect antibodies against lipoidal antigens, which can be elevated in conditions that are unrelated to syphilis infection. Conditions that could cause false positive nontreponemal test results include autoantibodies (lupus), viral infections (HIV, hepatitis C virus), bacterial infections (pneumonia, endocarditis), parasitic infections (malaria), and noninfectious causes (stroke, malignancy). Treponemal tests can cross-react with other treponemal diseases such as yaws or pinta. Other causes of false positive treponemal tests include autoimmune disease, diabetes, alcoholic cirrhosis, viral infections, pregnancy or cryoglobulinemia. False negative results can occur with either treponemal or nontreponemal tests. Nontreponemal flocculation tests are subject to the prozone effect, in which high titers of lipoidal antibody cause a false negative result by overwhelming the test antigen sites. Persons with latent syphilis, especially pregnant women with HIV, may have false negative nontreponemal tests. Treponemal tests may be falsely negative if the level of antibody is below the detectable limit of the assay. Co-infection with

HIV may cause false negative results with either treponemal or nontreponemal tests (Cortez & Greenwald, 2014).

Laboratory diagnosis of syphilis infection is challenging because current test methodologies are serologically based and depend on detection of antibodies. Syphilis antibody levels can vary depending on the disease stage; therefore, the CDC and APHL recommend a multi-tier or cascade approach to syphilis testing using either the traditional or reverse algorithm. The traditional algorithm starts with a nontreponemal test, such as the RPR, with reactive tests reflexed to a treponemal test, such as the Fujirebio Serodia TP-PA for confirmation. The reverse algorithm starts with a treponemal test, such as the Captia™ Syphilis IgG EIA, with reactive tests reflexed to a nontreponemal test, such as the RPR. If the RPR is nonreactive, then testing with a second, different treponemal test, such as the TP-PA, is necessary to resolve the discordance (APHL, 2009; CDC, 2011).

Because the treponemal EIA, CIA and MBIA tests can be automated, many clinical and some public health laboratories are offering these tests as the initial syphilis screening test, thus, employing the reverse algorithm. Automated testing platforms provide significant cost savings in laboratory staff time over the manual nontreponemal test methods. The decision to implement a new test methodology, such as an automated syphilis test, is based on several considerations such as cost, equipment, training, assay sensitivity and specificity. In addition, the laboratorian must know the prevalence of syphilis within the test population because it will impact the predictive value of the test (Zanto, 2010).

Syphilis is a complicated disease to diagnose because of its ability to mimic other diseases and the overlapping stages of disease progression. While there are multiple different serologic testing platforms available, there is variability in the sensitivity and specificity of each assay.

This variability is particularly important when reverse algorithm screening is utilized. If a highly sensitive treponemal test is used for initial screening, then CDC recommends that an equal or higher sensitivity test should be used for the second treponemal test in the event of discordant results with the nontreponemal test. If a lower sensitivity test is used and the second treponemal test is negative, there is a potential for assumption that the initial test was a false positive, which could have deleterious effects on the patient's health as well as allowing the spread of disease within the community (Binnicker, Jespersen & Rollins, 2012).

The semi-quantitative signal to cutoff (S/CO) or index value generated by automated treponemal assays may be useful for resolving discordant results with reverse algorithm testing. The S/CO value is proportional to the amount of *T. pallidum* antibody present in the infected person's serum. When there are discordant results between the initial treponemal screening and the nontreponemal reflex test, a high S/CO value could be predictive of a positive second treponemal test. If a standard cut off S/CO value for a "true positive" treponemal test could be identified for the automated initial test, then the second confirmatory treponemal test may not have to be performed. This could result in cost savings for the testing laboratory and expedite turnaround time. Currently, reporting the S/CO value is not recommended by syphilis test manufacturers and is not listed in the package insert approved by the US Food and Drug Administration (FDA). The CDC suggested in the 2015 Syphilis treatment guidelines that the usefulness of these semi-quantitative values should be investigated further (CDC, 2015).

Syphilis tests can be performed in any accredited clinical laboratory in the US; however, public health laboratories provide a unique testing environment because they are usually associated with a public health STD clinic. Individuals who are underserved in the traditional healthcare system predominantly access these clinics. STD clinic populations typically include

racial/ethnic minorities, young persons, MSM and women who have sex with women (WSW). These populations often have high STD rates. The prevalence of disease in the population tested has an impact on the predictive value of a diagnostic test. As prevalence increases, so does the positive predictive value of the diagnostic test, thus STD clinics would provide sufficient numbers of infected persons for determining the usefulness of syphilis tests (Galen & Gambino, 1975). STD clinics are particularly important for men who are less likely than women to seek preventive care. Patients prefer to access STD clinics rather than primary care due to confidentiality, expertise and convenience. These clinics serve a public health function of identifying index patients and preventing further spread of disease through contact investigation (Cellum et al., 1997; Pathela et al., 2015). Public health laboratories (PHLs) operate as diagnostic and reference labs as well as provide subject matter expertise to their public health agency partners. As diagnostic laboratories, PHLs ensure that testing is performed following the latest CDC disease guidelines and they communicate on a regular basis with agency partners to ensure that their laboratory services meet STD prevention and surveillance needs. As a reference laboratory, PHLs provide confirmatory testing and act as the conduit to CDC specialized laboratories for supplemental testing. PHLs also serve as subject matter experts for both clinical and private laboratory partners as well as public health agencies. PHLs are qualified to act as applied research centers and provide assistance to their agency partners for evaluation of the effectiveness of new test methods, such as direct syphilis testing using molecular methods (APHL, 2011; APHL 2017).

Research Question and Hypotheses

With the increase in syphilis infection in the US, it is imperative that laboratories utilize the most useful syphilis test methodology available. Providing the correct answer regarding syphilis

infection ensures treatment can be given early to prevent further disease spread or long-term health complications. Many laboratories in the US are utilizing automated testing platforms for syphilis testing in order to increase efficiency; and additional research is needed to provide data to laboratorians making decisions about test methods to use in their facilities.

There are gaps in the current body of knowledge regarding the accuracy and reproducibility of a specific syphilis EIA test, the Captia™ Syphilis IgG, for detecting syphilis infection. The Captia™ Syphilis IgG EIA method was one of the first EIA syphilis tests approved by the FDA in 2001. Though the sensitivity and specificity of the test have been reported, there is little research regarding the predictive value of the Captia™ Syphilis IgG test when utilized in the reverse algorithm for diagnosis (Halling et al., 1999; Lefevre, Bertrand & Baurlaud, 1990; Reisner, Mann, Tholeken, Waite & Woods, 1997; Siletti, 1995; Young, Moyes, McMillan & Patterson,1992; Young, Moyes & Ross, 1995). Additionally, a review of current literature did not find studies on the predictive value of the Captia™ Syphilis IgG S/CO value for detecting syphilis infection (Loeffelholz & Binnicker, 2012; Wong et al., 2011; Yen-Liberman, Daniel, Means, Waletzky & Daly, 2011). Most of the research on the usefulness of the reverse and traditional test algorithms has been performed either on stored serum samples or in the clinical setting. Further, there is limited research comparing the two algorithms in a public health setting with patients who may be at increased risk for syphilis infection. This study would be the first to evaluate the reverse algorithm test combination of Captia™ Syphilis IgG, RPR, and TP-PA for its usefulness in the diagnosis of syphilis. It would also be the first analysis of the usefulness of the Captia™ Syphilis IgG EIA assay S/CO value for detection of syphilis infection.

The purpose of this study was to determine the usefulness of a specific automated test method, the Captia™ Syphilis IgG EIA, and the syphilis reverse algorithm interpretation for

detecting syphilis infection among patients seeking care at public health clinics. The target construct was syphilis infection and the proposition for the research question is that more useful syphilis testing will improve syphilis detection. The process (syphilis test results and S/CO value) and outcome (traditional and reverse algorithm interpretations) domains of the Donabedian Quality Framework provide the theoretical framework of the study. The specific research question and hypotheses to be addressed in the study are:

RQ_1: What is the usefulness of the Captia™ Syphilis IgG EIA test method and the reverse algorithm for detection of syphilis infection in a public health population?

$H1_o$: Among patients seeking STD services at a large western region metropolitan public health clinic, there will be no difference between the Captia™ Syphilis IgG EIA and Fujirebio Serodia TP-PA test results.

$H2_o$: Among patients seeking STD services at a large western region metropolitan public health clinic, there will be no difference in diagnostic interpretation of the Captia™ Syphilis IgG EIA S/CO value and the Becton Dickinson MacroVue RPR titer test result.

$H3_o$: Among patients seeking STD services at a large western region metropolitan public health clinic, there will be no difference between the syphilis traditional and reverse algorithm interpretations.

Study Design

The study employed a retrospective, nonexperimental descriptive correlational design to address whether the Captia™ Syphilis IgG EIA test and the syphilis reverse algorithm provide useful diagnostic information about syphilis infection. All data used for the study was collected

as purposive convenience samples during two calendar years from January 1, 2012 through December 31, 2013. The study data was collected as part of standard public health practice within the STD clinic at a large western region metropolitan (greater than two million population) public health clinic and included syphilis test results from patients with no syphilis disease as well as those in early (primary, secondary, early latent), and late latent stages of disease.

Inclusion criteria for the study are males and females 18 years and older who were seen in the public health STD clinic, met the STD clinic criteria for syphilis testing and had all three syphilis tests (Syphilis IgG, RPR and TP-PA) performed with clinically diagnostic test results on each sample collected. Exclusion criteria included males and females under 18 years of age, those who did not visit the STD clinic, those that did not have three syphilis tests performed, samples that tested invalid by the Syphilis IgG test and samples that tested inconclusive by the TP-PA test method.

The study assumptions included the following: a) the study population was representative of patients seeking STD services at the public health clinic, b) all patient-related variables and laboratory outcome data were correctly entered into the study data spreadsheets, and c) all clinical samples were collected and handled following standard laboratory procedures.

The study variables included both descriptive and observed laboratory test variables. Descriptive variables included gender, age, reason for clinic visit, infection diagnosis, and syphilis stage. Observed variables included the Captia^{TM} Syphilis IgG test result and S/CO value, Becton Dickinson MacroVue RPR qualitative test result and titer value, Fujirebio Serodia TP-PA test result, syphilis traditional and reverse algorithm serologic laboratory interpretation, and clinical diagnosis of syphilis based on chart review.

Summary

Syphilis infection is on the rise in the US and there can be serious health implications for the men, women and babies that become infected. Accurate diagnosis of syphilis infection is necessary to ensure that treatment is started as early as possible. Current syphilis diagnostic tests are based on detection of antibodies in the infected person's blood. There is variability in the sensitivity and specificity of each of the current FDA approved test methods. New instrument testing platforms that allow for automation of treponemal syphilis testing provide an efficient, cost-saving mechanism for initial syphilis screening. This change to initial treponemal testing required establishment of a reverse syphilis test algorithm. There are limited studies on the usefulness of specific combinations of diagnostic serologic tests within the reverse algorithm. This study expanded the body of knowledge regarding syphilis testing by utilizing an automated test method, the Captia™ Syphilis IgG EIA, for initial syphilis screening followed by RPR and TP-PA within a reverse testing algorithm. This study was also the first to determine the usefulness of the Captia™ Syphilis IgG S/CO value for predicting syphilis confirmatory testing. The high STD rates among individuals seeking care at public health clinics and the diversity of the population provided a unique setting for the study.

Organization of Remaining Chapters

The study is presented in five chapters followed by a list of references. Chapter Two provides a review of the literature concerning the history of syphilis infection, laboratory diagnostic testing, description of laboratory test usefulness criteria and literature gaps. Chapter Three describes the study design and methodology including framework, selection of variables, data collection, validity, reliability, human subjects protocol and chapter summary. Chapter Four

presents the study results. Chapter Five concludes with discussion of the data analysis, limitations, conclusions and recommendations for potential future studies.

Chapter 2: Literature Review

The following chapter will present the background information necessary for a discussion of how to determine the usefulness of the Captia™ Syphilis IgG EIA test method for the detection of syphilis infection in a public health setting. General information regarding the history of syphilis infection, characteristics of *Treponema pallidum*, syphilis disease stages, syphilis diagnostic tests, laboratory test usefulness criteria, public health implications of syphilis infection, and clinical interpretation of diagnostic test results based on traditional and reverse algorithms will be provided. The accuracy of current syphilis test methodologies will also be discussed along with issues relating to resolving discordant results between test methods. Finally, a summary will be provided regarding gaps in literature related to the use of the Captia™ Syphilis IgG test method in a public health setting.

History of Syphilis

An ancient disease, syphilis was the cause of a major pandemic crossing international boundaries and affecting a multitude of people in the later years of fifteenth century Europe. The pandemic began in Europe in 1495, spread to India in 1498 and then entered China in 1505. This pandemic provided the first documented cases of syphilis infection. There are three theories of the origin of the syphilis pandemic – the New World or Columbian theory, the Old World or pre-Columbian theory, and the Unitarian theory (LaFond & Lukehart, 2006; Singh & Romanowski, 1999).

The Columbian theory dates the beginning of the syphilis pandemic to 1493 following the return of Christopher Columbus, his crew, and captured slaves from voyages to the New World (Haiti, West Indies and the Americas). This theory proposes that syphilis was endemic among the New World natives. The disease had developed into a mild form, which, when introduced into the naïve, susceptible Old World (Europe) population became aggressively virulent resulting in a worldwide pandemic. Anthropological findings on exhumed human bones have shown indications of venereal syphilis in New World bones but not in pre-Columbian Old World bones, thus lending support for the Columbian theory (Armelagos, Zuckerman, & Harper, 2012; Crosby, 1969; Rothschild, 2005; Singh & Romanowski, 1999).

The pre-Columbian theory proposes that syphilis was endemic in the Old World but was not recognized until the end of the 15th century when the organism mutated to a more virulent strain, and the pandemic began. This theory is disputed by testing of archeological remains that do not exhibit indicators of syphilis lesions in Old World exhumed bodies (Armelagos et al., 2012; Tampa, Sarbu, Matei, Venea, & Georgescu, 2014).

The Unitarian theory proposes that venereal syphilis is the same disease as the nonvenereal yaws, bejel and pinta treponemal illnesses. Each of these venereal and nonvenereal diseases is caused by a different *T. pallidum* subspecies. Crosby (1969) explains that the Unitarian theory is based on the assumption that treponematosis or infection with any *T. pallidum* subspecies is all the same disease but symptoms are manifested differently based on culture, climate, and hygiene. The treponemal organism only infects humans and cannot survive outside the body for more than a few minutes. Transmission and survival depends on its ability to adapt to different climates and human interactions. According to this theory, treponemal organisms first infected humans in Africa thousands of years ago, where the moist climate allowed the organism to survive on

human skin without causing significant damage. This disease manifestation is called yaws and is still present in Africa, Asia and Latin America. With the migration of humans to drier areas, the treponemal organism had to retreat off the skin surface into the human body, thus developing a nonvenereal disease which could be transmitted in unhygienic crowded conditions. This manifestation is called bejel and is still present in the Middle East. As human hygiene improved, the treponemal organism retreated further unto the human body attacking bones, blood vessels, and the central nervous system. Following Darwinian theory, the organism further adapted so that it could be transmitted only by contact with moist infected surfaces during intimate human contact such as sexual intercourse. This adaptation developed into venereal syphilis transmission, which is the most common *T. pallidum* infectious presentation in the world today. Modern genetic studies showing that the three treponemal subspecies are genetically distinct and had separate evolutionary paths have disproven the Unitarian theory (Frith, J., 2012; Harper, Zuckerman, Harper, Kingston & Armelagos, 2011).

Syphilis treatment options in the pre-antibiotic era were harsh and could be physically harmful. Since it was believed that syphilis was a New World disease, the initial treatments used New World plants such as the guaiac tree. The guaiac tree bark was made into a drink, which the infected person consumed and then sat in a sweat bath, which raised the person's body temperature for extended periods of time. Because *T. pallidum* is heat sensitive, the resulting fever from the sweat bath may have been more responsible for reducing syphilis symptoms than the guaiac tree drink. In 1917, Julius Wagner-Jauregg infected patients with another fever inducing organism, malaria, which at the time could be controlled by quinine and the resultant fevers appeared to have an effect on reducing neurosyphilis symptoms. Mercury and bismuth salts were also used as syphilis treatments during the 19th and 20th century. In 1908, Paul Ehrlich

discovered an arsenic compound, arsphenamine (Salvarsan), which was lauded as a "magic bullet" and the treatment of choice until the discovery of penicillin by Alexander Fleming in 1928. Penicillin was introduced as a treatment for syphilis in 1943 and is still the antibiotic of choice (Frith, 2012; Tampa, et al. 2014).

Biology of Syphilis Infection

Syphilis is a systemic sexually transmitted disease (STD) that is caused by the spirochete or corkscrew shaped bacterium, *Treponema pallidum* subsp. *pallidum.* Three other pathogenic *T. pallidum* subspecies cause nonvenereal illnesses in humans, *T. pallidum* subsp *endemicum* causes bejel; *T. pallidum* subsp. *pertenue* causes yaws; and *T. pallidum* subsp. *careteum* causes pinta. For the purpose of this dissertation, *T pallidum* will be referred to as *T. pallidum* when discussing syphilis infection, detection and treatment.

The primary host for *T. pallidum* infection is humans. While the organism can live for years inside humans, it does not survive for long outside the human body and cannot be cultured in a clinical laboratory. The organism can only be propagated in rabbits, which is not practical for routine testing in a clinical laboratory. The infected rabbits are the main source of material for laboratory research experiments and test method antigen. This inability to culture *T. pallidum* created difficulties for early researchers trying to discover the cause of syphilis infection and continues to create problems for modern syphilis researchers (LaFond & Lukehart, 2006; Singh & Romanowski, 1999; Tipple & Taylor, 2015).

Syphilis infection progresses through a series of overlapping stages (Figure 2). Initial transmission is primarily through sexual contact, either genital (vagina, penis) or extragenital (fingers, anus, tongue, mouth, throat). The *T. pallidum* organism is also able to cross the placental barrier resulting in maternal transmission to the fetus (congenital syphilis). Infection

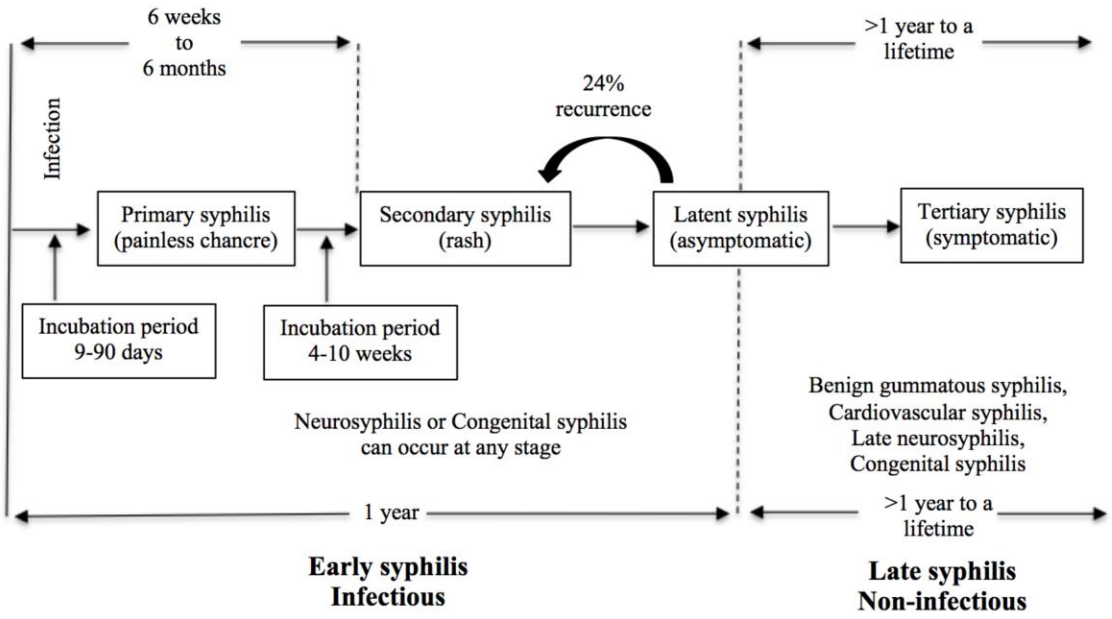

Figure 2. Schematic Representation of the Course of Untreated Syphilis in an Immunocompetent Person. Adapted from "Laboratory diagnosis of sexually transmitted infections, including human immunodeficiency virus", by Unemo, et al, 2013, *World Health Organization* publication, p.107. Retrieved from http://www.who.int/reproductivehealth/publications/rtis/9789241505840/en/ Copyright 2013 by World Health Organization. Adapted with permission.

could also occur at any stage of the disease by transfusion of contaminated blood, which rarely occurs today due to strict federal guidelines regarding pretesting of all blood products for syphilis prior to use. Because of their corkscrew shape, spirochetes are able to propel rapidly through liquid and gel-like materials. Once *T. pallidum* gains access to the human body through skin to skin or mucous membrane contact with infected ulcers, the organism rapidly moves through micro abrasions in the dermis, enters the bloodstream and is disseminated to other organs and tissues. *T. pallidum* begins replication at the inoculation site and induces an inflammatory response that results in a painless chancre or ulcer within about three to six weeks. This painless chancre is a reservoir of spirochetes and is, therefore, highly infectious. However, it may be in a location (vaginal, anal, or pharyngeal) that is not easily seen by the infected

person. This is the primary stage of syphilis, and the risk of infection among exposed contacts is approximately 30%. However, the chancre seen in primary syphilis can also be caused by other genital ulcer diseases (GUD) such as genital herpes, autoimmune disease, chancroid, trauma, or neoplasm which further complicates the diagnosis of a syphilis infection (Cohen, Klausner, Engleman, & Philip, 2013; Golden, Marra, & Holmes, 2003; Ho & Lukehart, 2011; LaFond & Lukehart, 2006).

If untreated, the syphilis chancre heals spontaneously in three to eight weeks; however, *T. pallidum* has now penetrated various anatomical sites such as the central nervous system, eye and placenta. LaFond and Lukehart (2006) theorize that because *T. pallidum* has a slow metabolism, it is able to survive under the immune system radar, thus allowing the organism to replicate slowly and re-seed other organ systems without generating a massive immune response. Once the chancre heals, the infection has moved to the secondary stage; however, the infected person is still capable of transmitting syphilis infection through contact with infectious body fluids. The infectivity rate was extremely high (50%-75%) among partners of persons with primary and secondary syphilis in the era before penicillin and has only dropped to 30% infectivity in the post-penicillin time period (Golden et al., 2003).

The most common symptom of secondary syphilis is a disseminated maculopapular rash. Other illnesses causing similar rashes include HIV, psoriasis, tinea versicolor, scabies, and streptococcal pharyngitis, thus making it difficult to differentiate syphilis from other infections based on symptomology alone. Syphilis is often referred to as the "great imitator" because of the variety of other symptoms that can be exhibited during the course of infection. These symptoms may include weight loss, muscle aches, hair loss, eye inflammation, stiff neck, headache, and a variety of skin eruptions. Secondary syphilis symptoms resolve without treatment; however,

about 24% of untreated cases will develop a recurrence of symptoms within four years from initial infection (Domantay-Apostal, Handog, & Gabriel, 2008; Golden et al, 2003; Ho & Lukehart, 2011).

After the secondary symptoms have resolved, the *T. pallidum* spirochetes are still present, mainly in the liver and spleen of the untreated person. If the person remains untreated, the disease enters the latent phase, which is characterized by a lack of symptoms. Latent disease is divided into early latent (the first year after secondary syphilis symptom resolution) and late latent (any year after the first year of symptom resolution). Early latent syphilis is considered to be infectious because there is a 24% chance of relapse to secondary syphilis. Late latent syphilis is considered noninfectious by sexual contact; however, *T. pallidum* can still cross the placental barrier and infect the fetus in utero. Additionally, at this stage, *T. pallidum* could still be transmitted by a blood transfusion. The latent stage can last many years without any symptoms (Fenton et al., 2008; Singh & Romanowski, 1999).

Tertiary syphilis (late syphilis with clinical manifestations) includes gummatous and cardiovascular syphilis. Tertiary syphilis progresses slowly, with clinical illness occurring from 1 to 46 years after initial infection, and is noninfectious. Because syphilis infection may be asymptomatic, an infected person treated with penicillin for another infection such as streptococcal pharyngitis, might not even know that they had been infected with syphilis. Widespread antibiotic use for other diseases is one reason that tertiary syphilis is rarely seen in developed countries. Cardiovascular syphilis typically involves the aorta with potential for aneurysms or valve insufficiency involvement. Approximately 10% of persons with untreated syphilis will develop cardiovascular symptoms. Gummatous syphilis lesions mainly affect the skin, bone, or liver and are typically benign; however, lesions can also occur in the brain or

heart, which may result in serious health complications. Approximately 15% of persons with untreated syphilis will develop gummatous lesions (CDC, 2015; Golden et al., 2003; LaFond & Lukehart, 2006).

Neurosyphilis can occur at any syphilis infection stage and is a result of *T. pallidum* entering the central nervous system (CNS). In early neurosyphilis, approximately 5% of infected patients will develop symptoms such as severe headache, confusion, and stiff neck, which are usually seen in stroke or meningitis. Neurological symptoms such as vertigo, personality changes, and insomnia may also be present. Up to 20% of untreated patients may develop deafness and/or ocular abnormalities. Late neurosyphilis occurs during the tertiary stage with onset from 2-50 years after initial infection (Golden et al., 2003).

Congenital syphilis is a result of transmission of *T. pallidum* through the placental barrier or during vaginal delivery if the infant comes in contact with an infected lesion. According to Soreng, Levy and Fakile (2014), pregnant women who are infected with syphilis have a higher risk for late-term miscarriage or still-birth (25%) or neonatal demise (11%). The infected child may be asymptomatic when born; therefore, it is important for the mother to be tested for syphilis prior to delivery. Untreated congenital syphilis symptoms are varied in the early stage and can include hepatomegaly, excessive nasal discharge, lymphadenopathy, and bone abnormalities. Deafness can occur in the late congenital syphilis stage along with facial and joint abnormalities (Genc & Ledger, 2000; Singh & Romanowski, 1999).

Syphilis and HIV have an interconnected relationship. Both infections are sexually transmitted and it is possible to have both infections at the same time. Lynn and Lightman (2004) described the "dangerous combination" of the two diseases. HIV infection may change the presentation of syphilis symptoms and disease progression, as well as disrupt syphilis

treatment. The genital ulcers present during early syphilis may increase the potential for HIV transmission. Zetola and Klausner (2007) reported that some experts believe there may be an increased risk of developing neurosyphilis if the person is co-infected with HIV. Chesson, Pinkerton, Irwin, Rein and Kassler (1999) used a mathematical model to estimate that syphilis infection could result in approximately 1000 new cases of HIV in the US. Bergman et al. (2013) reported that persons infected with syphilis are at increased risk (two-to-five-fold) of HIV co-infection. Branger, Van Der Meer, Van Katel, Juriaans and Prins (2009) reported that syphilis infection might increase the HIV viral load and decrease the T-lymphocyte cells bearing the cluster of differentiation 4 (CD4) receptor. The CD4 lymphocytes are referred to as "helper" cells and initiate the body's immune response to a pathogen. HIV infects and destroys the CD4 lymphocytes. A decrease in the CD4 count indicates further progression of the HIV infection, which can impact the syphilis treatment regimen (CDC, 2017; McKenzie & Williams, 2015).

Pathophysiology of Syphilis Infection

Upon introduction in the human body, *T. pallidum* induces an immune response primarily against lipoproteins TpN47, TpN17, and TpN15 which are embedded in the outer leaflet of the organism's cytoplasmic membrane. Antibodies produced against these lipoproteins are the basis for multiple treponemal test methods. Treponemal IgM antibodies develop in approximately two weeks and treponemal IgG antibodies appear within four weeks of infection. In primary syphilis, both treponemal IgM and IgG antibodies are detectable within three days to one week of the appearance of the painless chancre. Following successful treatment, detectable levels of treponemal IgG antibodies will persist for many years; whereas, IgM antibodies decrease to undetectable levels within 6 to 12 months. The treponemal antibodies do not provide immunity

and a person can be re-infected with syphilis after successful treatment (Cortez & Greenwald, 2014; Ho & Lukehart, 2011; LaFond & Lukehart, 2006; Sena, White & Sparling, 2010).

The human body will also produce nonspecific IgG and IgM antibodies in response to lipid material released by the damaged cells when invaded by *T. pallidum* and from cardiolipin on the *T. pallidum* cell surface. These nontreponemal antibodies, termed "reagin," typically appear in serum within 21 days to 6 weeks of initial exposure; and they subsequently disappear from serum following successful syphilis treatment. Detection of nontreponemal antibodies may be indicative of reinfection (Ratnam, 2005; Sena et al., 2010).

Because *T. pallidum* cannot be cultured, early researchers had to rely on infecting human subjects to understand syphilis biology. For many centuries, it was believed that the same organism caused all STDs. In the mid-18th century, a Scottish physician, John Hunter inoculated the penis of a healthy male with material from the infected urethral discharge of another person. The healthy person developed both syphilis and gonorrhea symptoms; thus, for many decades, physicians believed that both diseases were caused by the same organism. In 1831, Philippe Ricord proved scientifically that syphilis and gonorrhea were distinct diseases (LaFond & Lukehart, 2006; Tampa, et al, 2014).

History of Syphilis Diagnostic Testing

The history of syphilis diagnostic testing involves multiple breakthroughs in direct, nontreponemal, and treponemal testing over the past 100 years. These major breakthroughs are shown in Table 2 and are further described in this section.

Direct testing linking *T. pallidum* spirochetes to syphilis was successfully performed in 1905 when Schaudin and Hoffman used a modified Giemsa stain to observe spirochetes in lesions from infected persons. This method was not reproducible by other researchers, but in 1909,

Table 2

History of Syphilis Test Development

Date	Author	Accomplishment
	Direct treponemal tests	
1905	Schaudin and Hoffman	Linked *T. pallidum* with syphilis infection
1909	Coles	Described use of darkfield illumination
1964	Yobs, Brown and Hunter	Developed the DFA-TP test
1991	Noordhoek, Wolters, DeJonge and von Embden	Applied NAAT to the diagnosis of neurosyphilis in CSF
	Nontreponemal tests	
1906	Wasserman, Neisser and Bruck	Developed complement fixation test
1946	Harris, Rosenberg and Riedel	Developed VDRL test
1957	Portney, Carson and Smith	Modified the VDRL to create the USR test
1961	Portney,Bossak, Falcone, and Harris	Modified the USR to create RPR
1983	Petit, Larsen, and Harbec	Modified USR to create TRUST
	Treponemal antibody tests	
1949	Nelson and Mayer	Developed the TPI test
1957	Deacon, Falcone and Harris	Developed the FTA test
1964	Hunter, Deacon and Meyer	Modified the FTA by addition of sorbent, FTA-ABS test
1967	Rathlev	Developed the TPHA test
1969	Cox, Logan and Norins	Modified TPHA to a micromethod (MHA-TP)
1975	Veldekamp and Visser	Developed treponemal ELISA
2000	Fujirebio, Inc.	Developed first FDA approved EIA test
2001	Trinity Biotech	Modified MHA-TP by adding gelatin particles to create TP-PA test

Note: Adapted from *A Manual of Tests for Syphilis* (pages 42-52) by Larsen, Pope, Johnson, & Kennedy, 1998, Washington, DC: American Public Health Association, Copyright 1998 by American Public Health Association.

Coles used darkfield microscopy to observe the distinctive spirochetal motility in samples from infected lesions. A direct fluorescent antibody *T. pallidum* (DFA-TP) test was developed by Yobs, Brown and Hunter in 1964. Nucleic acid amplification tests (NAAT) such as the syphilis polymerase chain reaction (PCR) were described in 1991 by multiple researchers. Noordhoek, Wolters, De Jonge and von Embden applied PCR testing to the diagnosis of neurosyphilis in cerebrospinal fluid (CSF) (Larsen, Pope, Johnson, & Kennedy, 1998).

Wasserman, Neisser and Bruck developed the first nontreponemal serologic test, the complement fixation test, in 1906; however, it would be 40 more years before the test was standardized by Harris, Rosenberg and Riedel in 1946. The standardized test was named for the laboratory that developed the method, the Venereal Disease Research Laboratory (VDRL). The VDRL test provided a method for syphilis screening that could be performed in a clinical laboratory. Portney, Carson and Smith further modified the VDRL test in 1957 to remove the requirement for heating serum before testing and plasma was identified as an additional specimen type. The modified test was referred to as the unheated serum reagin (USR) test. Additional modifications to the USR test followed in 1961, when Portney, Bossak, Falcone and Harris developed the rapid plasma reagin (RPR) test, which improved test readability by adding charcoal particles and allowed use of a disposable slide. In 1983, Petit, Larsen and Harbec modified the RPR by adding a red paint pigment for visualizing a positive reaction and the test was referred to as the toluidine red unheated serum test (TRUST) (Larsen et al., 1998).

Treponemal serologic tests have also advanced since they were first developed in 1949. Currently, treponemal serologic tests can be performed on automated laboratory analyzers, which increases laboratory productivity and efficiency. The *T. pallidum* immobilization (TPI) test first developed by Nelson and Mayer in 1949, required the use of *T. pallidum* (Nichols strain) that had to be grown in rabbit testes and then purified into an antigen. The test was labor intensive and not applicable for general clinical laboratory use. In 1957, Deacon, Falcone and Harris developed a fluorescent treponemal antibody (FTA) test that could be standardized and performed with a fluorescent microscope. Hunter, Deacon and Meyer improved FTA sensitivity in 1964 by adding a sorbent, which removed nonspecific antigens produced by nonpathogenic treponemes present in normal human body flora. This fluorescent treponemal antibody

absorption (FTA-ABS) test was widely used by clinical laboratories for syphilis diagnosis; however, it still required use of a fluorescent microscope. In 1967, Rathlev developed the *T. pallidum* hemagglutination assay (TPHA), which could be macroscopically performed in a test tube. The micro-hemagglutination assay for antibodies to *T. pallidum* (MHA-TP) test was a modification of the TPHA developed in 1969 by Cox, Logan and Norris. The test could be performed in a microtiter plate, which allowed for use of smaller reagent amounts (Larsen et al., 1998). The MHA-TP was further modified by the Fujirebio Corporation in 2000 and incorporated gelatin particles in the test method, which eliminated nonspecific reactions with plasma samples and could be read macroscopically on a microtiter plate. The proprietary test is marketed in the US as the Serodia *T. pallidum* passive particle agglutination (TP-PA) test (Fujirebio, 2006).

In 1975, Veldekamp and Visser developed the first treponemal enzyme-linked immunosorbent assay (ELISA). The ELISA test method provides a measured colorimetric or color change final test result; thus, reducing the subjectivity involved with the visual interpretation of the MHA-TP or TP-PA serologic treponemal test result. The ELISA's 96-well microplate test format allows for many samples to be analyzed at the same time and the procedure could be automated for high-volume testing. Automation removes human error due to inaccurate pipetting, incomplete washing or incorrect incubation time periods from the test method, thus improving test sensitivity (Yolken, 1979). In 2001, the U.S. Food and Drug Administration (FDA) approved the first treponemal EIA test method, the Captia™ Syphilis-G test (Trinity Biotech). The Captia™ Syphilis-G utilized a wild type, whole cell sonicated *T. pallidum* antigen, Nichols strain for *T. pallidum* IgG antibody detection. Since 2001, multiple other FDA approved or cleared treponemal testing methods including CIA and MBIA have been

developed. These treponemal assays utilize either wild type or recombinant treponemal antigens (TpN15, TpN17, or TpN47) to detect IgG and/or IgM antibodies. These tests can be performed either manually or using an automated instrument platform. (Hoover & Park, 2015; Larsen et al., 1998; Malm, et al., 2015; Morshed & Singh, 2015, Sena, et al 2010).

Current Syphilis Diagnostic Testing

Current diagnostic laboratory testing for syphilis depends on either direct detection of *T. pallidum* or measurement of the serologic response to *T. pallidum* infection (Table 3). Each test has advantages and limitations depending on the stage of syphilis infection, and no single test is optimal for diagnosis of syphilis infection at every stage.

Direct detection methods. Darkfield microscopy, used for direct detection of *T. pallidum*, requires a specialized microscope with a darkfield condenser and qualified technical staff to perform the test. The specimen type is an exudate from an active lesion; however, testing cannot be performed on oral lesions due to inability to visually differentiate between benign oral spirochetes and *T. pallidum*. The presence of spirochetes with the morphological characteristics and motility of *T. pallidum* provides a definitive diagnosis of primary or secondary syphilis. Infected persons with darkfield microscopy positive tests may be serologically nonreactive due to delayed antibody response (Ratnam, 2005; Unemo et al., 2013).

The FDA does not currently approve DFA-TP testing; however, some public health and reference laboratories have completed non-FDA method validation studies and offer the test. A specialized fluorescent microscope and trained technical personnel are required to perform the test using a fluorescein-labeled anti-*T. pallidum* globulin. The specimen source is exudate from an active lesion. The test detects *T. pallidum* antigen, and a positive result provides a definitive

Table 3

Syphilis Test Characteristics

Characteristic	Direct Detection	Nontreponemal	Treponemal
Principle	Identify treponemes or treponemal DNA in lesion exudates	*T. pallidum* induces production of reagin or antibodies to lipoidal material released from damaged cells. •Reagins react with cardiolipin	*T. pallidum* infection induces production of specific antibodies (IgM, IgG) •Treponemal tests detect IgG, IgM or total IgM/IgG antibodies to *T. pallidum*
Examples	•Dark Field Microscopy •DFA •NAAT •PCR	•RPR •VDRL •TRUST	•TP-PA •MHA-TP •FTA-ABS •EIA •CIA •MBIA •POC
Advantages	Most specific diagnosis of early syphilis	•Automated RPR with high throughput •Rapid turnaround time and inexpensive reagent (manual) •Objective interpretation (automated methods) •Can revert to negative after specific therapy- useful to monitor response to therapy	•Automated (EIA,CIA,MBIA) with high throughput •Improved detection during early and late infection •Objective interpretation (automated methods) •Specific for treponemal infection
Limitations	•Labor intensive, subjective result (manual) •Specialized equipment •Technically competent staff required •Non-FDA approved methods •Can't differentiate treponemal subspecies	•Labor intensive, subjective result (manual) •Specialized equipment (automated) •Low throughput (manual) •Nonspecific-subject to biologic false positive results •Limited sensitivity in early primary syphilis and late latent syphilis •Potential for prozone effect	•Labor intensive, subjective result (manual) •Specialized equipment (automated) •Low throughput (manual) •Usually positive, even after successful treatment •Not useful in monitoring therapy

Note: Adapted from "Current Trends in Donor Testing to Detect Syphilis Infection" by Cortez & Greenwald (2014) in *Current Infectious Disease Reports*, p. 423. Copyright 2014 by Springer. Adapted with permission.

diagnosis of primary syphilis (Ratnam, 2005; Unemo, Ballard, Ison, Lewis, Ndoma & Peeling, 2013).

NAAT or PCR tests for *T. pallidum* are not currently approved by the FDA. As with DFA-TP tests, some public health laboratories offer the test following completion of method validation studies. Specialized test equipment for extraction and analysis, reagents with probe/primer sets specific for *T. pallidum*, and trained personnel are required to perform NAAT or PCR testing. The specimen sources include lesion exudate, tissue, or body fluid. The test detects *T. pallidum* deoxyribonucleic acid (DNA) and a positive result provides a definitive diagnosis of primary syphilis (Liu, Rodes, Chen, & Steiner, 2001; Ratnam, 2005; Unemo et al., 2013).

Serological test methods. Nontreponemal serological tests measure the non-specific reagin antibodies released in early syphilis infection. These IgG and IgM antibodies are not specific for treponemal infection and can be elevated in multiple other conditions such as autoantibodies (lupus); viral infections (HIV, hepatitis C); bacterial infections (pneumonia, endocarditis); parasitic infections (malaria) and noninfectious causes (stroke, malignancy). Due to their low specificity, there is potential for biological false positive test results; therefore, the CDC recommends that all reactive nontreponemal tests be reflexed to a treponemal test for confirmation of syphilis infection when using the traditional syphilis test algorithm. In the US, three FDA approved nontreponemal tests are currently available: RPR, VDRL, and TRUST. The tests utilize a reagent antigen mixture composed of cardiolipin, lecithin, cholesterol, and a visualization agent to bind the non-specific reagin antibodies present in the infected person's serum. The final reaction is read either microscopically or macroscopically, depending on the test method. The nontreponemal tests can be diluted or titrated to an endpoint, thus providing a numerical indicator for monitoring syphilis treatment. CDC recommends that posttreatment

titers should be performed by the same laboratory using the same initial test method. A rise in titer following treatment could indicate either treatment failure or syphilis reinfection. Following successful treatment, nontreponemal antibodies disappear within one to two years. In some instances, treated patients may become "serofast" resulting in a low reactive nontreponemal titer for life. This most often occurs among patients treated in late stages of syphilis infection or those who have multiple reinfections. Test limitations include prozone reactions, which occur when large amounts of nontreponemal antibodies in the patient's serum overwhelm the test antigens resulting in a nonreactive or false negative test result. (Larsen et al., 1998; Sena et al., 2010; Unemo et al., 2013).

Nontreponemal tests are usually performed manually; however, in 2016, the FDA cleared the first nontreponemal automated assay for clinical laboratory use. The Gold Standard Diagnostics AIX1000 automated analyzer uses proprietary Gold Standard Diagnostics RPR reagents to automate the entire testing process from sample dilution to result interpretation. The analyzer can perform qualitative RPR screens and semi-quantitative titers (1:2 to 1:256) with result interpretation based on a proprietary pattern recognition algorithm (FDA, 2015; Gold Standard, 2016). In 2017, BioRad announced FDA clearance for a dual treponemal total IgG and IgM antibody and RPR multiplex assay that can be performed on its proprietary BioPlex 2200 automated flow cytometry analyzer. This assay provides the first fully automated treponemal/nontreponemal dual assay which detects IgG/IgM antibodies to *T. pallidum* and reagin antibodies. It will also perform an RPR titer (1:2 to 1:64) which further reduces labor costs and allows laboratories to develop more efficient workflows (FDA, 2017).

Treponemal tests are highly sensitive and specific for *T. pallidum* antibodies and remain reactive for years with and without treatment. Treponemal tests cannot differentiate between

active and past infection; therefore, additional testing with a nontreponemal test is recommended to assist with determining disease stage. Treponemal tests also cannot differentiate between treponemal subspecies. Since 1975, multiple different types of treponemal tests that detect *T. pallidum* IgG, IgM or total IgG and IgM antibody levels have been developed and are FDA cleared or approved for use in the US (Table 3). All treponemal tests are based on binding of the treponemal antigen present in the test reagents with the treponemal antibody present in the infected person's serum. The FTA-ABS test is a labor-intensive manual method that utilizes a slide stain and a fluorescent microscope to visually detect the antigen/antibody reaction. The TP-PA and MHA-TP are labor-intensive manual methods that utilize treponemal antigen sensitized red blood cells or gel particles mixed with the patient's serum to visualize the resultant agglutination if treponemal antibodies are present. The treponemal immunoassays, EIA, CIA or MBIA, can be automated and involve use of either antigen coated wells or beads to detect treponemal antibodies present in the patient's serum. An automated analyzer detects the resultant color change or light emission and a test result is electronically generated. Point of care (POC) treponemal assays are easy to use, individual, manual tests that utilize a lateral flow strip or flow through device that is coated with treponemal antigen to detect treponemal antibodies. The test result is read visually. Only one POC treponemal test has received FDA clearance as a waived test, the Diagnostics Direct Syphilis Health Check test. (Larsen et al., 1998; Matthias et al. 2016; Sena et al., 2010; Unemo et al., 2013).

Syphilis Test Interpretation

Syphilis is a multisystem disease that progresses in stages and the symptoms can mimic multiple other diseases. The clinical diagnosis of syphilis infection requires correlation of the patient's history and physical examination with the laboratory test results. This is especially

problematic because in certain stages, e.g. secondary and latent, there may be no symptoms. Syphilis serologic testing requires an immune response from the host, and immunoglobulin class and concentration can vary depending on the disease stage. Further, diseases such as HIV can interfere with the immune response resulting in either false negative or false positive syphilis tests. Because of these issues, the diagnosis of syphilis infection must be determined by performance of a combination of treponemal and nontreponemal tests. There is a dynamic relationship between syphilis antibody production and the stage of disease with treponemal IgM antibodies appearing first at about 2 weeks post infection, followed by treponemal IgG, and then nontreponemal IgG/IgM at about 4 weeks post infection. There is a steady downward decrease in the presence of nontreponemal antibodies during the latent and tertiary syphilis stages, while treponemal IgG antibodies remain present at high levels during all stages (Figure 3). Because of this dynamic, a single syphilis serologic test is insufficient for syphilis diagnosis and all tests performed must be interpreted based on the patient's current symptoms, prior infections, sexual history, and STD risk (Soreng, Levy & Fakile, 2014).

Further complicating the establishment of serodiagnosis of syphilis infection is the wide variation in the sensitivity and specificity of each syphilis diagnostic test depending on the stage of infection, the disease prevalence in the population tested, and the standard used for comparison in the published study (Cantor, Pappas, Daeges & Nelson, 2016; Levett, et al, 2015). Additionally, the sensitivity and specificity of the treponemal EIA, CIA and MBIA tests may provide varied results based on the treponemal antigen(s) (e.g., wild type, TpN15, TpN17, or TpN47) utilized by the test kit manufacturer to detect IgG and/or IgM antibodies (Levett, et al, 2015). Darkfield microscopy sensitivity in the primary and secondary stages can vary from 50-100% due to the subjectivity of the test method and the technical skill of the microscopist. As

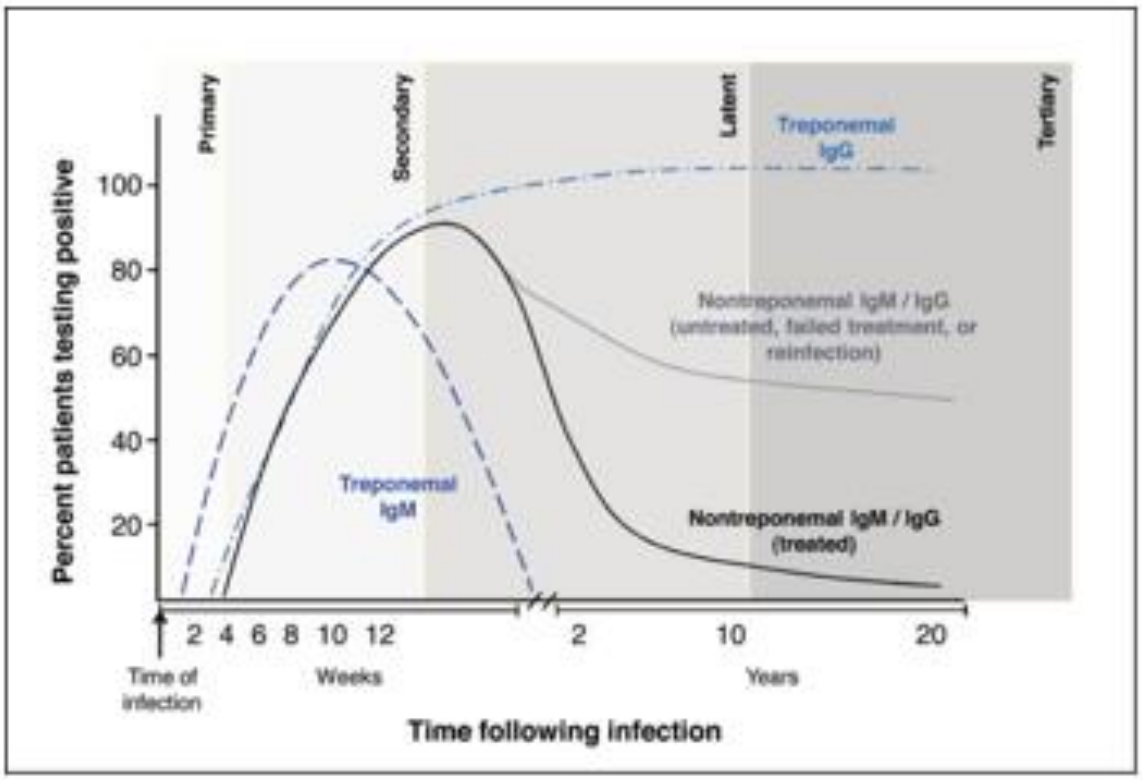

Figure 3. Syphilis Staging and Antibody Response.
From "Serologic Testing for Syphilis: Benefits and Challenges of a Reverse Algorithm" by Soreng, Levy & Fakile, 2014, *Clinical Microbiology Newsletter,36,* p.198. Copyright 2014 by Elsevier. Reprint with permission. Also "Diagnostic tools for preventing and managing maternal and congenital syphilis: an overview" by Peeling & Ye, 2004, *Bulletin World Health Organization, 82,* p.442. Copyright 2004 by World Health Organization. Reprint with permission.

shown in Table 4, nontreponemal and treponemal tests all have excellent sensitivity (99%-100%) for detection of syphilis infection in the secondary stage; but there is wide variability in the other stages, especially in the primary infectious stage. Specificity for all test methods is high ranging from 95% for darkfield microscopy to 98% for nontreponemal tests.

The CDC (2011) recommends a multi-tier or cascade approach to syphilis testing using either a traditional or reverse sequence algorithm. The traditional algorithm (Figure 4) starts with a nontreponemal test, such as the RPR, VDRL, or TRUST. Because the nontreponemal test does not detect *T. pallidum* specific antibodies, the possibility of a biological false positive result may

Table 4

Sensitivity and Specificity of Selected Syphilis Test Methods by Disease Stage

	Sensitivity during stage of infection, average % (range)				Specificity
Test	**Primary**	**Secondary**	**Latent**	**Tertiary**	**Nonsyphilis**
Direct					
Darkfield	76(50-100)[a,b,c,e]	76(50-100)[a,b,c,e]	Not available	Not available	95(85-100)[a,b]
Treponemal					
FTA-ABS	88(70-100)[a,c,e,f]	100[c,e,f]	100[c,e,f]	96[c,e,f]	97(94-100)[a,f]
TP-PA	90(69-100)[a,c,d,f]	100[c,d,f]	100[c,d,f]	94[c]	97(95-100)[a,f]
EIA,CIA,MBIA	89(57-100)[a,d,e,f,g,h]	100(97-100)[d,f,g,h]	93(75-100)[d,f,g,h]	100[f]	96(80-100)[a,e,f,h]
Nontreponemal					
VDRL	79(70-100)[a,c,d,e,f,g]	99(96-100)[c,d,e,f,g]	83(52-100)[c,d,e,f,g]	66(36-96)[c,d,e,f]	98(96-99)[a,e,f]
RPR	89(77-100)[a,c,e,f,g]	99(97-100)[c,e,f,g]	93(53-100)[c,e,f,g]	70(37-94)[c,e,f]	98(93-99)[a,e,f]

Note: Data from [a]Diagnostic tools for preventing and managing maternal and congenital syphilis: an overview by Peeling & Ye (2003) in *Bulletin of the World Health Organization*, p.440; [b]Comparison of Methods for the Detection of Treponema pallidum in Lesions of Early Syphilis by Cummings, et al (1995) in *Sexually Transmitted Diseases*, p.368; [c]Laboratory diagnosis of sexually transmitted infections, including human immunodeficiency virus by Unemo, et al (2013) in World Health Organization, p109;[d]The sensitivity of syphilis assays in detecting different stages of early syphilis by Manavi, et al (2006) in *International Journal of STD & AIDS*, p70;[e]Laboratory methods of diagnosis of syphilis for the beginning of the third millennium by Wicher,et al. (1999) in *Microbes and Infection*, p1037-1042;[f]Novel Treponema pallidum Serologic Tests: A Paradigm Shift in Syphilis Screening in the 21st Century by Sena et al. (2010) in *Clinical Infectious Diseases*, p.701;[g]It is Time to Use Treponema-Specific Antibody Screening Tests for Diagnosis of Syphilis by Loeffelholz & Binnicker (2012) in *Journal of Clinical Microbiology*, p.3.[h]New Proteins for a New Perspective on Syphilis Diagnosis by Smith, et al. (2013) in *Journal of Clinical Microbiology*, p.109

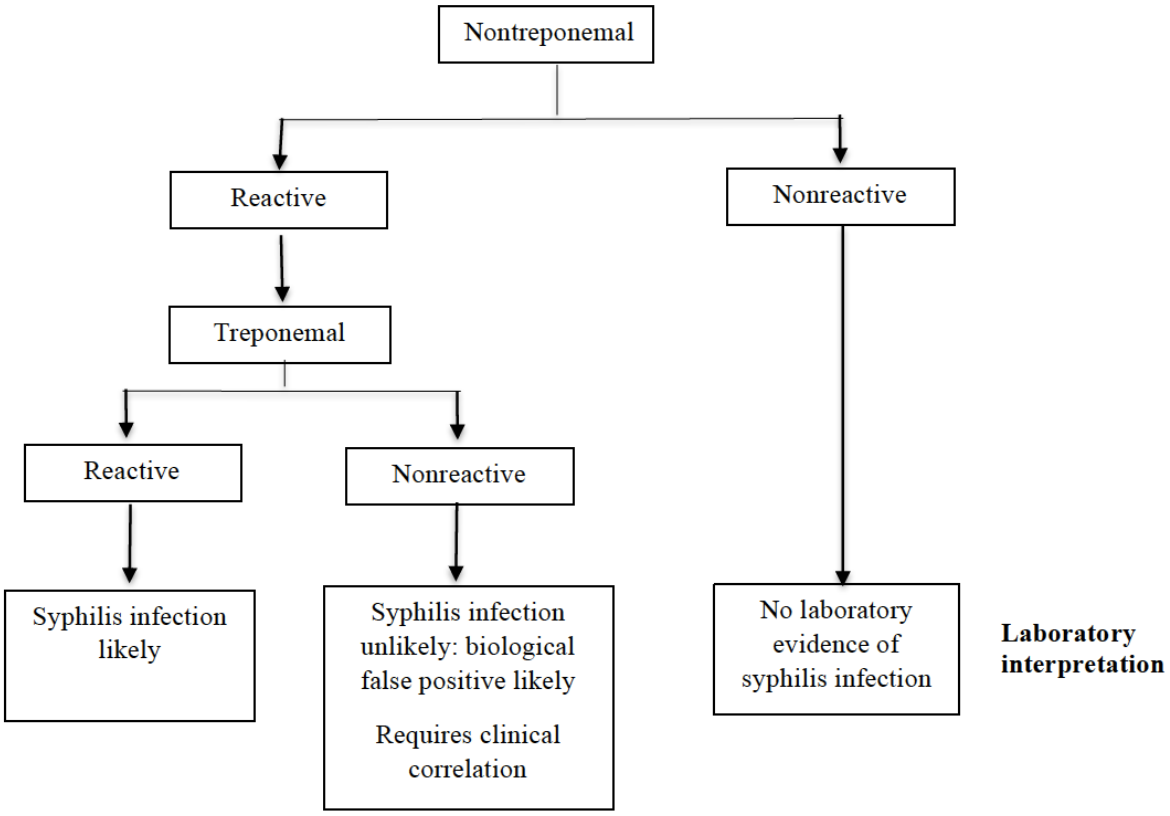

Figure 4. Traditional Syphilis Test Algorithm and Laboratory Interpretation.
Adapted from: "Discordant Results from Reverse Sequence Syphilis Screening-Five Laboratories, United States, 2006-2010" in CDC (2011) *Morbidity and Mortality Weekly*, p.7; and "Recent Trends in the Serologic Diagnosis of Syphilis" by Morshed & Singh (2015) in *Clinical and Vaccine Immunology*, p. 139. Copyright 2015 by American Society for Microbiology. Adapted with permission.

occur; therefore, the CDC recommendation is to reflex a reactive or positive nontreponemal test result to a treponemal test, such as TP-PA, for confirmation of syphilis infection.

The reverse sequence algorithm (Figure 5) begins with a treponemal test, such as an EIA, CIA or MBIA. Because the treponemal test cannot differentiate between past and current infection, a reactive or positive test must be reflexed to a nontreponemal test such as the RPR, VDRL or TRUST. If the nontreponemal test is nonreactive or negative, then the CDC recommendation is to test the sample with a second treponemal test, for example the TP-PA, as

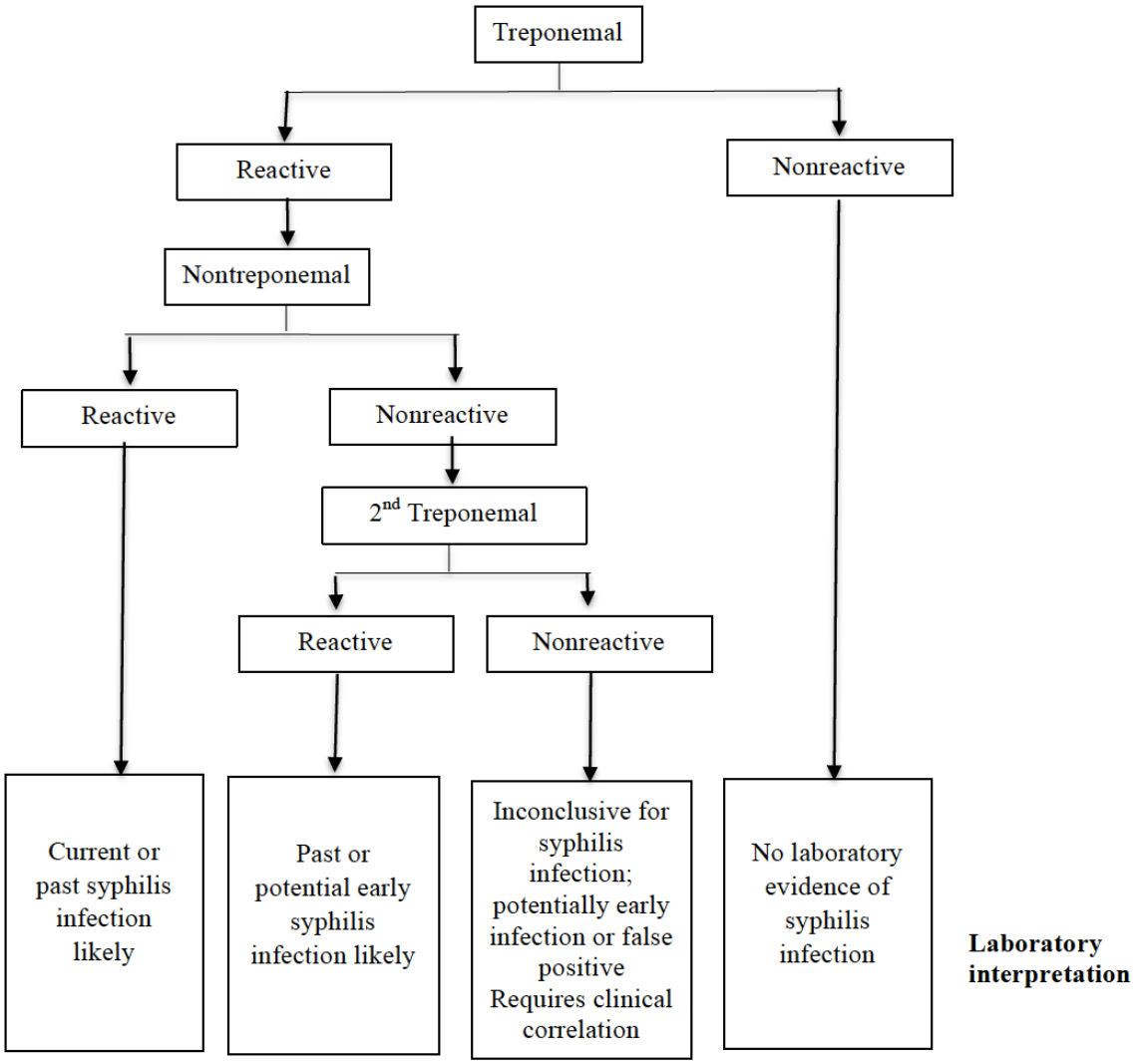

Figure 5. Reverse Sequence Syphilis Test Algorithm and Laboratory Interpretation.
Adapted from: "Discordant Results from Reverse Sequence Syphilis Screening-Five Laboratories, United States, 2006-2010" in CDC (2011) *Morbidity and Mortality Weekly*, p.7; and "Recent Trends in the Serologic Diagnosis of Syphilis" by Morshed & Singh (2015) in *Clinical and Vaccine Immunology*, p. 139. Copyright 2015 by American Society for Microbiology. Adapted with permission.

the "tie-breaker" (CDC, 2011; CDC, 2015; Morshed & Singh, 2015; Soreng, et al, 2014; Zanto, 2010).

There is a potential for discordant results with either algorithm because of varying sensitivity and specificity levels of each of the tests utilized as well as the disease stage when the samples were collected and the population tested. The issue of discordant results using the reverse

sequence algorithm for syphilis diagnosis has caused confusion in the medical and public health community. When three different tests are performed, the possible combinations for test result interpretations can become complex, especially when each test has a different sensitivity and specificity as shown in Table 4 (Binnicker, 2012; Caswell, Hathorn & Manavi, 2016; Morshed & Singh, 2015).

Due to the increase in automated treponemal testing, the CDC provided interpretation recommendations to healthcare providers in 2011 for testing performed using the reverse sequence algorithm. The recommendations were based on a CDC retrospective study of data (n=140,176) from five laboratories that performed the reverse sequence algorithm during 2006-2010. The five laboratories were divided into low prevalence populations (n=3), such as managed care organizations, and high prevalence populations (n=2), such as public health STD clinics. Of the total samples tested, 4,834 (3.4%) had a reactive EIA (Trep-Sure or Trep-Check) or CIA (Liaison) test. Of those reactive samples, 2,734 (56.7%) were RPR nonreactive. These discordant samples were further tested by either TP-PA or FTA-ABS with 866 (31.6%) testing nonreactive. This high discordance level may indicate higher false positive results with the reverse algorithm; however, the CDC study did not compare the reverse and traditional algorithms because the testing at each laboratory was performed on different samples. Also, without clinical correlation, the reverse algorithm positive results could be due to detection of latent syphilis, prior treated syphilis, early syphilis or false positive results. Based on prior published studies showing lower specificity and sensitivity of the FTA-ABS, the CDC recommended that the FTA-ABS should not be used as the second treponemal test in the reverse sequence algorithm to confirm discordant results. Because of its high sensitivity and specificity, CDC recommended use of the TP-PA test as a confirmatory treponemal test (CDC, 2011).

Mishra et al. (2011) conducted a retrospective, population based study (n = 3,092,938) of the impact of the reverse sequence algorithm testing at the Public Health Laboratory–Toronto (PHLT) in the Greater Toronto Area of Ontario, Canada. From 1998-2005, the PHLT utilized the traditional algorithm starting with an RPR for syphilis screening. In 2005, PHLT began using the reverse sequence algorithm starting with an EIA (Enzygnost Syphilis) followed by confirmatory testing of discordant results performed using either TP-PA, FTA-ABS or MHA-TP. The results of the study were similar to the 2011 CDC study with a higher reverse algorithm positive rate (2.24%) versus the traditional positive rate (0.59%). As with the CDC study, this could indicate detection of prior treated disease, but it could also be an indication of early (primary) or latent syphilis. In fact, the study further described the finding that early syphilis was detected in 0.59% of the 9,137 patients who were EIA reactive and RPR nonreactive. These cases would have been undiagnosed if only the traditional algorithm were utilized.

Tong et al. (2013) conducted a cross-sectional study of syphilis serology results for 24,124 patients in a high syphilis prevalence (11.40%) hospital setting in China between December 2011 and May 2013. Each patient sample was tested with RPR, TP-PA, and CIA. The subjects included outpatients, inpatients, and patients receiving routine health screening. Using the results of all three serological tests, the study compared three different testing algorithms: the US traditional algorithm, the US reverse sequence algorithm and the European Centers for Disease Prevention and Control (ECDC) algorithm. The ECDC algorithm differs from the US algorithms in that it starts with a treponemal test and reflexes to a second treponemal test if reactive. The nontreponemal test is not performed in the ECDC algorithm. The study evaluated the diagnostic accuracy of each of the algorithms against the clinical diagnosis. The traditional algorithm had a missed diagnosis (RPR nonreactive) rate of 24.2% (665 of 2,749 TP-PA reactive samples) and a

sensitivity of 75.81%. The sensitivity of the reverse sequence (99.85%) and ECDC (99.38%) algorithms was consistently higher than the traditional algorithm. The syphilis stage of each of the positive samples and the ability of each algorithm to detect positivity at each stage were determined. While the traditional algorithm performed well for detecting secondary syphilis at 99.18% (362/365), it performed poorly for all other stages with sensitivity ranges of 67.8% (primary) to 78.28% (early latent). Both the reverse and ECDC algorithms were highly diagnostic for syphilis infection at all stages with sensitivity ranges from 95.83% to 100%.

Binnicker, Jespersen & Rollins (2011) performed a direct comparison of the traditional and reverse sequence algorithms in a clinical setting on a population with a low prevalence (0.6%) for syphilis. Prospective samples (n=1,000) were sent to the testing laboratory for routine syphilis screening using the reverse algorithm. Initial treponemal testing was performed using the BioRad BioPlex 2200 syphilis IgG MBIA. All samples were also tested by RPR. With the reverse sequence algorithm, 15 of the 1,000 samples (1.5%) were reactive; four samples (0.4%) were reactive with the traditional algorithm. The four reactive RPR samples were also TP-PA and MBIA reactive. Of the 11 remaining MBIA reactive samples, five were reactive by TP-PA with three samples confirmed as prior treated syphilis and two samples staged as latent syphilis. The remaining six samples were TP-PA and RPR nonreactive and were determined to be false positive based on chart review for a false reactive rate of 0.6%. Because the reverse algorithm was able to identify two additional latent syphilis cases, the authors suggest that their study further supports the possibility of enhanced sensitivity for syphilis detection with the reverse sequence algorithm.

Park et al, (2011) performed a cross-sectional retrospective analysis of discordant reverse sequence algorithm testing within a low prevalence population in a large clinical setting.

Reverse algorithm testing was performed using the Diasorin Liason automated analyzer for initial CIA analysis. Reactive samples were reflexed to RPR, and discordant RPR nonreactive samples were then reflexed to TP-PA testing. Syphilis test results from 21,623 analyses were queried with 2% (439/21,623) CIA reactive. There were 58% (255/439) discordant results (CIA+/RPR-) within the low prevalence study population. All patients with discordant test results were clinically evaluated for syphilis infection and, if determined to be infected and not previously treated, offered antibiotic treatment. Reflex TP-PA testing performed on those patients resulted in 72% (184/255) reactive and 28% (71/255) nonreactive. Of the TP-PA reactive patients, 57% (105/184) had a history of treated syphilis. Of the 79 TP-PA reactive patients with no history of syphilis infection, 64% (51/79) received antibiotic treatment for syphilis infection based on clinical assessment and risk factors. There were 28 pregnant females in the study with 43% (12/28) having discordant test results (CIA+/RPR-/TP-PA+) and 17% (5/12) reporting prior syphilis infection. Antibiotic therapy was provided to 92% (11/12) of the pregnant women. The remaining pregnant women (52%) also had discordant test results however the TP-PA was nonreactive (CIA+/RPR-/TP-PA-). These women had no history of syphilis infection and were presumed to have false positive CIA test results. The authors concluded that, when using the reverse algorithm in a low prevalence population, performance of a second treponemal test provided valuable information for detecting syphilis infection when test results were interpreted based on a patient risk assessment and clinical examination.

Rourk, Nolte, and Litwin (2016) analyzed the impact of implementation of the reverse syphilis algorithm at the Medical University of South Carolina. South Carolina has a primary and secondary syphilis rate of 5.7 per 100,000 population which is considered to be a moderately high syphilis prevalence rate. Clinical samples were initially tested using a syphilis EIA

(BioPlex 2200 syphilis IgG multiplex MBIA) as the screening assay. Reactive EIA samples were reflexed to RPR with titer and discordant nonreactive RPR samples were reflexed to TP-PA testing. Between July 22, 2013 and February 16, 2015, 10,060 patients were tested for syphilis with 502 (5%) reactive on the initial EIA screen and 352 nonreactive by reflex RPR (70.1% discordance). When the discordant samples were tested with TP-PA, 103 samples were nonreactive and identified as potentially falsely reactive by initial EIA screening. However, as seen in studies discussed earlier, this study identified an increase in either latent syphilis or untreated/unknown treatment of past syphilis with 21 cases (8.7% of the EIA reactive/RPR nonreactive/TP-PA reactive) that would not have been identified if the facility were using the traditional algorithm. Additionally, two cases of early primary or secondary syphilis were also identified in the nonreactive RPR group. Using the traditional algorithm, a nonreactive RPR test would have resulted in no additional treponemal testing and no treatment. The authors emphasized the importance of performing a second treponemal test when results are discordant and correlation with medical history.

Gratix, et al (2012) also identified an increase in diagnosis of late latent syphilis using the reverse algorithm in their study performed in the Edmonton, Canada public health laboratory. Prior to 2007, public health laboratory syphilis testing was performed using the traditional (RPR first) algorithm. In September 2007, the laboratory switched to the reverse treponemal algorithm using the Abbott Architect Syphilis CIA in order to increase diagnostic sensitivity and specificity as well as utilize the high throughput screening capability of an automated analyzer. As previously shown in Figure 3, nontreponemal RPR antibody titers are slow to rise in early syphilis and decline over time becoming nonreactive in untreated late syphilis which leads to lower sensitivity for syphilis detection at this stage when using the traditional algorithm. This

study, using aggregated syphilis data from 2004 to 2009, was performed to analyze the impact on the identification of primary and latent syphilis cases following implementation of the reverse algorithm. The study found that the number of late latent syphilis cases doubled following implementation of the reverse algorithm. Using the traditional algorithm, there were 97 late latent syphilis cases (0.07%) identified from 2004-2007. Using the reverse algorithm, there were 137 cases (0.14%) identified from 2007-2009. Additionally, there were 3 cases of discordant primary syphilis (EIA reactive/RPR nonreactive) identified using the reverse algorithm. Of the 137 late latent syphilis cases, 81 were RPR nonreactive and diagnosis would have been missed with the traditional algorithm. Treatment of late latent syphilis cases can reduce the chance for tertiary disease complications. Of those 81 cases, there were 18 pregnant women who were at potential risk for maternal transmission of syphilis infection to their unborn child.

Laboratories using the reverse algorithm should consider the sensitivity of both the initial screen and the second confirmatory treponemal tests when choosing test methods. Zhang et al. (2012) performed an analytical sensitivity analysis of five different treponemal assays including BioPlex 2200 Syphilis IgG (Bio-Rad), Liaison Treponema (Diasorin), Trep-Sure (Phoenix Bio-Tech), Captia Syphilis-G (Trinity Biotech), and Serodia TP-PA (Fujirebio). For the analysis, 10 well-characterized treponemal positive clinical samples were serially diluted (two-fold dilutions) to produce a set of 10 dilutions with gradually decreasing syphilis antibody levels for each sample. All five methods were tested with each sample panel, that is, the initial undiluted serum and the 10 serial dilutions, to determine the number of dilutions required to reach a negative test result. All five methods were positive for the undiluted samples. The BioPlex, Liason and Captia methods were positive for three serial dilutions. The Trep-Sure was positive for seven dilutions and the TP-PA was positive at 10 dilutions suggesting high sensitivity for these test

methods. Laboratories have options when choosing which treponemal tests to include in reverse sequence algorithm testing. The authors suggest that if a highly sensitive test is used as the initial screening assay and a less sensitive test is used for confirmation, there is a possibility of a false negative result with the second test. This could further complicate the clinical management of the patient as the healthcare provider may consider the initial screening test to be a false positive.

Castro et al. (2013) performed an extensive study of 290 well-characterized stored serum samples analyzed with nine different treponemal assays to determine the analytical sensitivity of each assay. The assays included the FTA-ABS (Zeus Scientific), Liaison (DiaSorin), SD BIOLINE Syphilis 3.0 POC (Standard Diagnostics), INNO-LIA Immunoblot (Innogenetics), BioELISA (BioKit), Captia IgG (Trinity Biotech), Trep-ID (Trinity Biotech), and Trep-Sure (Trinity Biotech). Median end point titers were calculated for each test method. Results varied from the most sensitive (Trep-Sure 1:512) to the least sensitive (FTA-ABS 1:4). Based on the low sensitivity of the FTA-ABS, the authors recommended that its use as a confirmatory test should be reevaluated. Castro, et al. (2013) also recommended that laboratories considering using the reverse sequence algorithm should not use a less sensitive second confirmatory test.

The CDC 2015 STD treatment guidelines recognized the increase in laboratories performing automated treponemal assays as the initial syphilis screening test. These automated testing platforms increase laboratory efficiency due to high volume throughput and the ability to interface with Laboratory Information Management Systems (LIMS), which decreases typographical errors due to manual test result entry. An automated test platform also reduces operator error due to fatigue and provides an objective test result because the instrument interprets the optical reading to determine the final result. With the recent (2016-2017) FDA

clearance of a nontreponemal automated RPR analyzer (Gold Standard Diagnostics AIX1000) and approval of an automated RPR assay for the Bioplex 2200 analyzer, laboratories will now have another option for improving laboratory efficiency using the traditional algorithm.

There are advantages and disadvantages to the use of either the traditional or reverse syphilis screening algorithms. Both algorithms require additional testing to confirm the initial result and, as with any laboratory test, clinical correlation is required for disease diagnosis. Due to the potential for false biological positives with nontreponemal tests, the traditional algorithm requires performance of a second treponemal test for confirmation. If the second test is nonreactive, the initial nontreponemal test could be considered to be a false positive. The traditional algorithm has a long history of test performance and correlation with disease staging. CDC continues to recommend the traditional algorithm; however, multiple studies have indicated that the algorithm may have low sensitivity in various syphilis disease stages. The reverse algorithm also requires confirmation of an initial reactive treponemal test. In order to distinguish past from active infection, the CDC recommends follow-up with a nontreponemal test. If there is discordance between the treponemal and nontreponemal test results, then a second treponemal test is recommended to identify if syphilis infection is likely. This additional testing adds to laboratory costs, increases turnaround time, and may increase the cost of public health disease investigations due to either initial false positive tests or prior treated infections. However, recent studies have shown an increased sensitivity and specificity when using the reverse algorithm within certain populations and disease stages. (Cortez & Greenwald, 2014; CDC, 2011; CDC, 2015; Loeffelholz & Binnicker, 2012; Park, et al., 2011; Sena et al., 2010; Young et al., 1992).

The 2015 CDC STD syphilis treatment guidelines state that the reverse sequence algorithm can identify persons with past syphilis infection, those with untreated or partially treated syphilis,

and persons with false positive results that can occur in a low prevalence setting. Their 2011 recommendation continues to be that a person with a reactive treponemal screening test should have a quantitative nontreponemal test performed automatically by the testing laboratory, as the results of both tests will guide treatment management. If there is discordance, that is the nontreponemal test is nonreactive, then a second treponemal test, preferably a test based on different antigens from the initial treponemal test and with a higher sensitivity, should be performed. Additionally, healthcare providers should also perform a clinical evaluation, including sexual risk assessment, prior history of syphilis infection and prior syphilis treatment. If the second treponemal test is reactive (positive) and the person has a prior history of syphilis treatment, then no further treatment is needed unless the sexual history suggests potential reexposure. If there is a potential for reexposure, CDC recommends retesting of the patient with a nontreponemal test in two to four weeks to check for early syphilis infection. Those persons without a history of syphilis treatment should be treated for late latent syphilis. While late latent syphilis is not considered infectious, antibiotic treatment may prevent progression to tertiary syphilis in the infected person. If the physical examination suggests a recent infection, then the person would be treated as a primary syphilis case. If the second treponemal test is nonreactive and there is a low epidemiologic risk and clinical probability for syphilis, then further treatment is not indicated (CDC, 2015).

Treponemal Immunoassay Index Values

In order to calculate the test result for the patient's specimen, each of the automated treponemal immunoassays compares an optical reading for the patient's specimen to the reading for a standard that has an antibody concentration at a cutoff value determined by the test kit manufacturer. The cutoff value must be listed in the manufacturer's test kit package insert,

which was approved by the FDA. Optical readings below the cutoff value are considered to be nonreactive, and those above the cutoff value are considered to be reactive. Some immunoassays may also include an equivocal range of readings, which may require repeat testing with a new specimen. Each specimen tested will have a numerical (quantitative) optical reading, which, depending on the instrument, may be referred to as an Index value, Signal to Cutoff (S/CO) value, or Antibody Index (AI). The treponemal immunoassay tests report only qualitative test results (nonreactive, reactive, equivocal); however, the numerical or quantitative optical reading may be useful in resolving the discordant results that can occur with the reverse sequence algorithm, or to provide a semi-quantitative estimate of the likelihood of an infection. Reporting this numerical value is not included in the FDA cleared or approved manufacturer's instructions for any of the treponemal EIA, CIA or MBIA test methods and would require additional non-FDA method validation of the test method. The 2015 CDC STD Syphilis treatment guidelines state that recently published studies have demonstrated correlation of a high quantitative optical reading with a positive TP-PA test result. CDC further recommends that these findings should be investigated to determine their clinical significance (CDC, 2015).

Wong et al. (2011) compared the performance of the Trep-Sure EIA against the VDRL, using 674 specimens collected in a public health clinic with a high prevalence (9.4%) of syphilis. The authors found a strong correlation between the Trep-Sure index value and the TP-PA test result. For the Trep-Sure assay, an index value <1.0 is considered nonreactive. The study identified 14 specimens with an index value between 0.200 to 0.999, which were reported as nonreactive. Follow up testing of the 14 specimens revealed that 43% (6/14) were TP-PA reactive. The authors were unable to perform clinical correlations to determine if the patient was infected with

syphilis; however, they propose that flagging elevated index values on nonreactive samples may alert healthcare providers to perform additional clinical follow up.

Dai et al. (2014) performed a study to determine if the S/CO ratio of the Architect Syphilis TP CIA test could be used to predict the confirmatory TP-PA result with 100% accuracy, thus eliminating the requirement to perform a second confirmatory test in the reverse algorithm. The study was performed on serum samples collected from patients hospitalized at an integrated cancer center in China. The prevalence of syphilis within this population was 3.3%. All samples (n=8,980) were tested initially with the Architect Syphilis TP with reactive samples reflexed to TP-PA for confirmation. All Architect Syphilis TP reactive samples were also tested with the TRUST nontreponemal assay. The study concluded that 100% of the samples with an S/CO ratio of 9.9 were reactive when confirmed by TP-PA. Of the 319 screen reactive samples, 87 (27%) had S/CO ratio of 9.9 or greater and were TP-PA reactive. The authors proposed that these samples did not require confirmatory TP-PA testing because of the elevated S/CO value. The ECDC syphilis algorithm recommends reflex to a second treponemal test for all initially reactive treponemal tests. In this study, all 319 samples would have required a second test. The authors proposed a modified algorithm that included performing the nontreponemal TRUST test with reflex to TP-PA if negative and the S/CO value was between 1.0 and 9.9. Those samples with S/CO value greater than or equal to 9.9 were considered to be confirmed positive without running the TP-PA test. Following the modified algorithm, 34.8% (n=111) of the samples would require a second TP-PA and 65.2% (n=208) would not require a second test, resulting in savings in reagent costs and staff time for the laboratory and more rapid turnaround time for test results.

Yen-Liberman, et al. (2011) further confirmed the usefulness of syphilis MBIA index values for resolving discordant reverse algorithm test results. They performed an analysis of 142

samples collected from a low-prevalence (3%) clinic population that tested reactive with the BioPlex 2200 Syphilis IgG assay. The samples were further tested with the Trep-Sure EIA with 77% (110/142) testing reactive. The authors were able to determine that a quantitative AI value of 6.0 on the BioPlex 2200 was 100% specific for confirmation by a second treponemal (Trep sure EIA) test. Using this laboratory established cutoff value, the laboratory could decide to not perform the second treponemal test, thus, saving time and resources.

Berry and Loeffelholz (2016) performed a retrospective, descriptive analysis of specimens screened for syphilis at University of Texas Medical Branch between January and May 2014 to analyze the impact of instituting a modified reverse syphilis algorithm using the BioPlex Syphilis IgG MBIA antibody index (AI) to determine the necessity of additional TP-PA testing. The overall patient population included high risk incarcerated persons (28%) and miscellaneous inpatient and outpatient hospital patients (9%) along with low risk obstetrical/gynecological (OB/GYN) patients (63%). The syphilis reactive rates within the studied patient population were 1.8% for OB/GYN, 7.5% for incarcerated persons and 5% for miscellaneous hospital patients. Because prior published studies had identified a false reactive rate of 0.6% when using the reverse algorithm in low prevalence populations, the University modified the reverse algorithm used at their facility for all syphilis screening to reflex to both an RPR titer and a TP-PA when the initial treponemal test was reactive or equivocal. This modification differs from the recommended CDC algorithm which does not require performing a confirmatory TP-PA test if the RPR test is reactive and is similar to the modification utilized at the western regional laboratory where the dissertation study was performed. Among all patient populations tested in the study, 665 had an initial reactive Bioplex Syphilis IgG test result. Of those reactive samples, 430 (65%) had an AI greater than or equal to 8 and 99.3% (427/430) were confirmed reactive by

TP-PA. The authors conclude that based on this study, they could reduce laboratory workload by eliminating TP-PA testing on those 427 samples with an AI greater or equal to 8 thus resulting in laboratory cost savings and a reduction in turnaround time for completed test results. The authors also found nine instances among the incarcerated and OB/GYN populations where both the Bioplex and RPR tests were reactive, but the TP-PA was nonreactive. They conclude that this discordance could be due to the higher sensitivity of the Bioplex treponemal assay which has been documented in prior studies. The authors discuss the possibility of utilizing the AI value to decrease the number of confirmatory tests, but it could also be used to interpret discordant test results.

Laboratory Test Usefulness

There are multiple statistical measures that have been utilized to evaluate the usefulness of a laboratory diagnostic test. Various authors have used the terms usefulness and accuracy interchangeably, and there are no standard criteria for determining which statistical measures are associated with each term. (Aamir & Hamilton, 2014; Bossuyt, Reitsma, Linnett & Moons, 2012; Galen & Gambino,1975; Guyatt, Tugwell, Feeny, Hagman & Drummond, 1986; Hulley, Cummings, Bowner, Grady & Newman, 2013; Kaplan, 1990; Linnett, Bossuyt, Moons, & Reitsma, 2012; Lord, Staub, Bossuyt & Irwig, 2011; Vihinen, 2012). This section will discuss literature review of definitions various authors have used for diagnostic test usefulness.

A single statistical measure does not provide sufficiently complete information to determine the usefulness of a diagnostic test. According to Aamir and Hamilton (2014), a useful laboratory test should be able to detect or confirm a disease, provide information about disease progression, and/or quantify the disease abnormality. An abnormal (typically positive) test should be highly associated with disease presence. A normal (typically negative) test should be highly associated

with disease absence. The accuracy or efficiency of a laboratory test is determined by the percentage of patients correctly identified as either having or not having the disease for which they have been tested. The authors suggest the statistical measures of sensitivity, specificity, accuracy, predictive values, and likelihood ratios can be used to determine laboratory test usefulness.

Hulley et al. (2013) describes five determinants of usefulness: reproducibility (provides the same results when the test is repeated), accuracy (degree that the measurement corresponds to the true value or disease state), feasibility (costs, risks, acceptability), effect on clinical decisions, and effect on clinical outcome. A test may be accurate, but may not improve patient outcomes and, therefore, may not be useful.

Kaplan (1990) discusses five crucial test characteristics that impact clinical interpretation. He divides the five characteristics into two groups: characteristics that assist with test judgment (reliability and accuracy) and characteristics that assist with diagnostic probability (sensitivity, specificity, and predictive values, both positive and negative). According to Kaplan, clinicians approach disease diagnosis by successively revising diagnostic probabilities. Only the history and physical may be available when the clinician first assesses the ill patient; and as diagnostic test results become available, the clinician revises the disease probability by correlating the test results with the history and physical. A laboratory test result must always be interpreted based on clinical assessment of the patient.

Galen and Gambino (1975) state that laboratory test usefulness cannot be interpreted only in terms of normal or abnormal test results. They introduced the concept of the predictive value, how accurately a test predicts the presence or absence of disease, for determining the usefulness of a laboratory test. They also included sensitivity, specificity, and prevalence or incidence as

important variables for determining laboratory test usefulness. Galen and Gambino further emphasized the importance of a good physical examination and history as "essential for the effective use of laboratory tests" (p.18). The authors provided the following list of laboratory testing functions that can help define a "good" test from a "bad" test:

- Provide a correct diagnosis in a patient known to be sick.
- Provide a prognosis in a patient with a known disease.
- Provide an indication as to whether or not a disease is present in its early or subclinical stages in an otherwise "healthy" person.
- Provide data for monitoring the level of therapeutic drugs, or the effects of drugs , or both.
- Provide data that may indicate whether or not disease might develop at some future time, that is, delineation of risk factors.

Guyatt, et al. (1986) states that a clinically useful test is able to accurately distinguish between disease and non-disease states and quantitate disease severity. Accuracy should be established by independent comparison of the test against a gold standard with consideration of test method precision. If a gold standard is available, then sensitivity and specificity should be calculated to determine accuracy. Guyatt, et al. further states that the most powerful method for measuring a test's usefulness is to determine likelihood ratios. A likelihood ratio is the likelihood of a test result given disease (sensitivity) divided by the likelihood of a test result given no disease (specificity). Grimes and Schulz, (2005), Hulley et al., (2013), Jaeschke, Guyatt and Sackett (1994), Kent and Hancock (2016) all state that likelihood ratios are a powerful determinant of the clinical usefulness of a test. A likelihood ratio of one provides no information about the probability of disease presence; whereas the higher the likelihood ratio is

above one, the better the probability that the test result is ruling in the presence of disease. A guide for likelihood ratio magnitude as recommended by Jaeschke et al. (1994) and Kent and Hancock (2016) is presented in Table 5.

Table 5

Likelihood Ratio (LR) Interpretation

Positive LR	**Strong evidence for ruling in diagnosis** above 10	**Moderate evidence for ruling in diagnosis** between 5 and 10	**Weak evidence for ruling in diagnosis** between 2 and 5
Negative LR	**Strong evidence for ruling out diagnosis** less than 0.1	**Moderate evidence for ruling out diagnosis** between 0.1 and 0.2	**Weak evidence for ruling out diagnosis** between 0.2 and 0.5

Vihinen (2012) further discusses test accuracy as a statistical measure that can be calculated using the values in a 2x2 contingency table. Accuracy is calculated as the percentage of patients who are correctly identified by the test as either having or not having the disease (true positive plus true negative divided by the total contingency table number). Using the same calculation, Galen and Gambino (1995) define the term efficiency rather than accuracy to describe the ability of a test to detect a specific disease. Ball (2014) also uses the term efficiency as synonymous for accuracy to represent the "number of correct diagnoses as a percentage of all the diagnoses" (p. 210). Vihinen (2012) states that accuracy may provide a more comprehensive statistical measure than the row-wise (predictive values) and column-wise (sensitivity and specificity) measures within the contingency table. For this dissertation, accuracy, as described by Vihinen, will be the term used to calculate the percent of all results that are true results, whether positive or negative.

Contingency table. Sensitivity, specificity, predictive values, likelihood ratios, accuracy, and prevalence can be calculated from a 2x2 contingency table. The table contains dichotomous results showing the number of patients with the disease and a positive test result (true positive),

number of patients who do not have the disease and a positive test result (false positive), number of patients who do not have the disease and have a negative test result (true negative), and number of patients who have the disease and have a negative test result (false negative). The contingency table shown in Table 6 provides formulas for sensitivity, specificity, predictive values, likelihood ratios, accuracy, and prevalence calculations.

Table 6

Diagnostic Test Measures: Calculation of Sensitivity, Specificity, Predictive Values, Likelihood Ratios, Accuracy, and Prevalence

		True classification (Gold Standard)			
		Positive Test (Disease)	**Negative Test (No Disease)**	**Total**	**Measure**
Predicted classification	**Positive Test**	a True-positive (TP)	b False-positive (FP)	a + b	Positive predictive value (PPV) = a/(a+b)
	Negative Test	c False-negative (FN)	d True-negative (TN)	c + d	Negative predictive value (NPV) = d/(c+d)
	Total	a + c	b + d	a+b+c+d	
	Measure	Sensitivity = a/(a+c)	Specificity = d/(b+d)		Accuracy= a+d/a+b+c+d
		Positive likelihood ratio= sensitivity / 1-specificity	Negative likelihood ratio= 1-sensitivity / specificity		Prevalence = (a+c)/(a+b+c+d)

Adapted from: *Designing Clinical Research* by Hulley, et al (2013). Philadelphia, PA: Lippincott Williams & Wilkins, p.177. Copyright 2007 by Lippincott Williams & Wilkins; "How to evaluate performance of prediction methods? Measures and their interpretation in variation effect analysis" by Vihinen, M. (2012), *BMC Genomics*, 18(13), p. 5. Copyright 2012 by Vihinen: Licensee BioMed Central Ltd; "Refining clinical diagnosis with likelihood ratios" by Grimes, D.A. & Schulz, K.F. (2005), *The Lancet*,365, p.1501. Copyright 2005 by Elsevier Limited.

Sensitivity measures the ability of the diagnostic test to correctly identify or screen in the presence of disease within the tested population (positive test result). Specificity measures the ability of the diagnostic test to correctly identify or screen out those patients who do not have the disease (negative test result). These measures are independent of the prevalence of disease

within the tested population. Predictive values are influenced by the disease prevalence within the population and provide a measure of the proportion of tested patients who either have or do not have the disease. The positive predictive value (PPV) of a test is the percentage of true positives divided by the total of true and false positives. The negative predictive value (NPV) is the percentage of true negatives divided by the total of true and false negatives. Prevalence is estimated by dividing the total number of cases with the disease by the total number in the population at risk of having the disease. Typically, lower disease prevalence results in lower positive predictive values. Likelihood ratios (LR) measure how likely it is that the test result rules in the disease in question and uses both the sensitivity and specificity measure to calculate a single result. A positive LR is calculated as sensitivity/(1-specificity) and a negative LR is calculated as (1-sensitivity)/specificity. Likelihood ratios are independent of disease prevalence; these values do not vary in different populations or settings and can be applied to a specific patient. Likelihood ratios are reported as either positive (greater than one) or negative (less than one) (Aamir & Hamilton, 2014; Galen & Gambino, 1975; Grimes, D.A. & Schulz, K.F. 2005; Hulley et al., 2013; Kaplan, 1990; Motulsky, 2014; Polit & Beck, 2012).

Attia (2003) discusses the properties of diagnostic tests that can be calculated from the 2x2 contingency table formulas shown in Table 6. The measures of sensitivity, specificity and predictive values reflect population characteristics. In order to visually see the impact of disease prevalence on predictive values and LRs, calculations were performed on an example population of 10,000 using a test with 99% sensitivity (a very good test) and 95% specificity. As the prevalence increased from 1% to 10% in the example population with the sensitivity and specificity held constant, the PPV also increased; however, the LRs do not change. In the low prevalence population seen in Table 7, the PPV is 16.7% which means there is approximately a

Table 7

Impact on Predictive Value and Likelihood Ratio at 1% Disease Prevalence for a Test with 99% Sensitivity and 95% Specificity

	Positive Test (Disease)	**Negative Test (No Disease)**	**Total**	**Measure**
Positive Test	99	495	594	**16.7% PPV**
Negative Test	1	9405	9406	**100.0% NPV**
Total	100	9900	10000	
Measure	99% sensitivity **19.8 +LR**	95% specificity **.001 -LR**		95% accuracy **1% prevalence**

17% probability that a positive test result represents a person with the disease. Performing highly sensitive tests in a low prevalence population may lead to higher false positive test results, which can lead to over treatment, anxiety and additional testing.

As the prevalence increased to 10% as shown in Table 8, the PPV also increased to 68.8%, which means there is approximately a 69% probability that a positive test result indicates the person has the disease. In Tables 7 and 8, the likelihood ratios did not change as the prevalence changes. With a positive LR of 19.8, the likelihood of a patient with a positive test result having the disease tested for increased by approximately 20-fold, regardless of the prevalence in the population.

Because predictive values reflect disease presence in the population testing, Attia (2003), Grimes and Schulz (2005), and Kent and Hancock (2016) suggest that LRs should be used by the healthcare provider to interpret test results for an individual patient. Laboratory test results should be interpreted in correlation with a clinical assessment of the patient and LRs allow the healthcare provider to refine the clinical diagnosis for that specific patient. Clinical laboratories do not report LRs for each laboratory test; however, as a component of method verification for establishing a new FDA approved test method and during day to day operation, laboratories

Table 8

Impact on Predictive Value and Likelihood Ratio at 10% Disease Prevalence for a Test with 99% Sensitivity and 95% Specificity

	Positive Test (Disease)	**Negative Test (No Disease)**	**Total**	**Measure**
Positive Test	990	450	1440	**68.8% PPV**
Negative Test	10	8550	8560	**99.9% NPV**
Total	1000	9000	10000	
Measure	99% sensitivity **19.8 +LR**	95% specificity **.001 -LR**		95% accuracy **10% prevalence**

gather data that could be placed in a 2x2 contingency table. Therefore, all measures including LRs, could be calculated if requested. As shown in Table 5, LR values could be used to determine strong, moderate, or weak evidence of disease. Attia (2003) states that a useful test provides a high positive likelihood ratio (+LR) and a small negative likelihood ratio (-LR).

Bayes' nomogram. According to Attia (2003) and Kent and Hancock (2016), additional value for the usefulness of a clinical test for a specific patient can be obtained if the LR is plotted on a Bayes' nomogram to provide an estimation of the probability of disease. Use of the Bayes' nomogram requires an estimate of the pretest probability of disease, which is based on either clinical decision rules or the clinician's assessment after performing a history and examination of the individual patient, not the prevalence of disease in the population. For example, if a patient in a low syphilis prevalence (1%) population presents to the healthcare provider with a history of multiple unprotected sexual encounters but no specific symptoms to suggest syphilis, the healthcare provider may assign a pretest probability for syphilis infection of 30% based on history and clinical assessment. From the data presented in Table 7, the PPV for the syphilis test in a 1% prevalence population is 16.7%. Using this information, the healthcare provider would interpret a positive result in this low prevalence population at about 17% probability that the test

represented syphilis infection. An LR could provide additional information for interpreting the syphilis test result. The Bayes' nomogram shown in Figure 6 plots the 30% pretest probability from the clinical assessment performed by the healthcare provider on the first axis and the +LR of 19.8 for the test obtained from Table 7 on the second axis. A straight edge is used to draw a line connecting the two axes, and the connecting point on the third axis is the posttest probability. In the example in Figure 6, the posttest probability is now approximately 88% that the patient has syphilis. Calculating LR values and making them accessible could provide healthcare practitioners with another tool to incorporate the results of diagnostic testing into clinical decision making, especially with discordant syphilis test results.

Receiver operating characteristic (**ROC) plots.** Zweig and Campbell (1992) discuss the use of ROC plots for determining diagnostic test accuracy. A ROC plot or curve graphs the sensitivity and specificity of each test data point across the complete range of observed results. When determining sensitivity and specificity, a decision threshold must be used to differentiate a normal from an abnormal value. If the decision threshold is varied over the entire range of possible results, the sensitivity and specificity will move in opposite directions. When one increases, the other decreases; therefore, each decision threshold will have a corresponding sensitivity/specificity pair. The ROC curve provides a visual representation of these pairs and, therefore, the ability of the test to discriminate disease from non-disease. The *y*-axis of the ROC curve reflects sensitivity (true-positives) and is calculated only from the subgroup that is positive in the presence of disease. Sensitivity is defined as: (number of true-positive test results)/(number of true-positive + number of false-negative test results). The *x*-axis reflects 1-specificity (false-positive) and is calculated only from the subgroup that is negative in the in the absence of disease. A false positive is defined as: (number of false-positive

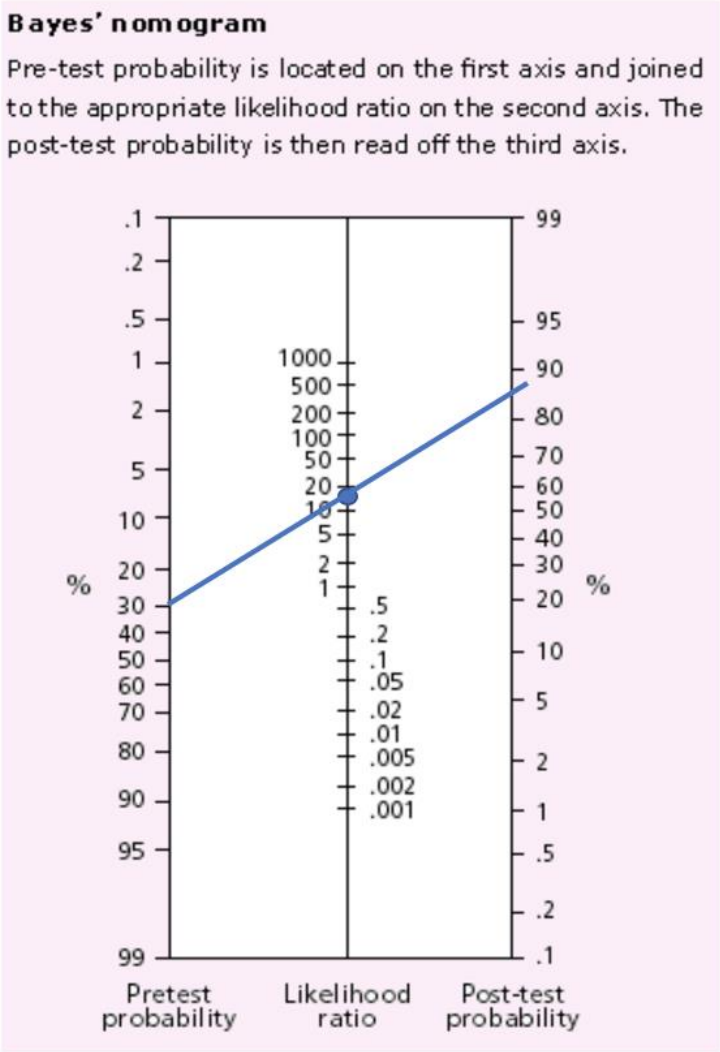

Figure 6. Example Bayes' Nomogram with 30% Pretest Probability and +LR 19.8
Nomogram template from: "Moving beyond sensitivity and specificity: using likelihood ratios to help interpret diagnostic tests" by Attia (2003) in *Australian Prescriber*, p. 112. Copyright 2003 by NPS Medicine.

results)/(number of true-negative + number of false-positive results). Disease prevalence is not a factor in calculating the ROC curve because the true- and false-positive values are calculated from two different subgroups. An ideal test exhibits a ROC curve that passes through the upper left corner (100% true-positive and 0% false-positive). A diagonal 45-degree line drawn from the lower left corner to the upper right corner indicates that the test cannot discriminate between a true- and false-positive result; therefore, tests following or below the diagonal line would be

considered useless or inaccurate. The area under the curve (AUC) on the ROC plot ranges from 0.5 (those tests on or below the diagonal line) for a useless test to 1.0 (those tests in the upper left corner) for a perfect test. The closer the AUC is to 1.0, the higher the diagnostic accuracy of the test. A ROC curve will determine which test is better when two test results are interpreted independently against the same gold standard (Guyatt et al., 1986).

Diagnostic test evaluation. Linnet et al. (2012) discuss their optimal approach for evaluating the accuracy of a diagnostic test. They suggest use of sensitivity, specificity, predictive values, ROC curves and likelihood ratios as statistical measures of accuracy. They further suggest that the diagnostic test evaluation should be carried out in a population that is suspected of having the target disease, preferably with study participants chosen on the basis of predefined symptoms and/or signs of the disease. All participants should be independently tested with both the gold standard reference and the index test with laboratory personnel blinded to the results of either test. Because laboratory test results can vary due to disease progression, it is important to correctly classify each participant as diseased or nondiseased and include the disease stage if possible. When comparing the accuracy of two tests, the statistical measures recommended by Linnet and colleagues are McNemar's test or comparison of AUC values.

Bossuyt et al. (2012) discusses clinical utility as another term for diagnostic test usefulness. Clinical utility refers to the extent that a new diagnostic test improves health outcomes in comparison with the current best alternative test. The four key elements of clinical utility of diagnostic testing include: a) preventing premature death and/or restoring or maintaining functional health, b) guiding treatment and clinical management decisions, c) evaluating tests at the group level, and d) comparison to current standard best practice. Accurately identifying patients with the suspected disease is an essential condition for determining clinical utility;

however, test accuracy can vary due to the spectrum of disease being tested. In studies of diagnostic test accuracy, the authors stress the importance of clearly defining the target condition. Lord et al. (2011) defines the target condition as the "classification of disease you wish to detect" (p. 2). This classification should be determined using the best available evidence based criteria for the disease. Clearly defining the target condition will enhance the validity of diagnostic test accuracy studies.

Because all laboratory tests must be interpreted based on clinical correlation with the ill person, it is important to also include the purpose for ordering laboratory tests in the determination of laboratory test usefulness. Kaplan (1990) describes three purposes for ordering laboratory tests: diagnosis, monitoring therapy, and screening.

Diagnosis would ideally require a test with high sensitivity (positive for most people with the disease) and high specificity (negative for most people without the disease), thus resulting in 100% PPV and NPV values. This rarely occurs with current diagnostic test methods as tests with high sensitivity typically have low specificity and vice versa. However, when ordering tests for diagnostic purposes, there are two main uses. The first use is to absolutely rule out a disease and the second use is to absolutely confirm a disease. In order to absolutely rule out that a person does not have a disease, a test with high sensitivity should be used because a negative result produces a high NPV, meaning the clinician can be confident that a negative test result indicates no disease. For the second use, in order to absolutely rule in or confirm a disease, a test with high specificity should be used because a positive result produces a high PPV, signifying that the clinician can be confident that a positive test result indicates presence of disease.

Monitoring therapy typically means performing the same test repeatedly during a treatment regimen time period. For this purpose, a test that is precise (reproducible) and accurate is

preferable. Sensitivity, specificity, and predictive values have a lower usefulness for this purpose.

Screening would require a test that detects early stages of a disease or condition. The disease may be in an asymptomatic stage, but early detection would prevent further spread or be in a stage that is easily treated or cured. Ideally a test would have 100% sensitivity, specificity, and predictive values. Because these tests do not currently exist, the clinician is faced with a tradeoff between sensitivity and specificity. A test with high sensitivity will also have a high NPV, which means that few false negatives would occur; however, the tradeoff in low specificity means a higher number of false positives would also occur. If screening a population with low prevalence of disease, there will be a number of persons with positive results; but only a small number will actually have the disease (low PPV). A highly specific test will have a higher PPV but a lower NPV, which means that a positive result indicates disease; but more false negatives and missed cases of the disease may be observed. The choice for determining which test to order for disease screening will be between a highly sensitive test that may result in a higher number of additional confirmatory tests or a highly specific test that may result in missing a higher number of diseased persons. The choice will depend on the benefit of early disease detection for the patients as well as the cost and risk of additional diagnostic follow up.

Ideally every diagnostic test would provide the correct result (presence or absence of disease) in 100% of the patients tested. However, currently there are no diagnostic tests that meet the ideal criteria. As sensitivity increases, specificity decreases and vice versa. Predictive value of laboratory tests is influenced by disease prevalence within the tested community with a higher disease prevalence associated with a higher predictive value of a positive test. Galen and Gambino (1975) provide suggestions shown in Table 9 for determining which statistical method

Table 9

Attributes for Determining Preferable Statistical Measure of a Diagnostic Test

Preferable statistical measure	**Attribute**
Highest Sensitivity (100%)	• Serious disease that should not be missed AND • Disease is treatable AND • False-positive results do not lead to serious psychological or economic trauma to the patient
Highest Specificity (100%)	• Serious disease but is *not* treatable or curable AND • Knowledge of absence of disease has psychologic or public health value AND • False-positive results can lead to serious psychologic or economic trauma to patient
High Predictive Value	• Treatment of false-positive might have serious consequences
Highest Efficiency (Accuracy)	• Serious disease that is treatable AND • False-positive and false-negative results are essentially equally serious or damaging

Adapted from: *Beyond Normality The Predictive Value and Efficiency of Medical Diagnoses* by Galen & Gambino (1975). New York, NY: John Wiley & Sons, pages 50-51. Copyright 1975 by John Wiley & Sons.

is preferred when deciding which laboratory test to order. According to Galen and Gambino (1975), for serious diseases that are treatable and false-positive results do not cause serious harm, it is preferable to use a test with high sensitivity. The consequences of missing the disease outweigh the consequences of a false positive test result, (e.g., treatable infectious diseases such as STDs can have serious public health consequences). For serious diseases that are not treatable, it is preferable to use a test with high specificity. In this case, false positive results can cause serious psychological or economic trauma to the patient (e.g. detection of multiple sclerosis). For diseases where treatment can have serious consequences, it is essential to use a test with high predictive values (e.g. detection of occult cancers can lead to disastrous consequences if treatment is given and the cancer is not there). For serious diseases that are treatable, it is preferable to use a test with high efficiency or accuracy. In this case false-positive

and false-negative results have essentially the same serious or damaging consequences (e.g. myocardial infarction or diabetes).

Cost analysis. Cost should also be considered when determining diagnostic laboratory test usefulness. Hernandez (2003) discusses the need for further studies that address both clinical efficacy and cost effectiveness of laboratory testing. He provides a review of a model developed by John Pfister, MS, (AAM), the acting director of the Wisconsin State Laboratory of Hygiene (WSLH) for assessing public health laboratory testing. The model is composed of three partly overlapping circles that represent assay performance, epidemiology, and costs. The assay performance circle contains the familiar laboratory performance criteria of sensitivity, specificity, quality assurance, and turnaround time. However, as laboratories are being required to do more with less as budgets are cut, it is imperative that directors and managers identify the value-added component for a given laboratory test. This involves looking outside the laboratory to determine the cost effectiveness, cost benefit, or cost utility of their operation. The epidemiology factors of disease prevalence and risk indicators within a tested population can affect how useful a laboratory test is for ruling in or ruling out a disease. The direct and indirect costs of laboratory testing also have an impact on assessing the value of laboratory tests. Pfister suggests that when developing testing strategies for newer tests the factors of assay performance, epidemiology, and costs should all be considered. Hernandez (2003) states that healthcare consumers, both employers and patients, are concerned about the quality of health care as well as reasonable costs. Pfister's schematic model (Figure 7) could provide a basis for laboratories to include not only laboratory test costs when determining the value of a test, but also the downstream costs of developing cost effective algorithms, identifying the costs associated with

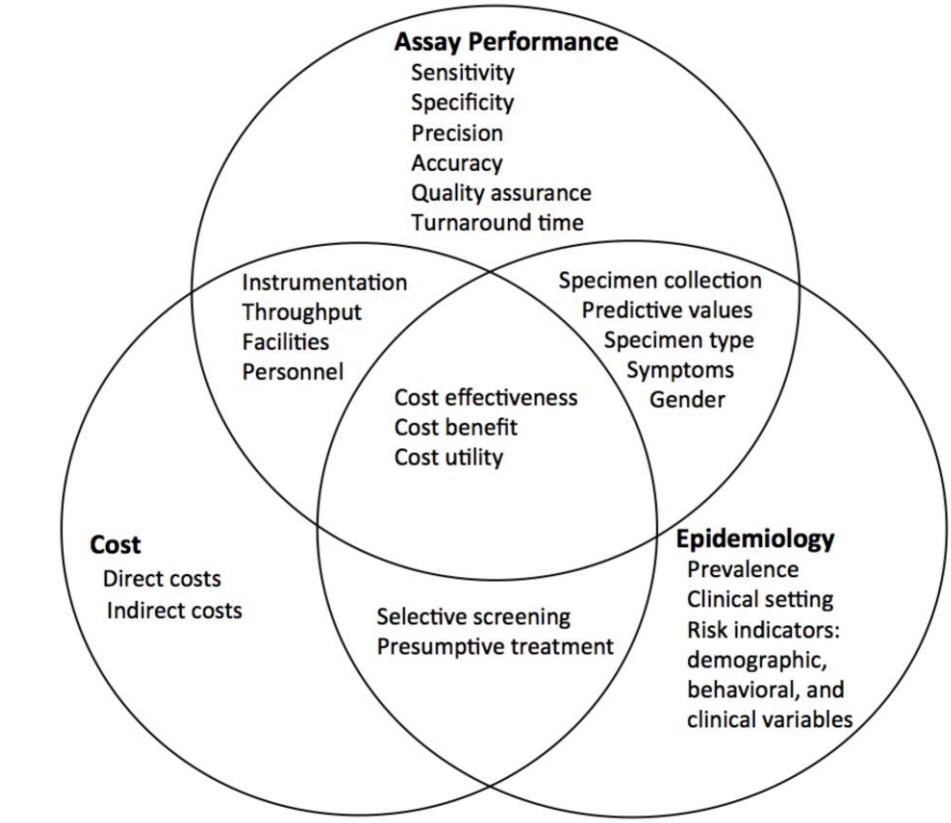

Figure 7. Schematic Model for the Appropriate Selection and Use of Laboratory Tests. Reprint from "Public health applications of new laboratory technologies for communicable diseases", by Pfister, J. (1997), *Wisconsin State Laboratory of Hygiene Results*, Spring, p. 7. Copyright 1997 John Pfister. Reprinted with permission.

false-negative and false-positive test results, and providing consultation services for suggesting effective testing strategies.

Owusu-Edusei, Peterman and Ballard (2011) performed a cost effectiveness analysis of the traditional and reverse syphilis screening algorithms using a cohort decision analysis to estimate the expected costs and effects (including follow-ups and overtreatment) of the two algorithms. A cohort of 200,000 individuals, which included 1,000 current infections and 10,000 previous infections were analyzed. The estimated cost effectiveness ratios were $1,671 for the reverse algorithm and $1,621 for the traditional algorithm. The study assumed a low prevalence rate of 0.5% with 5% of the population previously infected. The low prevalence rate implies that the

positive predictive value of the syphilis test would also be low. Using those assumptions, the higher cost of the reverse algorithm test could be due to the inability of syphilis EIA tests to differentiate active from prior disease, thus leading to additional testing. The authors suggest that further studies should be conducted to assess the cost effectiveness of the algorithms in higher prevalence populations. The authors used the Medicare reimbursement rate ($5 for RPR and $6 for EIA) rather than a laboratory calculated test cost to determine cost effectiveness.

Chuck, Ohinmas, Tilley, Singh and Jacobs (2008) performed a cost effectiveness analysis of the traditional and reverse algorithms using a cohort simulation model constructed with public health data from the Alberta Health Ministry. Labor was considered a cost driver in the delivery of syphilis testing services, which included laboratory testing and non-laboratory activities, such as community search for patients, contacting patients, patient treatment, chart review, and follow-up. The study was performed in prenatal (syphilis prevalence 0.08%) and nonprenatal (syphilis prevalence 1.9%) populations. The study determined that the overall (prenatal and nonprenatal populations) incremental cost effectiveness ratio for the reverse algorithm was less costly and more effective than the traditional algorithm. The reverse algorithm would save $461 (Canadian) to produce one additional correct diagnosis. This study used laboratory test costs based on internal laboratory accounting data. The study stated that cost per test included kit, labor and supplies, but did include the labor calculation details nor identify if indirect costs were included in the cost. The cost per test for RPR and EIA was listed as $2.22 and $4.75, respectively.

Owusu-Edusei, Koski and Ballard (2011) performed a cohort decision analysis to determine the health and economic outcomes of the traditional and reverse algorithms in both low and high prevalence settings. The study concluded that both algorithms detected and treated the same

number of individuals; however, the traditional algorithm was more cost effective in the low prevalence setting ($1,400 vs $1,500 per adverse outcome prevented) and more cost saving ($102,000 vs $84,000) in the high prevalence setting. The cost difference is likely due to the additional confirmatory tests that must be performed with the reverse algorithm. In this study, the authors varied the laboratory testing costs between $1 and $24 to identify a threshold cost. They identified that in a low prevalence setting, when the treponemal test cost was below $5.80, the reverse algorithm was more cost effective than the traditional algorithm. In high prevalence settings, when the treponemal cost was below $1.80, the reverse algorithm was more cost saving than the traditional algorithm.

Berry & Loeffelholz (2016) performed a cost analysis of the potential savings if the TP-PA confirmatory test was only performed on Bioplex Syphilis IgG MBIA reactive samples with an AI between 1.1 and 7.9. Based on results of their study, reactive samples with AI values above 7.9 were 100% confirmed positive by TP-PA. The cost analysis was performed using laboratory generated time studies for the analytical labor component and their reagent costs. The study calculated the cost savings using their current modified reverse algorithm (reactive or equivocal Bioplex Syphilis IgG reflexed to both RPR titer and TP-PA) versus an algorithm based on performance of TP-PA testing only on samples with an AI of 1.1 and 7.9. Samples with AI value equal to or above 8.0 would be assumed to be confirmed positive and additional TP-PA testing, as required by their modified reverse algorithm, would not be performed. The annual number of TP-PA tests performed using their modified algorithm was 1,786. There were 1,155 samples with AI values equal to or greater than 8.0. If those samples were assumed confirmed and not tested further, then there would be a 65% (1155/1786) reduction in TP-PA confirmatory testing. The laboratory labor cost of performing the 1,786 tests was $1,459.43 with reagent cost

of $5,994.32. The laboratory labor cost of performing 631 TP-PA tests was $510.80 with reagent cost of $2,118.29. The labor cost savings using the AI as a benchmark for reflex confirmatory testing would be $948.63 with reagent cost savings of $3,876.02 for a total annual savings of $4,824.65.

Other authors have used only the reagent costs to discuss pricing differences in the reverse and traditional syphilis algorithms. Sena, et al. (2010) states that the difference in reagent cost for nontreponemal RPR ($0.25) versus treponemal EIA ($5) along with test volume and labor costs are limiting factors in program decisions to adopt the reverse algorithm. Binnicker and Loeffelholz (2011) discuss the advantages of automated syphilis testing in reducing "hands on" staff time and increasing sample throughput. They further emphasize that labor is the laboratory's highest direct cost and recognize that laboratories with the highest test volume could benefit the most from automated syphilis testing. However, they also mention that even lower volume laboratories could benefit if multiple immunoassays could be combined on one instrument. This is especially true for diseases of public health significance, with many vendors automating HIV, syphilis, and hepatitis testing onto one instrument so that all tests could be performed on the same sample at the same time.

The previous studies provided a wide range of laboratory test costs; however, there is little backup evidence as to how the costs were calculated. When determining the cost effectiveness of a test method, it is important to incorporate well-defined laboratory test costs. Wilkinson, Dilts, Woolf, and Lifshitz (2011) define and identify the following costs for determining a laboratory test price:

- Cost: the supply, labor and overhead dollars spent on a product or service.
- Direct cost: expenses that can be traced back to a billable test (end product). Expenses

include reagents, consumables, equipment, and "hands on" staff time. Equipment may be rented, leased or purchased and direct costs should include any ongoing service agreements. Rental or lease payments may be paid on a monthly basis or bundled into the reagent kit purchase price. Depreciation costs for owned laboratory equipment that meets the capital cost criteria should be included in direct costs (Wawrzynski & Hall, 2013).

- Variable costs: expenses that change proportionately with test volume. Increased volume equals increased reagent costs.
- Fixed costs: expenses that do not vary with test activity. Fixed costs include staff salaries, fringe benefits, analyzer service, proficiency test service, or rent.
- Indirect costs: costs not directly related to the test being performed and usually calculated by administration using an allocation formula. Sometimes referred to as overhead, these costs can include utilities or waste disposal, licenses, secretarial support, supervisory or administration salaries, human resources, or information technology support (Wawrzynski & Hall, 2013).
- Salary costs: includes salary and fringe benefits. Salary can account for 60% to 80% of the laboratory budget. Fringe benefits include Social Security, health insurance, or pension plans and can add 16% to 28% to the base salary.
- Operating costs: expenses to produce a product; often a one-time use such as reagents, consumables, or electricity. Once the item has been used, it is of no further production value.
- Capital costs: items such as computers, analytical equipment, or physical laboratory facility that can be re-used again after production. There are three criteria that must be

met in order to qualify as a capital item:

- Time: item must have a useful life of longer than one year.
- Price: minimum dollar amount, usually $1,000 to $5,000, designated by the institution for qualification as a capital item.
- Purpose: usually replacement of old equipment or purchase of new equipment to use for new methods or services.

As proposed by Wilkinson, et al., (2011), microcosting is a mechanism used to derive the cost of producing a laboratory test. The total direct labor and supply cost of producing a single test is first determined and that cost is used as a starting point to determine the fully loaded (direct and indirect) cost based on test volume.

Most laboratory testing is performed in batches or continuous "runs" of multiple samples over one or more shifts. A run is determined by many factors, including clinical expectations, test type, the laboratory operating hours, laboratory workflow, and the manufacturer's FDA approved or cleared test procedure. A run typically includes all reagent, consumable, quality control (QC), and calibration costs necessary to produce a test result. A run may be performed all at once or many times during a shift. A batch may be a group of tests that are performed after the initial controls are performed and multiple batches could be performed during a shift. When microcosting a test, it is important to determine the workflow, control, and repeat requirements as they may impact the labor and supply costs. According to Wilkinson, et al. (2011), the cost per reportable result (CPRR) "distributes the total direct costs of a run over the patient 'reportable' results for that run. Testing efficiency is defined as the total reportable patient results/total test results. Thus, the more repeats and controls are performed, the lower is the efficiency, and the higher is the CPRR. As the testing efficiency increases, the CPRR decreases"

(p. 143). When a test is performed using an automated analyzer, the largest expenditure of labor is in the analyzer set up. Once set up with reagents and samples and started, the technologist can "walk away" as the analyzer completes the testing without further intervention. This allows the technologist time to perform other tasks. Typically, the only cost for adding an additional test to an already operating automated analyzer is the reagent cost; however, there may be a minimal labor expense for inserting a bar coded sample onto an automated analyzer. All other costs, such as controls or equipment are fixed costs. If the test is performed manually, then there may be additional technologist time during each test step, that time would be included in the test cost. In calculating labor costs, the staff time includes only "hands on" time, it does not include test incubation times for either an automated or manual assay. Because prior literature review did not provide clear detail regarding determination of syphilis test costs, the microcosting template provided by Wilkinson, et al., (2011) was used to calculate the total direct cost, cost per reportable result, testing efficiency, and fully loaded cost of the initial syphilis screening tests, MacroVue RPR and Captia™ Syphilis IgG tests, and the confirmatory Serodia TP-PA test as performed at the dissertation study western region public health laboratory. The following assumptions were used to complete the microcosting tables:

- Reagent and consumable costs for MacroVue RPR, Captia™ Syphilis IgG, and Serodia TP-PA were based on 2015 vendor negotiated prices with the western regional laboratory where the dissertation study was performed. The EVOLIS automated immunoassay analyzer used to perform the Captia™ Syphilis IgG tests was leased from the vendor under a cost per reportable contract. An additional vendor negotiated charge of $0.11 per test was included in the price of each Captia™ Syphilis reagent kit when purchased and is reflected in the price used for the microcosting calculations. Under this contract, the

vendor maintained ownership of the equipment and there was no additional charge for repair or preventive maintenance on the instrument during the contract time period.

- Medical technologist salary was based on the US Department of Labor Occupational Outlook Handbook (retrieved from https://www.bls.gov/ooh/healthcare/medical-and-clinical-laboratory-technologists-and-technicians.htm#tab-5). Medical technologist median annual wage as of May 2015 was $60,520 which calculates to hourly pay of $29.16
- Fringe benefits were calculated based on the range (16% to 28%) provided by Wilkensen, et al (2011). The range average is 22%, which calculates to $6.42 per hour ($29.16*0.22).
- Indirect costs were calculated based on the estimate provided by Wilkinsen, et al. (2011) as 2.5 x direct cost.
- Direct analytical labor time in minutes was based on time studies performed at the western regional laboratory where the dissertation testing was performed. Time studies were based on the standard workflow in use at the laboratory. Standard syphilis workflow included performing additional quality assessment checks by a second technologist at multiple steps in the analytical process and following the FDA approved manufacturer's instructions as written. Direct labor time did not include test incubation times which do not typically require "hands on" staff time.
- Captia™ Syphilis IgG EIA: automated test method using EVOLIS immunoassay analyzer. A run consisted of up to four 96-well reagent plates with five controls (four internal and one external) included in each plate. One run was performed per shift. Once the run was initiated, additional samples were not added.

- MacroVue RPR: manual test method. A run consisted of one 10 spot diagnostic test card with 3 external controls on a separate control card. Manufacturer instructions state controls must be performed once per shift. Following successful control performance, batches of 10 samples per diagnostic test card were performed continuously using the same reagent throughout the shift.
- Fujirebio Serodia TP-PA: manual test method. A run consisted of one 96 well plate with two external controls per plate. Each sample and control required four wells for test performance, therefore a maximum of 22 unknown patient samples and two external controls were performed per plate. One run was performed per shift.

Tables 10 and 11 provide a basis for calculating the cost of initial syphilis screening tests using either the MacroVue RPR or Captia™ Syphilis IgG with various test volumes. The MacroVue RPR used a 10-spot diagnostic card and the most efficient use of the card was to maximize the number of samples tested on the card, thus running all 10 samples at the same time in a batch. The Captia™ Syphilis IgG test used a 96-well plate and required use of five controls on each reaction plate; therefore, the maximum number of patients that could be performed on a single plate was 91 samples. Up to four 96-well plates with a maximum of 364 samples could be performed in one automated analyzer run. Performing less than the maximum number of samples would have an impact on test cost and efficiency.

Table 12 provides a basis for calculating the cost of confirmatory TP-PA testing. The TP-PA test method used a 96 well test plate for each batch run. Each sample required four wells to perform each analysis and the manufacturer required two external controls per day of use or batch run. The maximum number of samples and controls that could be performed per plate was 24. The test method could be performed with less than the maximum number of samples, which

Table 10

Microcosting: MacroVue RPR (Manual Method)

10.A. Microcosting: Manual run of one patient reportable result (RPR) A run consists of one 10 spot RPR diagnostic test card with 3 external controls on separate card. Per manufacturer instructions, controls performed once per shift.			
Direct Labor: Determine the total "hands on" time in minutes required to perform a manual "run" of one patient. Assume labor cost of $35.58/hour ($29.16 + $6.42 fringe)			
Category	**Minutes**		**Expense**
Test set up (specimens and reagents)	4		
Result analysis	2		
Result review and verification	2		
Result documentation	1		
Total direct labor ($35.58/60 minutes = $0.59/minute)	9		**$ 5.31**
Direct Supplies: List of consumables needed to perform the test. Note: 4 tests (1 sample and 3 controls) needed to produce 1 patient reportable result			
Category	**Unit cost**	**Units**	**Expense**
Reagent kit includes reagent, dispenstir and diagnostic cards ($235/500 tests)	$0.47	4	$ 1.88
Controls: external card ($51/30 controls)	$1.70	3	$ 5.10
Total direct supplies			**$ 6.98**
Total direct costs (direct labor + direct supplies)			**$12.29**
Cost per reportable result (total direct cost/reportable results)		$12.29/1	**$12.29**
Testing efficiency (reportable results/tests)		1/4	**25.0%**

10.B. Microcosting: Manual batch run of 100 patient reportable results (RPR)			
Direct Labor: Determine the total "hands on" time in minutes to perform batch runs of 100 patients. Batches of 10 samples per diagnostic card can be performed continuously using the same reagent throughout the shift. Note: additional 30 seconds (0.5 minutes) for each additional sample added to each manual step time determined in Table 10A.			
Category	**Minutes**		**Expense**
Test set up: 0.5x99 specimens = 49.5+4 minutes	53.5		
Result analysis: 0.5x 99 specimen= 49.5+2 minutes	51.5		
Result review and verification: 0.5x99 specimen=49.5+2 minutes	51.5		
Result documentation: 0.5x99 specimen=49.5+1 minute	50.5		
Total direct labor ($0.59/minute)	**207**		**$122.13**
Direct Supplies: Note: 103 tests (100 samples and 3 controls) needed to produce 100 patient reportable results. Fixed costs (controls) are spread over more than 1 sample			
Category	**Unit cost**	**Units**	**Expense**
Reagent kit ($235/500 tests)	$0.47	103	$ 48.41
Controls ($51/30 controls)	$1.70	3	$ 5.10
Total direct supplies			**$ 53.51**
Total direct costs (direct labor + direct supplies)			**$175.64**
Cost per reportable result (total direct cost/reportable results)		$175.64/100	**$ 1.76**
Testing efficiency (reportable results/tests)		100/103	**97.1%**

10.C. Fully Loaded Cost: MacroVue RPR for 100 sample batch run			
Category	**Unit cost**		**Expense**
Direct cost per reportable result for 100 sample batch run	$1.76		
Indirect cost (estimated at 2.5 x direct costs)	$4.40		
Fully loaded cost per reportable result (direct + indirect costs)			**$ 6.16**

Table 11

Microcosting: Captia™ Syphilis IgG (EVOLIS Analyzer)

11.A. Microcosting: Automated run of one reportable result (Syphilis IgG) A run consists of up to four 96-well reagent plates with 5 controls (4 internal and 1 external) included in each plate. One run performed per shift			
Direct Labor: Determine the total "hands on" time in minutes required to perform an automated "run" of one patient. Assume labor cost of $35.58/hour ($29.16 + $6.42 fringe)			
Category	**Minutes**		**Expense**
Prepare specimens	2		
Prepare reagents	10		
Prepare instrument	10		
Computer and worksheet setup	5		
Documentation of results/QC/maintenance	10		
Clean up	10		
Total direct labor ($0.59/minute)	47		**$ 27.73**
Direct Supplies: List of consumables needed to perform the test. Note: 6 tests (1 sample and 5 controls) needed to produce 1 patient reportable result			
Category	**Unit cost**	**Units**	**Expense**
Reagent kit ($1315/960 tests) includes $0.11 reagent rental charge/test for analyzer use	$1.37	6	$ 8.22
300 ul pipette tip ($92/960) 1 tip/sample	$0.10	6	$ 0.58
1100 ul pipette tip ($141.75/960) 5 reagent tips/plate	$0.15	5	$ 0.74
Internal controls (included in reagent kit)	$0.00	4	$ 0.00
External control ($84/12 controls)	$7.00	1	$ 7.00
Total direct supplies			**$ 16.53**
Total direct costs (direct labor + direct supplies)			**$ 44.26**
Cost per reportable result (total direct cost/reportable results)		$44.26/1	**$ 44.26**
Testing efficiency (reportable results/tests)		1/6	**16.7%**

11.B. Microcosting: Automated run of 91 reportable results (Syphilis IgG)			
Direct Labor: Determine the total "hands on" time in minutes required to perform an automated "run" of 91 patients (one full test plate). Note:15 seconds (0.25 minutes) added to sample preparation step from Table 11A for additional samples inserted on test plate.			
Category	**Minutes**		**Expense**
Prepare specimens: 0.25 x 90 = 22.5 + 2 minutes	24.5		
Prepare reagents	10		
Prepare instrument	10		
Computer and worksheet setup	5		
Documentation of results/QC/maintenance	10		
Clean up	10		
Total direct labor ($0.59/minute)	69.5		**$ 41.00**
Direct Supplies: Note: 96 wells (91 samples and 5 controls) needed to produce 91 patient reportable results per plate. Fixed costs (controls and reagent tips) are spread over more than 1 sample			
Category	**Unit cost**	**Units**	**Expense**
Reagent kit ($1322/960 tests)	$1.37	96	$131.50
300 ul pipette tip ($92/960) 1 tip/sample	$0.10	96	$ 9.20
1100 ul pipette tip ($141.75/960) 5 reagent tips/plate	$0.15	5	$ 0.74
Internal controls (included in reagent kit)	$0.00	4	$ 0.00
External control ($84/12 controls)	$7.00	1	$ 7.00
Total direct supplies			**$148.44**
Total direct costs (direct labor + direct supplies)			**$189.44**

Table 11. *Continued*

Cost per reportable result (total direct cost/reportable results)		$189.44/91	**$ 2.08**
Testing efficiency (reportable results/tests)		91/96	**94.8%**

11.C. Fully Loaded Cost: Captia™ Syphilis IgG EIA for 91 sample batch run			
Category	**Unit cost**		**Expense**
Direct cost per reportable result for 91 sample batch run	$2.08		
Indirect cost (estimated at 2.5 x direct costs)	$5.20		
Fully loaded cost per reportable result (direct + indirect costs)			**$ 7.28**

Table 12

Microcosting: Fujirebio Serodia TP-PA (Manual Method)

12.A. Microcosting: Manual run of one patient reportable result (TP-PA) A run consists of one 96 well plate with 2 external controls per plate. Each sample and control require 4 wells for test performance, therefore a maximum of 22 unknown patient samples and 2 external controls can be performed per plate. Per manufacturer instructions, external controls performed once per day or batch run.			
Direct Labor: Determine the total "hands on" time in minutes required to perform a manual "run" of one patient. Assume labor cost of $35.58/hour ($29.16 + $6.42 fringe)			
Category	**Minutes**		**Expense**
Test set up (specimens and reagents) – includes dilutions	4.5		
Result analysis	2		
Result review and verification	2		
Result documentation	1		
Total direct labor ($0.59/minute)	9.5		**$ 5.60**
Direct Supplies: List of consumables needed to perform the test. Note: 3 tests (1 sample and 2 controls) needed to produce 1 patient reportable result			
Category	**Unit cost**	**Units**	**Expense**
Reagent kit includes reagents, external controls ($239.94/100 tests)	$2.40	3	$ 7.20
Test plate ($142.00/100 plates)	$1.42	1	$ 1.42
Pipette tips for individual pipettor ($100.00/960 tips) 2 tips per sample or control	$0.10	6	$ 0.60
Pipette tip for multichannel pipettor ($113.00/960 tips) 1 per tip per sample or control	$0.12	3	$ 0.36
Total direct supplies			**$ 9.58**
Total direct costs (direct labor + direct supplies)			**$15.18**
Cost per reportable result (total direct cost/reportable results)		$15.18/1	**$15.18**
Testing efficiency (reportable results/tests)		1/3	**33.3%**

12.B. Microcosting: Manual batch run of 22 patient reportable results (TP-PA)			
Direct Labor: Determine the total "hands on" time in minutes to perform batch run of 22 patients. Note: additional 30 seconds (0.5 minutes) for each additional sample added to each manual step time determined in Table 12A.			
Category	**Minutes**		**Expense**
Test set up: 0.5x21 specimens = 10.5+4.5 minutes	15.0		
Result analysis: 0.5x21 specimen= 10.5+2 minutes	12.5		
Result review and verification: 0.5x21 specimen=10.5+2 minutes	12.5		
Result documentation: 0.5x21 specimen=10.5+1 minute	11.5		
Total direct labor ($0.59/minute)	**51.5**		**$ 30.38**

Table 12. *Continued*

Direct Supplies: Note: 24 tests (22 samples and 2 controls) needed to produce 22 patient reportable results. Fixed costs (plate and tip) are spread over more than 1 sample			
Category	**Unit cost**	**Units**	**Expense**
Reagent kit includes reagents, external controls ($239.94/100 tests)	$2.40	24	$57.60
Test plate ($142.00/100 plates)	$1.42	1	$ 1.42
Pipette tips for individual pipettor ($100.00/960 tips) 2 tips per sample or control	$0.10	48	$ 4.80
Pipette tip for multichannel pipettor ($113.00/960 tips) 1 tip per sample or control	$0.12	24	$ 2.88
Total direct supplies			**$ 66.70**
Total direct costs (direct labor + direct supplies)			**$ 97.08**
Cost per reportable result (total direct cost/reportable results)		$97.08/22	**$ 4.41**
Testing efficiency (reportable results/tests)		22/24	**91.7%**
12.C. Fully Loaded Cost: Serodia TP-PA for 22 sample batch run			
Category	**Unit cost**		**Expense**
Direct cost per reportable result for 22 sample batch run	$4.41		
Indirect cost (estimated at 2.5 x direct costs)	$11.02		
Fully loaded cost per reportable result (direct + indirect costs)			**$ 15.43**

would impact test cost and efficiency. A batch run could be performed daily or only on specific days which would maximize the number of samples in the batch.

Table 13 provides a summary of the costs associated with various initial syphilis screening workflows based on test volume. The initial direct and fully loaded cost of one reportable test is listed for each test method, then different runs or batch volumes were calculated based on the values in Tables 10-12 (calculations not shown). A run of one patient sample had the highest costs and lowest efficiency for both test methods. Additional calculations were performed for a run of 10 patient samples, which maximizes the RPR diagnostic card but only allows for partial use of the Syphilis IgG reaction plate. With the increase in test volume, the direct cost decreased rapidly from the one sample test scenario for both test methods ($12.29 vs $2.71 for RPR and $44.26 vs $5.90 for Syphilis IgG). Additionally, the efficiency increased for both methods when the sample volume increased (25% vs 76.9% for RPR and 16.7% vs 66.7% for Syphilis IgG). These calculations could reflect the costs and efficiency associated with a low volume laboratory

Table 13

Comparison of MacroVue RPR and Captia™ Syphilis IgG Costs, Efficiency, Time

One reportable result	**MacroVue RPR**	**Captia™ Syphilis IgG**
	1 reportable result	1 reportable result
	1 partial diagnostic card	1 partial reaction plate
Direct cost	$12.29	$ 44.26
Fully loaded cost (direct+indirect)	$43.01	$154.91
Efficiency	25.0%	16.7%
Labor time	9 minutes	47 minutes
Batch or run reportable results	**MacroVue RPR**	**Captia™ Syphilis IgG**
	10 reportable results	10 reportable results
	1 full diagnostic card	1 partial reaction plate
Direct cost	$2.71	$ 5.90
Fully loaded cost (direct+indirect)	$9.48	$20.65
Efficiency	76.9%	66.7%
Labor time	27 minutes	49 minutes
Batch or run reportable results	**MacroVue RPR**	**Captia™ Syphilis IgG**
	100 reportable results	91 reportable results
	10 full diagnostic cards	1 full reaction plate
Direct cost	$1.76	$2.08
Fully loaded cost (direct+indirect)	$6.16	$7.28
Efficiency	97.1%	94.8%
Labor time	207 minutes	69 minutes
Batch or run reportable results	**MacroVue RPR**	**Captia™ Syphilis IgG**
	182 reportable results	182 reportable results
	18 full + 1 partial diagnostic cards	2 full reaction plates
Direct cost	$1.71	$1.85
Fully loaded cost (direct+indirect)	$5.98	$6.47
Efficiency	98.4%	97.3%
Labor time	371 minutes	92 minutes

running only 10 samples per day or batch. Additional calculations were performed to determine the effect of further increases in test volume such as might be seen in high volume laboratories. With an increase in test volume, the direct costs decreased for both test methods with the lowest direct costs for RPR at $1.71 and Syphilis IgG at $1.85 when 182 samples were performed in a run. The efficiency also increased with RPR at 98.4% and Syphilis IgG at 97.3%. The total direct and fully loaded cost differences between the RPR and Syphilis IgG also shrank as the test

volume increased. At the 182 sample level, there was only a 49 cent difference between the fully loaded cost of RPR ($5.98) and Syphilis IgG ($6.47). This emphasizes the requirement that test volume must be considered when comparing the cost of syphilis testing methods.

However, cost and efficiency are not the only factors to look at in the calculations in Table 13. While the costs decreased for both test methods, the labor time increased for the RPR test method with 371 minutes of staff time required to perform 182 samples versus 92 minutes for the Syphilis IgG. This time difference highlights the advantage of the automated analyzer in reducing staff time as the technologist performing batches of manual RPR testing will spend 371 minutes (6 hours 11 minutes) hands on time at the bench, while the technologist performing the automated testing will spend 92 minutes (1 hour and 32 minutes) hands on time and can utilize the other four plus hours to perform other laboratory related tasks. There would be further labor savings if the automated analyzer could simultaneously perform other STD tests such as HIV on the same sample that was being tested for syphilis.

The cost effectiveness analysis performed by Owusu-Edusei, Koski and Ballard (2011) identified that in a low prevalence setting, when the treponemal test cost was below $5.80, the reverse algorithm was more cost effective than the traditional algorithm. In high prevalence settings, when the treponemal cost was below $1.80, the reverse algorithm was more cost saving than the traditional algorithm. Based on the calculations provided in Table 13, there was a difference in the direct versus fully loaded costs for each test method. If only the direct cost method was used to determine cost effectiveness, then the reverse algorithm using the Syphilis IgG test method at $1.85 for 182 reportable results met the cost effectiveness criteria for low prevalence areas ($5.80) and was very close for the high prevalence areas ($1.80). If the fully loaded cost at $6.47 was used, the Syphilis IgG test method would not meet the cost

effectiveness level for either low or high prevalence settings. This demonstrates the importance of using clearly defined costs for all syphilis test methods when cost effectiveness analyses are performed. The fully loaded cost included an indirect calculation of 2.5 times the direct cost which significantly increases the cost of performing any laboratory test.

Usefulness measure summary. As indicated in the previous paragraphs, various authors have described multiple statistical measures for determining the usefulness of diagnostic laboratory tests. Additionally, many authors use the terms accuracy and usefulness interchangeably. For the purposes of this dissertation, accuracy was included as a component of usefulness, as no single test can provide a definitive answer of diagnostic test usefulness. The following statistical measures were used in the study to define test usefulness: sensitivity, specificity, predictive values, likelihood ratios, accuracy, and/or ROC analysis. Because test feasibility is also a usefulness measure, test cost was included in the discussion. For public health purposes, a highly sensitive test that can determine syphilis infection in an early disease stage has implications not only for treatment of the infected patient but also for preventing disease spread within the community. A highly accurate test is necessary to reduce the financial and emotional consequences of a false test result. Strong likelihood ratios provide the healthcare provider with information for determining the clinical diagnosis for a specific patient. For patients seeking care at STD public health centers, syphilis tests that can provide the best combination of these usefulness measures would be most desirable (Aamir & Hamilton, 2014; Galen & Gambino, 1975; Hulley et al., 2013; Vihinen, 2012).

Public Health and Syphilis

Syphilis has serious public health consequences and is considered a sentinel public health event by the CDC. Syphilis infection is a reportable disease in all 50 states. The CDC as well as

local and state health authorities use the data to monitor syphilis infection rates. In 1999, the CDC launched the National Plan to Eliminate Syphilis. The CDC determined this was a unique opportunity to eliminate the disease because syphilis rates were at an all-time low, treatment and testing were effective and easily accessible, and high incidences of disease were limited to specific geographic areas. However, the CDC recognized that due to the worldwide spread of syphilis, it would not be feasible to set a goal of eradicating syphilis infections in the US; therefore, the goal was to eliminate or reduce the incidence of syphilis in specific geographic areas to zero. Because syphilis typically affects isolated groups involved in high-risk activities such as drug use, trading sex for money or drugs, unprotected anal intercourse, and multiple sex partners, the National Plan targeted a defined geographic area and utilized deliberate efforts to reduce syphilis infection (CDC, 1999).

The National Plan proposed the following five goals:

- Enhanced surveillance for detection, monitoring and data analysis of cases and contacts.
- Strengthened community involvement and partnerships in both the public and private sectors and development of locally relevant syphilis elimination plans.
- Rapid outbreak response to quickly interrupt syphilis transmission.
- Expanded clinical and laboratory services including multi-level activities to provide access for persons infected or exposed to syphilis.
- Enhanced health promotion to develop multi-level preventive sexual and health care behavior activities for at risk groups.

CDCs rationale for implementation of the 1999 National Plan to Eliminate Syphilis was based on the public health importance of syphilis infection as well as the biological and epidemiological feasibility for elimination. High rates of syphilis infection within a community act as the “canary

in the coal mine", indicating a breakdown in public health services. By enhancing public health efforts to eliminate syphilis, control of other infectious diseases, such as HIV, within the affected community and improvement of reproductive health services will also be achieved. Syphilis elimination has an impact on fetal and maternal health through reduced stillbirths and congenital syphilis cases. Based on 1999 dollars, the direct and indirect cost of syphilis infection in the US was $996 million each year; thus, eliminating syphilis would allow that funding to be used for other healthcare purposes. Multiple biological characteristics of syphilis infection made elimination feasible. These characteristics included inexpensive and effective antibiotic treatment with penicillin, easily accessible diagnostic testing, lack of nonhuman reservoirs, and a long incubation period, which provided sufficient time for case finding and epidemiological interventions to prevent disease transmission (CDC, 1999).

In 1999, the time looked right for launching an all-out public health campaign to eliminate syphilis. CDC reported that syphilis occurs in seven to ten year cycles and the 1999 plan hoped to catch the end of a 10-year cycle to knock out syphilis infection. The 1990 syphilis epidemic had the highest rates of primary and secondary (P&S) syphilis infection (20.3/100,000 population) in 40 years. By 1998, the P&S syphilis rate had dramatically decreased to 2.6/100,000 population. Congenital syphilis rates also showed a significant decline from a high of 107.3/100,000 live births in 1991 to 20.6/100,000 live births in 1998. CDC theorized that the dramatic decreases in syphilis infection were due to multiple factors including the STD public health prevention efforts to stop the 1990 syphilis epidemic along with additional public health activities to prevent HIV transmission and substance abuse. In 1998, the geographic concentration of syphilis was highest in the Southern states with almost 75% of the rest of the nation's counties reporting zero syphilis cases (CDC,1999).

Significant federal dollars were allocated to fund the syphilis elimination effort. CDC provided $107 million dollars in funding to High Morbidity Areas (HMAs) to implement the syphilis elimination efforts. Seven years later, the CDC reviewed the results of their efforts in the 2006 Syphilis Elimination Effort (SEE) report. The report focused on the gains that had been made since 1999 in reducing syphilis incidence in selected groups. There were reductions in syphilis infections in women and newborn infants with congenital syphilis rates dropping by 92% from 107.3/100,000 live births in 1991 to 8.8/100,000 live births in 2004. Syphilis infections in women also declined dramatically from 2.0/100,000 to 0.8/100,000 population during the 1991-2004 time period. Racial disparities were also reduced during that time period with a decrease in African American syphilis rates from 14.3/100,000 to 8.9/100,000 population. However, despite the best public health efforts, syphilis had not been eliminated and was again increasing, with 60% of new infections among the MSM population in 2006 (CDC, 2006).

Current review of P&S syphilis infection numbers shows a continued increase in US rates. The 2016 P&S syphilis rate (8.7/100,000 population) was the highest recorded since 1994. This was a 17.8% increase in P&S syphilis compared to rates reported in 2015 (7.5/100,000 population). Men accounted for almost 90% of all P&S syphilis cases in 2016. Cases increased among MSM with a doubling of the rate of P&S syphilis, the most infectious stage, between 2000 and 2016. In 2016, MSM accounted for 80.6% of all P&S syphilis. This increase is of public health concern because exposure to genital sores caused by syphilis can increase the risk of acquiring HIV infection. The CDC reports 47% HIV coinfection among MSM diagnosed with P&S syphilis in 2016 (CDC, 2017).

The connection between syphilis and HIV rates can be traced to the beginning of the HIV epidemic within the MSM community in the early 1980s. The public health prevention efforts

focusing on safer sex and community wide behavior changes to prevent HIV transmission during the 1990s resulted in a significant decrease in syphilis infection. The CDC's 1999 National Plan to Eliminate Syphilis was initiated because the syphilis rate had fallen to a historic low by 2000, but the rates began rising again within the MSM community shortly afterward. According to Smith (2014), there are three initial areas to explore when trying to identify the reason for an increase in disease infection rates:

a. Is there a change in the organism? The treatment for syphilis infection at any stage has not changed since penicillin was first used in 1943, and the organism continues to remain susceptible (CDC, 2015; Frith, 2012; Ho & Lukehart, 2011).
b. Did the incubation period or infectiousness of the organism change? Syphilis disease progression has been well documented, and there have been no changes in the incubation period or virulence of the organism (CDC, 2015).
c. Was there a public health or psychosocial change? There have been documented changes in high-risk behaviors, especially within the MSM community. The CDC 2006 SEE report and multiple authors identified the following high-risk behaviors that may contribute to this continued increase in syphilis infection especially among MSM:
    - Increase in unprotected penetrative sex, including oral-genital intercourse. Zetola and Klausner (2007) reported that among the MSM community, oral sex is considered to be "safer" sex because it is rarely associated with HIV infection; however, other STDs including syphilis can be transmitted orally.
    - Serosorting. Choosing to have unprotected sex with partners with the same HIV negative or positive serostatus. This practice potentially reduces the risk of contracting HIV, but other STDs can be transmitted (Rowniak, 2009).

- HIV treatment optimism. The success of highly active antiretroviral therapy (HAART) for HIV treatment resulted in a decrease of the infected person's total HIV burden, maintenance of a functional immune system, and prevention of opportunistic infections that often lead to death (Moore & Chaisson, 1999). New HIV viral load test methods provide healthcare practitioners with quantitative tools for making decisions about starting or modifying HIV treatments. According to Halkitis, Wilton, and Drescher (2005), these HIV treatment and testing advances may have a "widespread effect on the sexual behavior of MSM by decreasing the perceived severity and consequences of HIV transmission" (p.18). Rowniak (2009) provided a review of multiple research studies that documented an association between HIV treatment optimism and increased high risk behaviors of unprotected sex and drug use. Treatment optimism can also be applied to syphilis infection which is curable with antibiotic use; therefore, the infection risk is minimalized and may lead to an "if you get it, it can be treated" attitude among the MSM population (CDC, 2006).
- Safer sex fatigue. Rowniak (2009) describes safer sex fatigue as "the inability of individuals to maintain safe sex practices" (p. 32). Ostrow, et al., (2008) performed a prospective study of attitudinal and relationship predictors of sexual risk among HIV positive and negative MSM. The study showed that a higher proportion of MSM reporting unprotected sexual partners was associated with decreased HIV concern, possibly due to HAART, and increased safer sex fatigue. The increase in safer sex fatigue could be due to repeated exposure to outdated and simplistic HIV prevention messages, failure of public health prevention programs to develop new approaches to prevention that recognize psychosocial changes within the community, or that older

MSM no longer have the constant reminder of friends dying of acquired immunodeficiency disease syndrome (AIDS) as occurred at the beginning of the AIDS epidemic, and younger MSM have not experienced the fear of dying from AIDS because of advances in HIV treatment (Halkitis, Wilton & Drescher, 2005; Ostrow, et al., 2008).

- Substance abuse. Halkitis, Wilton and Drescher (2005) discuss studies that have shown an increased rate of substance abuse among MSM. Methamphetamine, a "club" drug, can be used to lower sexual inhibitions and enhance sexual experiences. Its use has been associated with risky sexual practices such as unprotected anal intercourse.
- Internet. The Internet has provided new opportunities for MSM to meet sex partners anonymously 24 hours a day, 7 days per week. Multiple websites are available that allow men to identify a preference for unprotected sex and often provide for disclosure of HIV status to allow for serosorting. Multiple authors have demonstrated an increasing association between early syphilis infection and Internet use. MSM who meet sex partners online report more sex partners, more likely use of methamphetamines, more likely to have unprotected sex, and more likely to have had an STD in the prior year (CDC, 2003; Halkitis, Wilton & Drescher, 2005; Klausner, 2000).

Public health programs at the local and state level include both clinical and laboratory services necessary to perform syphilis detection, treatment and prevention. Because of the interaction between syphilis and HIV infection, the CDC recommends that STD and HIV programs at local health departments should continue to merge to allow for synergistic

opportunities for disease detection and prevention (CDC, 2006). Loss of funding for STD programs at the state and local level due to the recent US economic downturn has had a significant negative impact on the ability of public health agencies to provide clinical and laboratory services. A study by Willard, Shah, Leep and Ku (2012) found that by 2010, over 50% of local health departments (LHDs) had cuts to their core funding. Between 2008-2009, over 45% of LHDs had lost staff to layoffs or attrition, thus challenging the ability of remaining staff to support programmatic activities. Between 2001 to 2006, the federal government cut funding for infectious disease control programs and HIV programs by 1.9% and 21%, respectively. The loss of funding required public health programs to streamline services by finding alternative providers or forming partnerships, which most likely involved the private sector. The Affordable Care Act (ACA) may allow more patients to access STD treatment through the private sector; however, this will require increased collaborations between the public and private sector to ensure that private sector providers understand their public health responsibilities. The responsibilities related to STD prevention and control include appropriate laboratory testing, correct disease staging, appropriate treatment, public health disease reporting, and participation in effective partner services. Laboratory services include an urgent need to improve diagnostic testing capabilities with new tools for rapid and accurate detection. The 2006 CDC SEE report included a set of recommended laboratory service activities for syphilis elimination including establishing a regional laboratory network for performance of syphilis PCR testing and policy guidance on the use and interpretation of the reverse sequence algorithm (CDC, 2006).

Public health interventions have traditionally been provided through public health STD clinics. Celum, et al. (1997) performed a cross-sectional study of 2,490 patients attending five

urban STD clinics to characterize the clinic clientele and determine why they sought care in a public health setting. The results of the study showed that 51% of the patients were under 25 years of age; 64% were non-white; 43% earned less than $10,000 per year; and 59% were uninsured. The STD morbidity among the patients was high with 66% diagnosed with one or more STD. The main reasons for attending the STD clinic were for convenience (walk in appointments and location), cost, confidentiality, and expert care.

With the implementation of the ACA in 2010 and the 2007 economic recession, many changes have occurred in the healthcare environment. In order to answer questions about current access to public STD clinics, the diseases detected and services provided, Pathela, et al. (2015) performed a retrospective analysis of 608,536 clinic visits to 40 geographically diverse clinics associated with the CDC Sexually transmitted Surveillance Network (SSuN) from 2010 through 2011. The SSuN geographically diverse areas are located along the four US boundaries: East (New York and Virginia), West (San Francisco and Washington), North (Minnesota) and South (Colorado). The SSuN participants followed common protocols for data collection and included enhanced data elements such as anatomic sites of infection, risk behaviors and treatment (Rietmeijor, et al., 2009). Pathela, et al. (2015) found that 61.9% of STD clinic patients were male, 47.1% were 20 to 29 years old and 56.2% were non-Hispanic black. There were 212,765 STD diagnoses made during clinic visits. While the ACA may provide insurance for many US citizens, underserved populations such as undocumented immigrants and underinsured individuals still need access to STD care. STD clinic staff provide a two-phase disease control approach, which includes detection and treatment of syphilis infection and effective partner services such as notification and contact treatment. Private healthcare providers may not understand their public health role in this type of community based disease prevention. STD

clinics will need to evaluate their partnerships to ensure they can provide proactive training and support to these new ACA healthcare paradigm participants. Elimination of syphilis infection in the US continues to be an achievable goal because the disease can be detected with laboratory diagnostic tests and is treatable. The current high levels of syphilis are localized in specific populations that can be targeted with public health interventions. However, federal funding is needed to accomplish the goals at the state and local level if the elimination services remain in the public realm.

Summary

Syphilis is a complicated disease to diagnose because of its ability to mimic other diseases and the overlapping stages of disease progression. Current laboratory diagnostic testing capabilities rely on detection of an immune response to *T. pallidum* infection. While there are multiple different testing methods available, there is variability in the sensitivity and specificity of each assay. This variability is particularly important when reverse algorithm screening is utilized. If a highly sensitive treponemal test is used for initial screening, then CDC recommends that an equal or higher sensitivity test should be used for the second treponemal test in the event of discordant results with the nontreponemal test. If a lower sensitivity test is used for discordant testing, then there is a potential for assumption that the initial test was a false positive, which could have deleterious effects on the patient's health as well as allowing the spread of disease within the community. The CDC does not recommend the use of the FTA-ABS test as a second treponemal test due to its low sensitivity and recommends use of the TP-PA test. The decision to implement a new test methodology, such as an automated syphilis test, is based on a number of considerations including cost, equipment, training, assay sensitivity and

specificity. In addition, the laboratorian must know the prevalence of syphilis within the population because it will impact the predictive value of the test (Zanto, 2010).

The quantitative S/CO or index value generated by automated treponemal assays may be of benefit in resolving discordant results with reverse algorithm testing. The S/CO value is proportional to the amount of *T. pallidum* antibody present in the infected person's serum. In instances where there are discordant results between the initial treponemal screening and the nontreponemal test, a high S/CO value could be predictive of a positive second treponemal test. If a standard cut off value for a "true positive" treponemal test can be identified for the automated test, then the second tests would not have to be performed which would result in a cost savings for the laboratory. Additionally, nonreactive samples with S/CO values within a specific range could be flagged by the laboratory with a comment so that the healthcare provider would be alerted to perform additional follow up testing (Wang, et al., 2011). The CDC suggested in the 2015 Syphilis treatment guidelines that the usefulness of these semi-quantitative values should be investigated further (CDC, 2015).

Public health STD clinics provide a unique environment for detecting syphilis infection. The clinics are predominately accessed by individuals who are underserved in the traditional healthcare system. STD clinic populations typically include racial/ethnic minorities, young persons, MSM and WSW. These populations often have high STD rates. STD clinics are particularly important for men, who are less likely than women to seek preventive care. Patients prefer to access STD clinics rather than primary care due to confidentiality, expertise and convenience. These clinics serve a public health function of identifying index patients and preventing further spread of disease through contact investigation. Useful laboratory testing is

an important diagnostic tool for determining individual syphilis infection and preventing community-wide disease spread.

There are gaps in the current body of knowledge regarding the usefulness of the Captia™ Syphilis IgG EIA assay for syphilis detection. The Captia™ Syphilis IgG method was one of the first EIA methods approved by the FDA in 2001. While there are research articles from that time period regarding the sensitivity and specificity of the test, there is little or no research regarding the usefulness of the Captia™ Syphilis IgG test within the reverse algorithm constellation of tests. Additionally, review of current literature did not reveal studies that explored the predictive value of the Captia™ Syphilis IgG S/CO value for determining syphilis infection. According to the College of American Pathologists (CAP) 2016 Syphilis serology G-B proficiency test survey summary, there were 104 US participating laboratories that reported Syphilis IgG EIA testing with 48% (50/104) of the laboratories using the Captia™ Syphilis IgG assay (CAP, 2016). Additional research on test usefulness would be helpful as laboratories make decisions on maintaining or updating test methodology. Most of the research that has been conducted on the usefulness of the reverse and traditional algorithm has been performed either on stored serum samples or in the clinical setting. There is minimal research comparing the reverse and traditional algorithms in a public health setting with patients who may be at risk for syphilis infection. Studies of the cost effectiveness of the reverse and traditional algorithms often do not include definitions of the methods used to determine the cost of laboratory testing. Differences between the direct and fully loaded cost of syphilis test methods can have an impact on determining the cost effectiveness of either algorithm in detecting syphilis infection.

Chapter 3: Methodology

This chapter will present the methodology for determining the usefulness of the Captia™ Syphilis IgG EIA test method and reverse algorithm for detection of syphilis infection in a public health setting. The theoretical framework for the dissertation study will be presented along with research design details including subjects, variables, human studies protocol, data collection, data analysis, validity, and reliability.

Theoretical Framework

This study utilized the Donabedian Quality Framework to evaluate the usefulness of the Captia™ Syphilis IgG EIA test and the syphilis reverse algorithm for detection of syphilis infection. The Donabedian model is composed of three domains: structure, process, and outcomes. Each domain is considered equally important and linked hierarchically. Structure is the foundation and primarily influences process; however, both structure and process influence the outcomes. Since 1980, the model has been successfully used as a framework for quality assessment and systems monitoring of healthcare systems (Donabedian, 1980; Polit & Beck, 2012; Shi & Singh, 2012).

The Donabedian model is very similar to the Quality Management System (QMS) Model for Laboratory Services, an approved guideline developed by the Clinical and Laboratory Standards Institute (CLSI). The QMS model describes the sequential path of workflow within a clinical laboratory showing the steps from receipt of a laboratory test order to a reportable test result (CLSI, 2004). The path of laboratory workflow is further defined in federal regulation 42 Code

of Federal Regulations (CFR) 493 Centers for Medicare and Medicaid Services (CMS) Clinical Laboratory Improvement Amendments (CLIA) Programs: Laboratory Requirements Relating to Quality Systems and Certain Personnel Qualifications, Final Rule published on January 24, 2003. All US laboratories that perform diagnostic testing on humans must follow the CFR 493 CLIA regulations, and perform quality assessment through all phases of laboratory testing within the path of workflow (CMS, 2017). These regulations also utilize the QMS pathway with slightly different terminology. The QMS pathway describes the path of workflow as preexamination, examination, and postexamination; whereas, CLIA regulation describes the path of workflow as preanalytic, analytic, and postanalytic. While the terms are slightly different, the definitions are the same. For this dissertation, the CLIA terms will be used. In reviewing the Donabedian model and the QMS/CLIA model, a similarity between the three domains and the laboratory workflow path can be identified (Figure 8).

Structure is defined as resource outputs in the Donabedian model. QMS/CLIA describe the preanalytic component as all the activities from the time a test is ordered until it is collected and transported to the laboratory for processing (CLSI, 2004; CMS, 2017). These activities include the use of a variety of qualified, trained staff to collect, process and transport the sample; delivery systems such as courier or commercial shippers; facilities for sample collection procedures; and equipment to collect the sample. These preanalytic activities (staff qualifications, staff levels, delivery, facilities, and equipment) are included in the Donabedian model for the structure domain.

Process is defined as actual delivery of healthcare in the Donabedian model. QMS/CLIA describe the analytic component as follows: selecting the appropriate test method including a

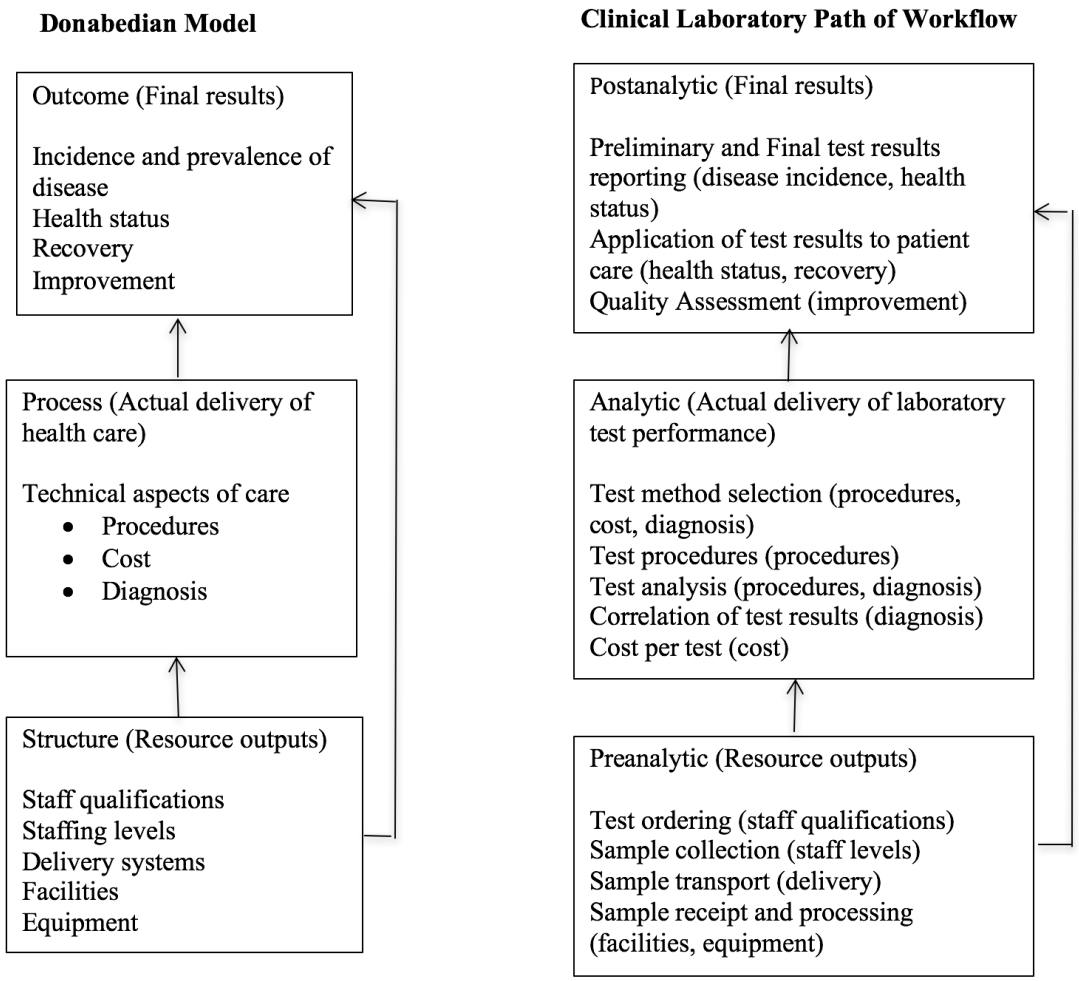

Figure 8. Comparison of Donabedian Quality Framework Model and Clinical Laboratory Path of Workflow.
Adapted from *Delivering Health Care in America* by L. Shi & D.A. Singh, 2012, p. 506, Copyright 2012 by Jones & Bartlett Learning, LLC and *Application of Quality Management System Model for Laboratory Services: Approved Guideline – Third Edition* by CLSI, 2004, pg 7-12 Copyright 2004 by The National Committee for Clinical Laboratory Standards.

cost per test evaluation, verifying test performance, creating standard operating procedures (SOPs), performing the test using trained and competent staff, and correlating final test results (CLSI, 2004; CMS, 2017). These analytic activities (cost, procedures, and diagnosis) are included in the Donabedian model for the process domain.

Outcome is defined as the final results in the Donabedian model. QMS/CLIA describe the postanalytic component as the activities of test result reporting, which may be the healthcare

provider and/or a public health agency (CLSI, 2004; CMS, 2017). The healthcare provider will utilize the results to treat the patient while the public health agency will utilize the results to track incidence and prevalence of disease within the community. Other postanalytic activities include consultation to healthcare providers regarding the test interpretation or any follow up testing that may be recommended. Abnormal or critical results are communicated to the healthcare provider so that appropriate treatment can be started as soon as possible. These activities could have an impact on the health status of the patient and affect their recovery. Postanalytic activities may also impact public health investigations if further testing is required to characterize a disease outbreak or emerging pathogen. Quality assessment (QA) is a CLIA requirement for all clinical laboratories and includes specific standards that must be included in a laboratory quality assessment program. Adherence to these standards ensures continual laboratory quality improvement (CMS, 2017). These postanalytic activities (disease incidence and prevalence, health status, disease recovery, and improvement) are included in the Donabedian model for the outcome domain.

The Donabedian model domains are hierarchical with structure influencing process and both structure and process domains affecting outcomes. This study utilized the process (analytic) and outcome (postanalytic) domains of the Donabedian framework to evaluate the usefulness of the Captia™ Syphilis IgG EIA test and the syphilis reverse algorithm interpretation to detect syphilis infection in a public health setting. For this dissertation, there was an assumption that the structure domain was performed properly with trained and competent staff following established SOPs for sample collection and handling. The western region laboratory where the dissertation testing was performed provided SOPs and training to all staff involved with the preanalytic

process. The laboratory maintained documentation for samples that were rejected due to improper handling or processing.

The target construct for the study was syphilis infection, and the proposition for the research question was that more useful syphilis testing will improve the detection of syphilis infection. The following laboratory analytic and postanalytic domain activities were utilized for the study:

- Analytic (process) domain activities: syphilis test results and S/CO value (predictive value).
- Postanalytic (outcome) domain activities: syphilis reverse and traditional algorithm serologic laboratory interpretation; syphilis diagnosis (predictive value).

Research Question and Hypotheses

The specific research question and hypotheses addressed by this study are:

RQ_1: What is the usefulness of the Captia™ Syphilis IgG EIA test method and the reverse algorithm for detection of syphilis infection in a public health population?

- $H1_o$: Among patients seeking STD services at a large western region metropolitan public health clinic, there will be no difference between the Captia™ Syphilis IgG EIA and Fujirebio Serodia TP-PA test results.
- $H2_o$: Among patients seeking STD services at a large western region metropolitan public health clinic, there will be no difference in diagnostic interpretation of the Captia™ Syphilis IgG EIA S/CO value and the Becton Dickinson MacroVue RPR titer test result.
- $H3_o$: Among patients seeking STD services at a large western region metropolitan public health clinic, there will be no difference between the syphilis traditional and reverse algorithm interpretations.

Research Design

The study was a retrospective, nonexperimental descriptive correlational design. It addressed whether useful diagnostic information about syphilis infection can be obtained utilizing the Captia™ Syphilis IgG EIA test and the syphilis reverse algorithm. According to Polit and Beck (2012) the aim of correlational studies is to describe relationships rather than to determine causality. The design was nonexperimental because there was no manipulation of an independent variable. It would be unethical to manipulate a syphilis test result, as it would have a deleterious effect on determining if the patient was infected with syphilis. All data used for the study was collected during two calendar years January 1, 2012 to December 31, 2013 as purposive convenience sampling; therefore, the study design was retrospective. The study data was collected as part of standard public health practice within the STD clinic at a large western region metropolitan (greater than two million population) public health agency. According to CDC STD surveillance criteria, the western region includes the following states: Alaska, Arizona, California, Colorado, Hawaii, Idaho, Montana, Nevada, New Mexico, Oregon, Utah, Washington, and Wyoming (CDC, 2015). During 2012-2013, the western region public health agency enhanced their standard syphilis surveillance protocols to include the performance of three different syphilis tests (Captia™ Syphilis IgG, MacroVue RPR, and Serodia TP-PA) on patients who exhibited high-risk behavior, who were contacts to syphilis cases, who had symptoms of syphilis infection, who had an initial reactive or equivocal treponemal test result, or who had a nonreactive Syphilis IgG S/CO value greater than or equal to 0.450.

The public health laboratory associated with the public health agency began performing initial syphilis testing using the syphilis reverse algorithm in 2009; prior to this time, the traditional algorithm was used. The syphilis reverse algorithm starts with an initial treponemal

test (EIA) and reflexes to nontreponemal tests (RPR and titer) if the results are reactive or equivocal. If the nontreponemal tests are nonreactive, then a second but different treponemal test (TP-PA) is performed as the tiebreaker. The CDC recommended performance of the TP-PA test only if the reverse algorithm reflex RPR test was nonreactive. However, there was concern among the western region public health agency STD clinic staff that the reverse algorithm was not as sensitive as the syphilis traditional algorithm; and the staff was confused by the reverse algorithm interpretation when the results were discordant. Therefore, during the study period, the public health laboratory coordinated with the public health STD clinic to modify the reverse algorithm recommended by CDC so that an initial reactive or equivocal treponemal test was reflexed to RPR, RPR titer, and TP-PA tests, which were all run simultaneously. This modification was similar to the modified reverse algorithm described in the study conducted by Berry and Loeffelholz (2016). The ability to order a three-test panel consisting of Syphilis IgG, RPR, and TP-PA based on their clinical assessment of the patient provided STD clinic staff with the option to clinically correlate the test results using either the traditional or reverse syphilis algorithms. Additionally, in 2010, the public health laboratory and STD clinic staff had detected numerous instances of nonreactive Syphilis IgG EIA tests in patients who presented with symptoms of primary syphilis. A pilot study was performed by the laboratory and it was determined that nonreactive samples with an S/CO value greater than or equal to 0.450 were more likely to be falsely negative when correlated with clinical symptoms (P.Armour, personal communication, January, 10, 2011). Based on this pilot study and discussion with the STD clinic staff, the public health laboratory director approved performance of additional reflex testing to RPR and TP-PA for those nonreactive Syphilis IgG samples with a S/CO value greater than or equal to 0.450. This additional reflex testing was in use during the study time period.

All syphilis test results were reported for each patient and were used by the public health agency STD clinic staff to determine the presence or absence of syphilis infection based on clinical correlation of test results. Obuchowski (1998) states that one strategy to improve the power of diagnostic test accuracy studies is to use a paired design with all tests performed on the same sample. Additionally, the power of a retrospective study comparing diagnostic tests can be improved by including patients with "subtle" as well as normal manifestations of the disease. In the case of syphilis diagnosis, infected patients were clinically staged within the disease progression based on laboratory test results, symptoms and risk factors. This study included patients with no syphilis disease, those in the early (primary) stages of syphilis disease as well as those in later (latent) stages, which enhanced the study design.

Design Validity

Polit and Beck (2012) describe four types of research design validity: statistical conclusion, internal, construct, and external. Each type will be discussed in relation to the study design.

Statistical conclusion validity. Statistical conclusion validity of the design is the ability of the study to detect relationships that "exist in reality and can be reliably detected" (Polit and Beck, 2012, p. 241). Threats to statistical conclusion validity can be reduced by ensuring the study has adequate statistical power and maximizing precision. Statistical power will be discussed in the population and sample section that follows. Hulley et al. (2013) discuss the following five strategies for increasing precision:

- Standardize the measurement methods in an operations manual. The public health laboratory performing the testing on the study samples was registered with the Centers for Medicare and Medicaid Services (CMS) CLIA program and was required by federal regulation to maintain SOPs. Additionally, CLIA regulations state that all testing must

be performed following the test manufacturer instructions (CMS, 2017). The FDA is the regulatory authority for all clinical diagnostic tests sold in the US and approves all test manufacturer instructions to ensure that the test provides accurate results when performed as instructed.

For the chart review clinical diagnosis and staging measurement, the STD clinic staff determined all syphilis disease staging based on standard CDC guidelines, which were documented in the STD clinic SOP. For the reverse and standard algorithm interpretation, the western region laboratory staff and/or the student researcher determined the interpretation based on the CDC algorithms (Figures 4 and 5).

- Training and certifying the observer. CLIA regulations require that the laboratory director must provide written certification that all testing staff are trained and competent prior to performing any diagnostic test (CMS, 2017). The laboratory director certified all public health laboratory staff performing the testing for this study as trained and competent to perform the tests and enter test results into the Laboratory Information Management System (LIMS).

  For the chart review clinical diagnosis and syphilis staging, the public health agency required that all STD clinic staff performing clinical assessments must be trained and competent to determine syphilis disease stages. The staff was further trained to enter data correctly into the patient's electronic medical record. The public health laboratory staff and student researcher performing the information abstraction from the electronic medical record received training from the clinic and information technology (IT) staff prior to performing the abstraction.

- Refining the instrument. The FDA approved all three diagnostic tests (Captia™ Syphilis IgG, MacroVue RPR, and Serodia TP-PA) used in this study for testing human serum to provide serologic evidence of syphilis infection. The staff at the public health laboratory performing the tests in this dissertation study followed the FDA approved manufacturer's instructions with no modifications. An additional external control was added to the Captia™ Syphilis IgG test; however, the additional control was not considered to be a modification by CLIA. FDA approved manufacturer's directions for test performance of the three syphilis tests used in this dissertation study are detailed below.
    - Captia™ Syphilis IgG is an EIA test for the qualitative detection of *T. pallidum* IgG antibodies in serum specimens. It is intended to be used in conjunction with nontreponemal testing to provide serological evidence of *T. pallidum* infection. According to the manufacturer's package insert (Trinity Biotech, 2003), the assay principle is as follows:

      > Microtitration wells with *T. pallidum* antigens are exposed to test specimens which may contain specific IgG antibodies. After an incubation period, unbound components in the test sample are washed away. Specifically-bound IgG reacts with a conjugated horseradish peroxidase (HRP) monoclonal antibody (mAb) during the second incubation period. Following a second wash cycle, specifically-bound enzyme conjugate is detected by reaction with tetramethylbenzidine (TMB). The enzymatic reaction is stopped using 1N sulfuric acid. The assay is measured spectrophotometrically to indicate the presence or absence of IgG treponemal antibodies (p. 1).

      Immunoassay methods may exhibit a possible interference phenomenon known as the hook effect. The effect occurs when there is excess antibody blocking all the binding sites on the antigen. This may result in a falsely low test result which could be reported as a false negative. The hook effect typically occurs in single step immunoassays where antigen, antibody, and marker are all incubated at the

same time. According to Dodig (2009), Miller (2004), and Selby (1999) the hook effect does not typically occur in two-step competitive assays, such as the Captia™ Syphilis IgG test method, that include a wash step prior to the addition of the second antibody, which is HRP-mAb conjugate for the Captia method. An additional mechanism for eliminating the hook effect is to dilute the sample at a previously established dilution prior to analysis. According to the manufacturer's package insert, all serum samples were diluted 1:21 automatically by the EVOLIS instrument prior to analysis. This sample dilution step added to the assay design also reduces the possibility of a hook effect.

The assay can be performed manually or on a qualified automated analyzer using a 96 well format. The western regional laboratory utilized the BioRad EVOLIS automated analyzer for all Captia™ Syphilis IgG testing performed on the study samples. Each assay kit contains microwell strips, conjugate, substrate, diluent, and controls (high titre reactive [HTR], low titre reactive [LTR] and nonreactive [N]) to perform either 96 or 960 tests depending on package configuration. The kit materials may not be interchanged with kits with different lot numbers because the kits are optimized and balanced for a specific kit lot number. All test specimens were diluted 1:21 with diluent prior to analysis. The kit controls were provided at working strength and were not diluted. All three controls (HTR, LTR and N) were included on each 96 well plate and placed in specific locations on the test plate so that the analyzer calculated plate run validity. The LTR control was tested in duplicate and was used to calculate the cutoff value for distinguishing nonreactive, equivocal and reactive test results.

Because the LTR was used as a single point calibrator on each 96 well plate, it was no longer considered a control according to CLIA guidelines; therefore, an external well-characterized low-titre reference control was included in each run to act as a replacement low titer control. The external low-titre reference control (Virotrol, Syphilis total) used by the public health laboratory was purchased from Bio-Rad. It was treated the same as the sample and was, therefore, diluted prior to analysis. At the western region laboratory where the dissertation testing was performed, three levels of control (N, external reference low-titre, and HTR) were performed each day of use on each plate. Additionally, according to the manufacturer's instructions, the LTR was included in duplicate on the 96 well plate as a calibrator. Each 96 well plate analysis was validated as follows:

- The absorbance of N must be less than or equal to 0.25
- The absorbance value of HTR must be greater than or equal to 0.8
- The mean absorbance of LTR must be greater than or equal to 2.5 x the absorbance of N.

All three parameters must be met for an assay plate to be valid. If any parameter was not met, the entire assay plate was invalid and was repeated. Additionally, the external low-titre reference control must provide reactive results or the entire assay plate was invalid and was repeated. No patient test results were reported for invalid test runs. Calculation of patient results were performed as follows:

- Calculate the mean absorbance value of the duplicate LTR. According to manufacturer's package insert (Trinity Biotech, 2003), "this is the cut-off value for the test method and was derived from clinical trials as the value

giving optimum discrimination between specimens which are reactive or nonreactive for antibodies to *T. pallidum* as characterized by a range of standard serological techniques" (p. 3).

- Calculate the Antibody Index (AI) or Signal to Cutoff (S/CO) ratio for each sample by dividing the absorbance of the test specimen by the mean absorbance of the kit LTR.
    - For example: Test serum absorbance = 0.75 and mean LTR absorbance = 0.30. The S/CO value = 0.75/0.30 = 2.50
- Following calculation of the S/CO value for each test sample, the final result was interpreted as follows:
    - S/CO value less than or equal to 0.9 is considered nonreactive
    - S/CO value between 0.9 and 1.1 is considered equivocal
    - S/CO value greater than or equal to 1.1 is considered reactive
        - In the example above, an S/CO value of 2.50 would be reported as reactive.
- Initially equivocal or reactive results were repeated in duplicate. The BioRad EVOLIS included a computerized program that automatically ordered repeat Syphilis IgG testing and interpreted the final test result based on all three Syphilis IgG test results as shown in the algorithm in Figure 9. An invalid test result was reported if the duplicate repeat test results were not reproducible. Invalid test results may be due to instrument, reagent, and/or operator error. An invalid test result cannot be interpreted as either reactive or nonreactive and requires additional testing

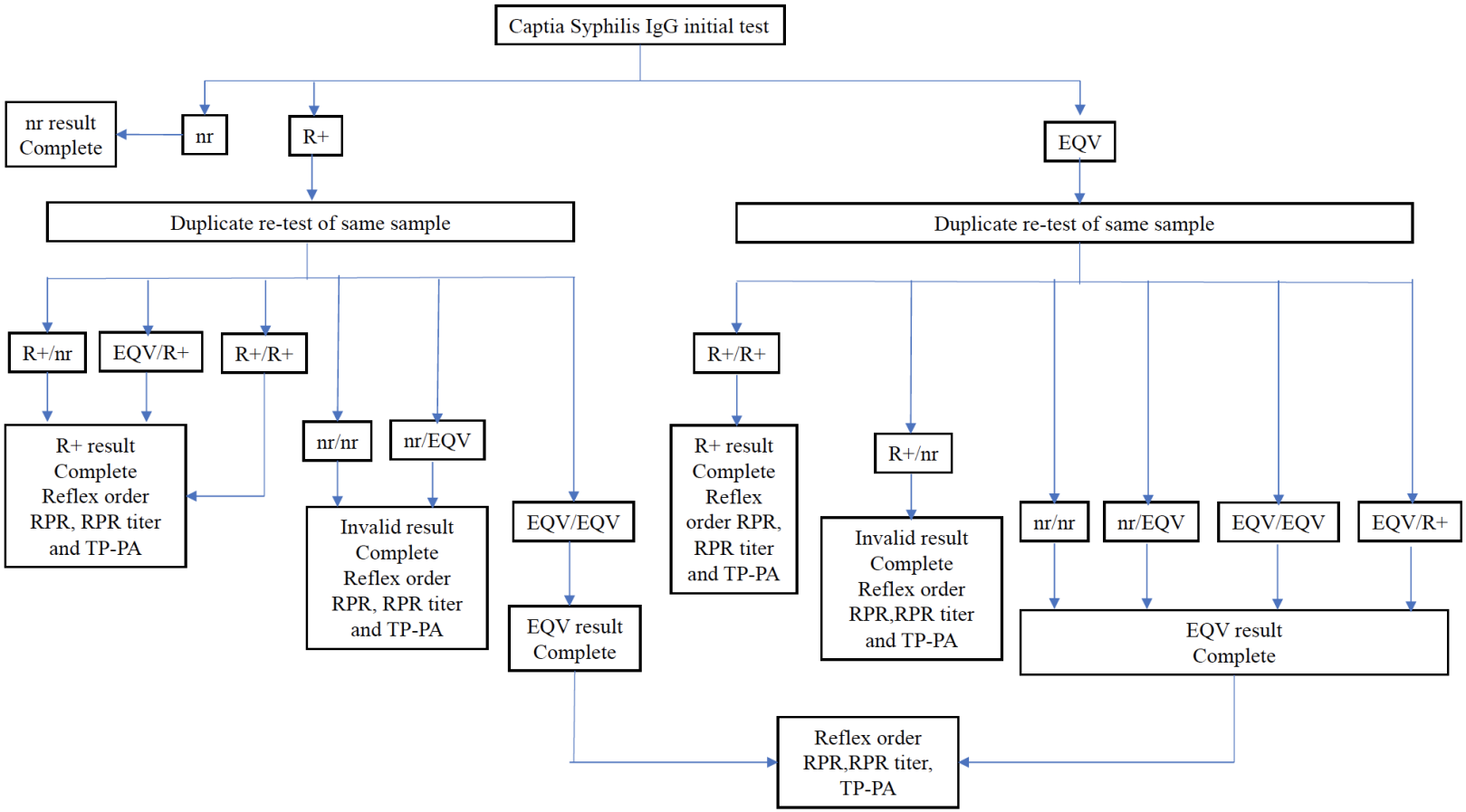

Figure 9. Western Region Public Health Laboratory Captia™ Syphilis IgG Repeat Algorithm.
nr = nonreactive; EQV = equivocal; R+ = reactive**.** P. A. Armour, Captia™ Syphilis repeat algorithm, personal communication, 11-18-2008.

or recollection to determine syphilis infection. The western region laboratory protocol for invalid test results required reflex to RPR, RPR titer, and TP-PA testing. Invalid test results were excluded from the study as they cannot be used for clinical diagnosis. Additional discussion of invalid test results is in the measurement reliability and validity section. According to the manufacturer's package insert, only the qualitative (nonreactive, equivocal, reactive, or invalid) results were reported. The S/CO value was only for internal laboratory use for analysis calculation and results interpretation. The value was not reported to the healthcare provider.

- Becton Dickinson MacroVue RPR 18mm card test is a manual nontreponemal macroscopic flocculation test for the detection of syphilis infection. The assay principle according to the manufacturer's package insert (Becton Dickinson, 2010) is:

> The RPR card antigen suspension is a carbon particle cardiolipin antigen which detects "reagin", an antibody-like substance present in serum or plasma from syphilitic persons, and occasionally in serum or plasma of persons with other acute or chronic conditions. The reagin binds to the test antigen, which consists of cardiolipin-lecithin-coated cholesterol particles, causing macroscopic flocculation (p. 1).

The test kit included antigen, test cards, dispensing bottle, and silicone coated dispensing needle. The accuracy of the dispensing needle was verified prior to use by checking that the needle delivered 30+/- 1 drops of 0.5 ml of test antigen. The needle was discarded if accuracy was not verified. External QC material was purchased separately. According to CLIA regulations, each day of use, at least a negative and positive control must be performed. At the western region laboratory where the dissertation testing was performed, three levels of RPR qualitative control

(nonreactive, minimally reactive, and reactive) were performed each day of use and covered the full range of qualitative reactions that could be expected for an individual patient sample. All three control results must meet expected values in order for testing to proceed. Additional equipment required for testing included a mechanical rotator which circumscribed a circle approximately two centimeters in diameter in the horizontal plane and included a humidifying cover. The recommended speed for the rotator was 100+/- 2 rpm and the speed was verified prior to analysis. An example of the RPR QC documentation form is shown in Figure 10. The form included day of use documentation required for instrument checks, temperature checks, RPR qualitative, and RPR titer quantitative controls. All QC checks were reviewed by a second technologist and passed all expected parameters before patient testing was performed. Final RPR qualitative results were reported as either nonreactive (no flocculation or clumping) or reactive (any degree of flocculation or clumping). The manufacturer's package insert recommended that all reactive RPR tests should be repeated using an alternate procedure. The CDC traditional algorithm recommended that all reactive RPR tests should be reflexed to a quantitative RPR titer and a treponemal test. The treponemal test was for confirmation of syphilis infection and the RPR titer was used to establish a baseline for treatment monitoring. The RPR titer was performed in the same manner as the qualitative test, using the same reagents; however, a two-fold dilution of the sample was performed and the final result was the highest dilution giving a reactive result.

A prozone effect may occur in some patients infected with syphilis. In a prozone reaction, the undiluted sample exhibits a complete or partial inhibition of reactivity

Public Health Laboratory
RPR Day of Use/QC Worksheet

Titer Result Legend:
N = nonreactive
R = reactive
Rm = reactive minimally

Quantitative QC	Expected range: Titer must be 1:4 or 1:8.			
Virotrol Control Panel Member03 Lot #:_____ Exp. Date:_____				
1 Result: ____Titer: 1:1	**2 Result: ____Titer: 1:2**	**3 Result: ____Titer: 1:4**	**4 Result: ____Titer: 1:8**	**5 Result: ____Titer: 1:16**
Circles 6-10 not used				
This form MUST be used each day of use. If any reagent change is made a new form is completed and all of the data below MUST be filled in.				

Room temp _____°C (23 – 29 °C) **Note with each run**
Rotator speed _____ rpm (100 ± 2 rpm)
Needle: _____ drops/0.5 mL (30 ± 1 drop / 0.5 mL)
RPR Antigen Lot # _____; Exp. Date _____
0.9% NaCl Lot # _____; Exp. Date _____

NOTES:

Qualitative RPR--N=nonreactive; R= reactive

Control Card
Lot#: _____ Exp. Date: _____

Reactive	Nonreactive	Reactive (min.-mod.)
Expected result=**R**	Expected result=**N**	Expected result=**R**
Result: ____	**Result: ____**	**Result: ____**
Note: To be run minimally each day of use.		

Page #_____
Performed by: _____
Date Performed: _____
Reviewed by: _____
Date: _____

Figure 10. Western Region Public Health Laboratory RPR QC Documentation Form. P. A. Armour, RPR Day of Use/QCworksheet, personal communication, 7-15-2011

(nonreactive result); however, if the sample was diluted (RPR titer) there was maximum reactivity shown in the sample, thus high levels of reagin antibody may produce a nonreactive qualitative RPR test result. This is similar to the hook effect described above for the immunoassay test method. All samples that exhibited a "rough" visual appearance with the qualitative RPR test method were diluted using the RPR titer method. Additionally, if the healthcare provider suspected syphilis and the qualitative RPR was nonreactive, an RPR titer should be ordered to rule out the prozone effect (Larsen, et al, 1998). As recommended by the CDC traditional and reverse algorithms, all reactive qualitative RPR tests had a quantitative RPR titer performed.

- The Fujirebio Serodia TP-PA is a "qualitative gelatin particle agglutination assay intended to be used for the detection of *T. pallidum* antibodies in human serum or plasma as an aid to the diagnosis of syphilis" (Fujirebio Diagnostics, Inc., 2006, p.1). The test is designed for a manual microtitration 96 well plate method that is read visually. The test kit included reactive and nonreactive control serum, sample diluent, sensitized particles, unsensitized particles, and reconstituting solution. The 96 well microplate was purchased separately. In accordance with CLIA regulations, positive and negative controls were performed each day of use. Due to the two-hour incubation time required for the test, the controls were run at the same time as the patient samples. Both control values must meet expected values in order to report the final results. Samples and controls were serially diluted in microplate wells. Gelatin particles sensitized to *T. pallidum* (Nichols Strain) antigen and unsensitized gelatin particles were added to the respective wells; the plate was mixed and incubated for

two hours at room temperature. After the two-hour incubation, the microtiter plate was read macroscopically using a magnified, lighted viewbox. Samples containing *T. pallidum* antibodies (IgG and/or IgM) will react with the antigen-sensitized gelatin particles and form a smooth mat in the tray well indicating a positive reaction. Nonagglutinated particles will form a compact button indicating a negative reaction. According to manufacturer package insert, final results were read macroscopically and interpreted as follows:

- Reactive: specimen showing negative with unsensitized particles but demonstrating a positive reaction with any dilution at 1:80 or over with sensitized particles.
- Nonreactive: specimen showing negative with sensitized particles at a 1:80 final dilution regardless of a reaction with the unsensitized particles.
- Inconclusive: specimen showing negative with unsensitized particles but with a plus/minus reaction (button with a hole in the center) with sensitized particles at 1:80 dilution.

According to the manufacturer's package insert, an inconclusive result should be repeated using the same sample. If the repeat test was reactive or nonreactive, the repeat results were reported. If the repeat test was again inconclusive, it was reported as inconclusive with a recommendation for further follow up by testing with another treponemal method, such as FTA-ABS or recollection of another sample for analysis in two weeks. An inconclusive test result cannot be interpreted as either reactive or nonreactive and is not clinically diagnostic; therefore, inconclusive test results were excluded from the study and are discussed further in the measurement and reliability section. Figure 11 shows the

Public Health Laboratory
TP-PA Worksheet

Legend:
U – Unsensitized
S – Sensitized
R – Reactive
NR – Nonreactive
Ctl – Control

Room temp _____°C (15 – 30 °C)
Incubation start time: _____
Incubation end time: _____
TP-PA Kit Lot # _____; Exp. Date_____.

		Acc #	U	S	Interp.		Acc #	U	S	Interp.		Acc #	U	S	Interp.
Well	1	2	3	4		5	6	7	8		9	10	11	12	
A	+	R Ctl				Unk 7					Unk15				
B	-	NR Ctl				Unk 8					Unk16				
C	Unk 1					Unk 9					Unk17				
D	Unk 2					Unk10					Unk18				
E	Unk 3					Unk11					Unk19				
F	Unk 4					Unk12					Unk20				
G	Unk 5					Unk13					Unk21				
H	Unk 6					Unk14					Unk22				

Settling Patterns of Particles	Reading	Interpretation
Particles are concentrated in the shape of a button at the center of the well with a smooth round outer margin.	(-)	Non-Reactive (NR)
Particles are concentrated in the shape of a compact-ring with a very small "hole" in the center and a smooth round outer margin	(-)	Non-Reactive (NR)
Particles are concentrated in the shape of a compact ring with a "hole" in the center and a smooth round outer margin.	(+/-)	Inconclusive (I)
Defined large ring with a rough multiform outer margin and peripheral agglutination.	(+)	Reactive (R) NOTE: If a serum sample demonstrates a positive reaction with BOTH the Sensitized and Unsensitized Particles, then retest using the Absorption Procedure. Include the unabsorbed sample in the repeat run.
Agglutinated particles spread out covering the bottom of the Well uniformly, edges sometimes folded.	(++)	

REPORTING:

REACTIVE: A specimen showing Non-reactive with Unsensitized particles but demonstrating + or ++ with Sensitized particles is interpreted as Reactive

INCONCLUSIVE: A specimen showing Non-reactive with Unsensitized particles but demonstrating a +/- reaction with Sensitized particles is regarded as inconclusive. The test must be repeated. If the repeat result is the same, then result test as Inconclusive.

NON-REACTIVE: Regardless of the reaction pattern with Unsensitized particles, a specimen showing Non-reactive with Sensitized particles is regarded as Non-reactive

Performed by: _____
Date: _____

RUN VALIDATION: all controls must perform as listed below for valid run
- Both Controls are Nonreactive to Unsensitized Particles
- Positive Control is REACTIVE to Sensitized Particles
- Non Reactive Control is NONREACTIVE to Sensitized Particles

Valid Run? Yes No
If No, entire run must be repeated
Reviewed by:_____
Date _____

Figure 11. Western Region Public Health Laboratory TP-PA Day of Use Documentation Form. P. A. Armour, TP-PA worksheet, personal communication, 7-8-2011.

TP-PA day of use documentation form utilized by the western region laboratory that performed the dissertation study testing. All QC and test results must be valid and were confirmed by a second technologist prior to reporting test results.

For the chart review clinical diagnosis and syphilis staging, the STD clinic staff followed the agency SOP which was reviewed annually and revised, as needed, based on CDC guidelines for syphilis diagnosis and treatment. The public health laboratory staff used a standardized data collection form with clear criteria for handling ambiguously recorded data. Inter-rater reliability for chart review is discussed in the measurement reliability and validity section.

- Automate the instrument. The EVOLIS automated analyzer used for performing the Syphilis IgG EIA testing was maintained by the manufacturer, BioRad; and preventive maintenance (PM) was performed every six months by BioRad technicians to ensure high quality performance. Additional daily, weekly, and monthly PM was performed by trained laboratory staff. The analyzer also performed automated checks during the entire daily testing process which ensured all operations were completed within the manufacturer's specifications. Any assay plates that did not meet the automated checks during the analysis period for the incubator, pipettor, reagent and sample racks, plate transport, reader, and washer were reported as failed runs and sample results were not calculated. The analyzer detected clotted and short sample tubes and did not calculate results for those samples. If the test plate did not meet the validation criteria for the N, LTR and HTR controls, the entire plate was considered a failed run and individual patient results were not calculated. For failed runs, the instrument did not report the absorbance values for the samples and therefore, manual calculation of test results could not be

performed. The manual TP-PA and RPR tests required minimal automation. The automated equipment such as rotators and mixers used for the TP-PA and RPR tests were all maintained according to manufacturer's specifications. The RPR dispense needle volume and the rotator speed were verified each day of use. External controls were performed with each analysis and had to meet expected values or the patient results were not reported.

- Repeating the measurement. According to the manufacturer's instructions, the Captia™ Syphilis IgG EIA test was repeated in duplicate if the initial test result was reactive or equivocal. This ensured measurement reproducibility for the automated instrumentation. The initially reactive RPR test result was reflexed to undiluted and two-fold serial dilution RPR titer testing until an endpoint was reached. This repeated testing provides measurement reproducibility for the manual RPR test method. An inconclusive TP-PA was repeated in duplicate, and the results were reported based on the repeat test interpretation.

Internal validity. Internal validity of the design refers to any extraneous factors that may vary the study results. Threats to internal validity included selection bias, temporal ambiguity, testing, and instrumentation. Selection bias for the study was controlled because the study population was homogeneous and included all persons seeking care at a public health clinic for STD screening due to either symptoms, high risk behavior, contact to an infected person, or other healthcare reason. The STD clinic staff followed standard protocols for ordering laboratory tests and determining syphilis infection and disease stage. The STD clinic staff followed the laboratory SOP for sample collection and handling. Samples were transported by a trained laboratory courier following the laboratory SOP. The laboratory staff performing the testing

were blinded to the results of the other syphilis tests as well as the initial clinical diagnosis or reason for screening, which reduced outcome bias. The laboratory staff performing the chart review abstraction used a standardized format to obtain the syphilis diagnosis and staging information.

Temporal ambiguity and testing validity was controlled because all three syphilis tests in the study were performed simultaneously on each patient based on initial clinical assessment, reason for the clinic visit or reflex testing following laboratory protocol; therefore, each patient acted as their own matching control.

Instrumentation was controlled by the use of the same automated analyzer throughout the entire study period. The analyzer was not modified or upgraded during the study period. There were no updates or revisions by the manufacturer to the syphilis diagnostic tests (Captia™ Syphilis IgG, Becton Dickinson MacroVue RPR, or Serodia Fujirebio TP-PA) used during the entire study period. Additional discussion on instrument validity is provided in the measurement reliability and validity section.

For the chart review clinical diagnosis and staging, the STD clinic staff followed the 2012 national CDC syphilis case definition for determining syphilis disease stages, and no changes were made to the protocol during the study time period (Table 14). The laboratory staff performing the data abstraction from the electronic chart only entered the syphilis staging categories as shown in the chart; there were no independent judgments made by the laboratory staff regarding the syphilis stage entered in the electronic chart.

Construct validity. Construct validity of design examines the relationship between the study theory and observations. According to Cronbach and Meehl (1955), this relationship is a network that involves how the study proposition relates to the construct and how the construct

Table 14

CDC Syphilis Case Definition by Stage (2012)

Syphilis Stage	**Clinical description**	**Case classification**
Primary	A stage of infection caused by *T. pallidum* characterized by one or more chancres (ulcers)	• Clinically compatible case with one or more ulcers (chancres) consistent with primary syphilis and at least one reactive serologic test for syphilis or • Clinically compatible case with one or more ulcers (chancres) consistent with primary syphilis and *T. pallidum* demonstrated by darkfield, DFA or equivalent method
Secondary	A stage of infection caused by *T. pallidum* and characterized by localized or diffuse mucocutaneous lesions, often with generalized lymphadenopathy. The primary chancre may still be present.	• Currently has symptoms clinically consistent with secondary syphilis and *T. pallidum* demonstrated by darkfield, DFA or equivalent method or • Currently has symptoms clinically consistent with secondary syphilis with reactive nontreponemal titer greater than or equal to 1:4
Latent	A stage of infection caused by *T. pallidum* in which organisms persist in the body of the infected person without causing signs or symptoms	• Asymptomatic with no history of syphilis with a reactive nontreponemal and treponemal test or • Past history of syphilis therapy and current nontreponemal test titer demonstrating fourfold or greater increase from the last nontreponemal test titer
Latent, early	A subcategory of latent syphilis when initial infection has occurred within the previous twelve months.	Latent syphilis classification with one or more of the following criteria: • Documented seroconversion or fourfold increase in nontreponemal titer during the previous twelve months. • History of symptoms consistent with primary or secondary syphilis within previous twelve months • History of sexual exposure to a partner with primary, secondary or early latent syphilis • Reactive nontreponemal and treponemal tests from a person whose only possible exposure occurred within the preceding twelve months

Table 14. *Continued*

Syphilis stage	Clinical description	Case classification
Latent, late	A subcategory of latent syphilis when infection has occurred more than one year previously	Latent syphilis classification with no evidence of having acquired disease within preceding 12 months (see Latent, early) and whose age (>35 years) and titer (<1:32) do not meet the criteria specified for latent syphilis of unknown duration.
Latent, unknown duration	A subcategory of latent syphilis when the date of initial infection cannot be established as having occurred within the previous year and the patient's age and titer meet the case classification	Latent syphilis classification that does not meet the criteria for early latent syphilis and patient: • is age 13-35 years and • has a nontreponemal titer greater than or equal to 1:32
Tertiary (late benign and cardio-vascular)	A stage of infection caused by *T. pallidum*, which usually becomes clinically manifest only after a period of 15-30 years of untreated infection.	Characteristic abnormalities or lesions of the cardiovascular system, skin, bone, or other structures with: • reactive treponemal test, • absence of other known causes of these abnormalities, and • without CSF abnormalities and clinical symptoms or signs consistent with neurosyphilis

Adapted from *STD Surveillance Case Definitions*, by CDC (2012), retrieved from http://www.cdc.gov/std/stats12/2012-casedefinitions.pdf, pages 141-142.

relates to the observable components of the study. The construct for this dissertation study was syphilis infection, and the proposition was that more useful syphilis testing (observable components) will improve syphilis detection. The current CDC case definition (Table 14) for diagnosing and staging syphilis infection required the use of both syphilis diagnostic laboratory testing and clinical correlation with a thorough patient clinical assessment. There are multiple FDA and non-FDA approved laboratory test methods to detect syphilis infection. This study used the results of three FDA approved tests, the Captia™ Syphilis IgG, Becton Dickinson MacroVue RPR, and Fujirebio Serodia TP-PA, whose intended use was to detect syphilis

infection, in conjunction with the 2012 CDC syphilis case definition to determine if syphilis infection (construct), either current or past, was present in the patient receiving care in a public health setting.

All syphilis detection and treatment procedures are performed within the US healthcare delivery system. In this system, syphilis tests are usually ordered by a licensed healthcare professional who will also diagnose and treat the infected patient. In recent years, there have been governmental changes as to who has the authority to order laboratory testing. On the federal level, the FDA determines whether a test kit can be utilized in the US for human testing; the CMS through the CLIA regulations oversees clinical laboratories to ensure they are providing accurate and reliable patient test results; and the Federal Trade Commission (FTC) investigates deceptive laboratory marketing practices. Each state determines who has the authority to order laboratory tests. State oversight of laboratory test ordering has changed in recent years with 37 states and the District of Columbia allowing the general public (consumers) to directly order some or all of their laboratory tests, without the involvement of a licensed healthcare provider. While the consumer has the ability to order laboratory tests, federal and state regulations still require that diagnosis and treatment, either invasive or with prescription medication, for an illness or condition must be provided by a licensed healthcare practitioner.

CLIA regulation does not differentiate between laboratories that perform patient direct access testing (DAT) or those that perform provider ordered testing; therefore, all laboratories must obtain and maintain a CLIA certificate that satisfies the federal regulations for all tests reported (American Association of Clinical Chemists[AACC], 2015; CMS, 2017; Ilahi, 2016). Whether the syphilis test was ordered by the patient or healthcare practitioner, diagnosis and treatment or follow up for an infection would still be provided by a healthcare practitioner; therefore, both the

laboratory and healthcare provider components are within the healthcare delivery system. The study proposition that more useful syphilis testing will improve syphilis detection would, therefore, require use of a theory that measures the healthcare delivery system.

According to Shi and Singh (2012), there are three interdependent components of healthcare delivery: cost, access, and quality. Cost and access are usually correlated, that is if access is increased, then cost also rises. However, quality can influence cost and access, with high-quality care often associated with cost effective care. Quality is a difficult concept to define and measure. The Institute of Medicine (IOM) defines health care quality as "the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge" (National Academies of Sciences, Engineering and Medicine [NASEM],2013, p. 1). According to a 2015 NASEM study, accurate and timely clinical diagnosis is a requirement for the delivery of high-quality healthcare. The study identified that diagnostic errors, defined as "the failure to establish or communicate an accurate and timely assessment of the patient's health problem" (p. 3) had an impact on the delivery of quality health care with five percent of US adults having one diagnostic error per year (National Quality Forum [NQF], 2017). Diagnostic errors could be a factor in about 10% of deaths per year and up to 17% of adverse hospital events. In order to characterize and potentially reduce these diagnostic errors, a draft framework was proposed by NQF to measure diagnostic quality and safety within the US healthcare delivery system. The framework included all six dimensions of quality identified by the IOM: safety, effectiveness, patient-centeredness, timeliness, efficiency, and equitability. The conceptual framework chosen by NQF was Donabedian's Quality Framework, which has been used since 1980 in multiple assessment studies of healthcare quality. A limitation of Donabedian's Quality Framework is that it does not

account for individual social behaviors and economic factors that are outside the healthcare delivery system. The focus of the framework is specifically on the linkage between those areas under the control of healthcare providers and the effect on patient outcomes. Focusing on areas that can be controlled by healthcare providers allows researchers to develop quality measures that can be monitored and revised to improve healthcare quality (NQF, 2017; Shi and Singh, 2012).

This dissertation study used Donabedian's Quality Framework, which is composed of structure, process and outcome domains, as the theoretical basis for evaluating the construct of syphilis infection and proposition that more useful syphilis testing will improve syphilis detection. The Donabedian model domains are hierarchical with structure influencing process and both domains affecting outcomes. For this dissertation, the Donabedian model was compared to the QMS model which is specific to laboratory workflow (figure 8). The two observable QMS model components utilized for the research hypotheses were the analytic (syphilis test results and S/CO values) and the postanalytic (traditional and reverse algorithm interpretations and syphilis diagnosis). An example of laboratory testing workflow within the Donabedian model would be as follows:

- Structure (preanalytic): a sample was improperly collected, handled and transported to the laboratory.
- Process (analytic): the laboratory utilized a test with poor sensitivity and/or specificity to perform analysis of the unsatisfactory sample resulting in either false negative or false positive test results.
- Outcome (postanalytic): the healthcare provider utilized the inaccurate and/or low sensitivity test results to determine the patient's diagnosis and treatment.

This outcome would have an impact on the quality of healthcare received by the patient and could lead to either under or over treatment, as well as have a potential financial and/or social impact on the patient and public health in the case of an infectious disease.

Polit and Beck (2012) state that construct validity also requires characterization of persons and settings utilized in the study to ensure that they appropriately represent the construct. The setting for this study was a large metropolitan western region public health agency which focused on detection and treatment of STDs within the community. The CLIA certified public health laboratory performing the testing was owned by the public health agency and specialized in STD testing. The public health agency provided diagnosis and treatment of STDs following CDC guidelines. Both settings, clinic and laboratory, were within the US healthcare delivery system. The study population included patients who sought care at the public health clinic due to contact with an STD, exhibited symptoms of an STD, or sought follow up care for an STD previously identified by an outside facility or public health agency outreach. The settings and population provided a homogeneous group of patients with a range of syphilis infection stages which enhanced the construct validity of the study.

Polit and Beck (2012) list the following five threats to construct validity of design:

- Reactivity to the study situation: refers to changes in study participants behavior if they are aware of their study role. For this dissertation study, all syphilis laboratory testing was performed by technologists who were blinded to the other syphilis test results as well as the reason for performing the tests. Additionally, the Captia^{TM} Syphilis IgG test was performed by an automated analyzer which provided an objective test result based on an optical reading performed by a computer. The RPR and TP-PA tests had subjective readings, but the public health laboratory required 100% agreement by two trained

technologists of the visual results before a final result was reported. Laboratory testing was ordered by the clinic staff using CDC recommended public health guidelines and study data was retrospectively gathered so staff were not aware of who was included in a study.

- Researcher expectancies: refers to the researcher's influence on participant responses to elicit a desired outcome. The study testing was performed in a real-world public health clinic setting following CLIA regulatory guidelines, established SOPs and QA processes. All study samples were based on a retrospective analysis which eliminated any influence by the researcher on the test results. The patients who had all three tests performed were chosen based on clinical practices at the public health agency STD clinic or samples were reflexed following laboratory protocols.
- Novelty effects: refers to use of a new treatment. The staff at the western region laboratory had been performing the RPR, RPR titer, and TP-PA testing since 2007 and Captia^{TM} Syphilis IgG since 2009. The reverse algorithm was initiated in 2009 over two years before the retrospective data was gathered. The testing was not new to laboratory staff and the laboratory director determined that all technical staff were competent to perform all tests. The STD public health clinic had been in operation for almost 50 years and had established policies for STD detection. The clinic staff followed the 2012 CDC syphilis diagnosis guidelines and were trained in use of the reverse syphilis algorithm when it was initiated at the public health laboratory in 2009.
- Compensatory effects: refers to compensation in intervention studies. This dissertation study did not include an intervention.

- Treatment diffusion or contamination: refers to blurring of alternate treatment conditions. There were no alternative treatment conditions related to this study. If the STD clinic ordered all three syphilis tests on a clinic patient it was based on their clinical assessment following standard public health practice and all three tests were performed as ordered. Typically, these patients were from high risk groups, such as MSM, drug use, contact with a person infected with an STD, exhibited syphilis symptoms or infection with another STD such as HIV, trichomonas, herpes, gonorrhea, or chlamydia. In the laboratory, all three syphilis tests were performed if the initial syphilis EIA was either reactive or equivocal. This was based on the western region public health laboratory director approved modification of the CDC reverse algorithm which added simultaneous TP-PA, RPR and RPR titer testing based on the reactive or equivocal EIA test. The CDC reverse algorithm recommended performing the TP-PA only if the reflex RPR was nonreactive. An additional modification allowed the laboratory to reflex nonreactive Syphilis IgG samples with a S/CO value greater than or equal to 0.450 to RPR and TP-PA testing.

External validity. External validity was enhanced because the study data was collected in a "real-world" public health clinic setting as part of standard public health practice for detecting syphilis infection (Polit & Beck, 2012). During the study period, the STD public health clinic staff followed standard syphilis surveillance and detection practices that were recommended by the CDC and are followed by multiple other US public health STD clinics. The public health laboratory followed the manufacturer's instructions for performance of all syphilis tests and operation of the BioRad EVOLIS analyzer. The settings of this study would, therefore, be representative of other public health settings where the findings might be applied.

Measurement Reliability and Validity

Measurement, whether subjective or objective, requires assignment of a value, often numerical, that represents the attribute amount being measured. Because attributes are not constant, there may be a measurable variability either by time (day to day) or by individual (person to person). In the healthcare field, assigning numbers to measurements allows healthcare providers to differentiate normal from abnormal values to assist in diagnosing ailments or conditions. Biophysical attributes such as temperature, weight, or blood pressure follow well established clinical guidelines so that two people performing the same measurement on the same patient should obtain the same result. For laboratory testing, the FDA establishes standards for tests that measure body system functions such as organ systems or infectious processes so that the test results are reproducible in different laboratory settings. Because biophysical attributes and laboratory testing allow for precise measurements, they can be accurately communicated and understood across multiple healthcare disciplines. Measurement reliability refers to how precisely the device measures the target attribute and does not require comparison to external criteria (internal consistency). Measurement validity refers to how accurately the device measures the target attribute and requires comparison to external criteria that reflect the target attribute (external consistency).

Any measurement can have an element of error and this error may be caused by influences that can or cannot be successfully controlled. The three main influences that can be sources of measurement error include:

- the attribute being measured: syphilis infection is detected by measuring antibody levels which can fluctuate depending on the illness stage and immune status of the infected person;

- the observer: the technologist performing the syphilis test was not trained appropriately, ignored the failed QC results and reported the potentially incorrect patient results;
- the measuring device: the laboratory staff did not perform required preventive maintenance on the immunoassay analyzer used to perform the syphilis test, which resulted in random instrument pipetting errors and potentially incorrect test results.

These influences can occur individually or all at once resulting in validity and/or reliability measurement errors. Some errors are random and are not attributable to a specific cause. Random errors typically affect the reliability of a measuring device. Systematic errors are usually due to a specific cause and typically fall in the same direction reflecting bias. These errors typically affect the validity of a measuring device. (Karras, 1997; Polit and Beck, 2012; University of Ottawa, 2017).

Example of interdependency of measurement validity and reliability. The interdependency of validity and reliability in assessing laboratory device measurements can be seen in the following hypothetical example. A hypothetical laboratory is assessing the validity and reliability of four different creatinine assays, one of which will be added to their current analytical instrument test menu. The laboratory utilized a reference creatinine standard (1.2 mg/dL) for external criteria comparison and performed repeat analysis of the standard to determine internal consistency, thus the laboratory was able to efficiently perform both validity and reliability assessments of the assays. Because it is difficult to obtain a 1.2 mg/dL value every time an analysis is performed due to instrument and/or assay variability, the laboratory established an expected range of 1.0-1.4 mg/dL for the reference standard. Using the reference standard, the laboratory performed 10 creatinine analyses for each assay method on the same instrument. The laboratory will consider an assay to be valid if the test results are within the

range established for the external reference standard. The assay will be considered reliable if it provides repeatable results. Basic statistical measures calculated in Table 15 provide information regarding the validity and reliability of each assay. The mean value reflects the average of the 10 reference standard analyses performed for each assay.

Table 15

Statistical Measures for Four Creatinine Assays Performed on the Same Instrument

	Assay 1	Assay 2	Assay 3	Assay 4
Creatinine (mg/dL)	1.2	0.5	0.9	0.6
	1.2	0.5	1.0	0.6
	1.2	0.5	1.0	1.3
	1.2	0.5	1.1	0.8
	1.2	0.4	1.2	1.4
	1.1	0.4	1.3	1.4
	1.1	0.6	1.3	0.7
	1.3	0.6	1.4	0.5
	1.3	0.6	1.5	1.6
	1.3	0.6	1.5	0.5
Mean	1.2	0.5	1.2	0.9
SD	0.1	0.1	0.2	0.4
Range	1.1-1.3	0.4-0.6	1.0-1.4	0.5-1.3
CV	6%	14%	17%	44%

The mean in Table 15 for Assay 1 and Assay 3 is 1.2 mg/dL which is comparable to the reference standard reflecting a valid measure. Assay 2 (mean 0.5 mg/dL) and Assay 4 (0.9 mg/dL) are both below the reference standard value and would be considered invalid measures. The standard deviation (SD) reflects the dispersion or variability of the test values around the mean. Assay 1 and Assay 2 both have an SD of 0.1 which reflects minimal test value variability, and each assay appears to provide reliable repeat measures. Assay 3 (SD 0.2) and Assay 4 (SD 0.4) show more variability and less reliable repeat measures. The assay range is calculated by

adding and subtracting the assay SD from the assay mean. The range for Assay 1 (1.1-1.3) and Assay 3 (1.0-1.4) are within the expected range reflecting accurate or valid results, while the other assays are outside the range. Assay 2 has a range 50% less than the expected range (0.4-0.6) and the range for Assay 4 is very large (0.5-1.3), both reflecting a lack of validity. The coefficient of variation (CV) describes the variation within a test and is calculated as CV = (SD/mean)100. The CV can be used to compare the precision of two methods, with the method providing the lowest CV considered to be the most precise or reliable. Assay 1 has the lowest CV (6%) and would be considered to be the most reliable of the four assays. A graphical display of the 10 reference standard analyses performed for each assay depicting the reliability and validity of each is shown in Figure 12.

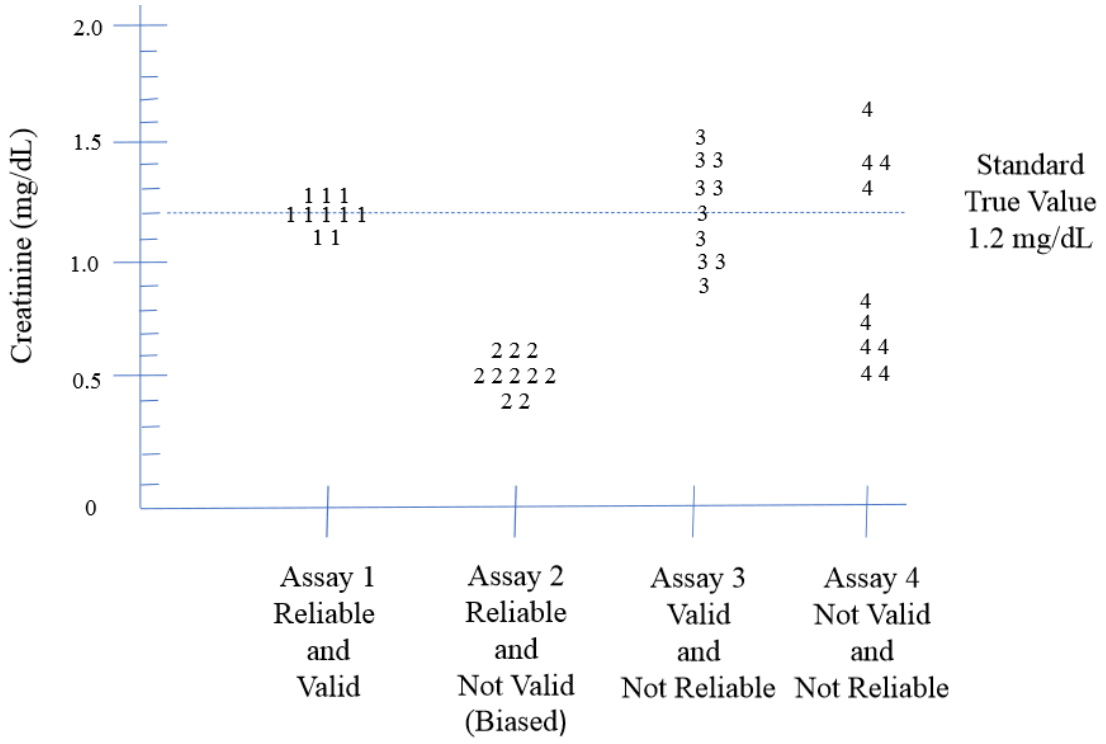

Figure 12. Graph of Validity and Reliability of Four Creatinine Assays Performed on the Same Instrument.
Adapted from "Statistical Methodology II: Reliability and Validity Assessment in Study Design, Part A," by D.J.Karras, 1997, *Academic Emergency Medicine, 4*, p.65. Copyright 1997 by John Wiley and Sons. Adapted with permission.

Assay 1 shows both reliability and validity which is reflected in the tight cluster of data points around the 1.2 standard value. The mean, SD, range, and CV statistical measures shown in Table 15 also reflect both reliability and validity. Assay 2 provided reliable, repeatable invalid results with an assayed value 50% less than the reference value, as seen in the cluster of values ranging from 0.4 to 0.6. Assay 3 provided valid results within the expected reference standard range; however, there is a wide dispersal of values and the high CV (17%) and SD (0.2) seen in Table 15 indicate it is less reliable than Assay 1. Assay 4 lacks both validity and reliability with a wide dispersal of data points both above and below the standard value. Table 15 shows that Assay 4 has the largest variability of test values (SD 0.4), largest range (0.5-1.3) and highest CV (44%) of the four assays (Karras, 1997; Polit & Beck, 2012; Zady, 1999).

According to Polit and Beck (2012), the reliability of a laboratory test can be evaluated using a test-retest method, in which the same test is administered on two or more occasions to the same subject. This method evaluates the stability of the test, which assesses how consistently or precisely the test measures the target attribute. As described in the hypothetical example above, both reliability and validity can be measured in a laboratory test using repeated measures of a sample of known value. For laboratory tests that report a quantitative or numerical value such as the hypothetical example above, standards of known value can be prepared and utilized for both reliability and validity studies. Immunoassay tests, such as the Captia™ Syphilis IgG, RPR, and TP-PA, report non-numerical or qualitative test results; and numerical values or concentrations are not included in the test report. Results are reported as reactive, nonreactive, equivocal, inconclusive, or invalid, depending on the test method. Known samples, either from an outside vendor or tested in house by a different method, that show the presence or absence of the test attribute can be used for both reliability and validity studies. The automated Captia™ Syphilis

IgG test method also provides a nonreportable numerical absorbance or optical density (OD) reading for each sample, which may be used as a measurement of test precision. The RPR titer is reported as an undiluted (1:1) or two-fold dilution number (1:2, 1:4, 1:8, 1:16, etc.) and is performed with the same reagent used for the qualitative RPR. The RPR titer also provides a numerical measurement of test precision.

Method verification. According to CLIA (2017) regulations, each US clinical laboratory that introduces an unmodified, FDA-cleared or approved test must establish and verify the test performance specifications prior to reporting any patient test results. The initial method verification must provide evidence that the accuracy, precision, reportable range, and reference intervals (normal range) of the test system meet acceptable parameters established by the laboratory director. The laboratory director may use the manufacturer's performance specifications as a guideline. For qualitative methods, the laboratory can test known positive and negative samples to verify that the method will identify the presence or absence of an analyte. Accuracy can be established by testing reference materials, comparing results of tests performed using an established reference method or comparing split sample results with results obtained from another comparable method. Precision can be established by assessing day-to-day, run-to-run, and within-run variation of the same samples as well as operator variance (CLIA, 2017; Pandori, 2017). Once the initial verification is completed and accepted by the laboratory director, the laboratory can utilize daily control procedures to monitor the accuracy (validity) and precision (reliability) of test performance over time. The daily control procedures should detect immediate errors due to test system failure, adverse environmental conditions, or operator performance. The frequency of control procedures must not be less than the frequency specified in the manufacturer's package insert. CLIA (2017) regulation CFR 493.1256(e)(1) also requires

that the laboratory must demonstrate and document that each lot number and shipment of reagents performs within the initial method verification specifications established for the test procedure. This requirement can be satisfied either by concurrent testing of reagent lot numbers (parallel testing) or acceptable QC results for the newly opened lot.

The western region public health laboratory successfully completed the initial method verification for the Becton Dickinson MacroVue RPR and Fujirebio Serodia TP-PA in May 2007. Initial method verification for the Captia™ Syphilis IgG test was successfully completed in November 2008. Following successful method verification, the laboratory director implemented performance of day of use QC as stated in the manufacturer's instructions for all syphilis tests with the addition of an external low titre reactive control for the Syphilis IgG test. As stated in the CLIA regulations, each reagent shipment was tested for acceptable qualitative (reactive/nonreactive) QC results when placed in use to identify any lot to lot variation in the test kits. Additionally, each box within the shipment was tested for acceptable qualitative QC results when it was opened. Documentation for syphilis reagent QC testing lot to lot and box to box is shown in Tables 16-18.

Measurement reliability. Reliability of the Captia™ Syphilis IgG test method was determined using the external low reactive titre control (BioRad Virotrol, Syphilis total), which was tested with each daily Syphilis IgG run. Table 16 shows that from January 2012 to December 2013, eight different reagent lot numbers in 13 shipments for a total of 57 boxes of Captia™ Syphilis IgG reagent were used with 100% (57/57) of the external low reactive titre control testing as expected with reactive results. Only the external control was used for reliability determination because the kit controls could only be used with the same kit; therefore, parallel testing was not performed. The kit controls were received at working strength and were

Table 16

Captia™ Syphilis IgG Quality Control Results January 2012 to December 2013

Captia™ Syphilis IgG reagent use dates	Syphilis reagent lot #	# boxes of reagent lot #	Valid Run Yes or No Expect	Nonreactive kit control (N)/each box Expect nonreactive	High reactive titre kit control (HRT)/each box Expect reactive	Low reactive titre external control/each box Expect reactive
1-3-12 to 2-10-12	IgG A	1	Yes	nonreactive	reactive	reactive
2-13-12 to 4-20-12	IgG B	5	Yes	nonreactive	reactive	reactive
4-23-12 to 6-6-12	IgG C	5	Yes	nonreactive	reactive	reactive
6-7-12 to 8-14-12	IgG D	2	Yes	nonreactive	reactive	reactive
8-15-12 to 9-25-12	IgG D	5	Yes	nonreactive	reactive	reactive
9-26-12 to 12-6-12	IgG E	5	Yes	nonreactive	reactive	reactive
12-7-12 to 2-25-12	IgG F	5	Yes	nonreactive	reactive	reactive
2-26-13 to 3-21-13	IgG F	5	Yes	nonreactive	reactive	reactive
3-22-13 to 5-22-13	IgG G	5	Yes	nonreactive	reactive	reactive
5-23-13 to 7-22-13	IgG H	5	Yes	nonreactive	reactive	reactive
7-23-13 to 9-24-13	IgG G	4	Yes	nonreactive	reactive	reactive
9-25-13 to 11-19-13	IgG G	5	Yes	nonreactive	reactive	reactive
11-20-13 to 12-31-13	IgG G	5	Yes	nonreactive	reactive	reactive
Total # of reagent boxes		57				

Table 17

Serodia TP-PA Quality Control Results December 2011 to January 2014

Serodia TP-PA reagent use dates	TP-PA reagent lot #	# boxes of reagent lot #	Valid Run Yes or No Expect Yes	Nonreactive kit control/each box Expect nonreactive	Reactive kit control/each box Expect reactive
12-23-11 to 1-23-12	TPPA 1	2	Yes	nonreactive	reactive
1-24-12 to 5-18-12	TPPA 2	10	Yes	nonreactive	reactive
5-21-12 to 8-29-12	TPPA 3	10	Yes	nonreactive	reactive
8-30-12 to 11-16-12	TPPA 4	10	Yes	nonreactive	reactive
11-19-12 to 3-4-13	TPPA 5	10	Yes	nonreactive	reactive
3-5-13 to 6-21-13	TPPA 6	5	Yes	nonreactive	reactive
6-24-13 to 8-2-13	TPPA 7	5	Yes	nonreactive	reactive
8-5-13 to 10-30-13	TPPA 8	5	Yes	nonreactive	reactive
10-31-13 to 1-8-14	TPPA 9	5	Yes	nonreactive	reactive
Total # of reagent boxes		62			

Table 18

MacroVue RPR Quality Control Results December 2011 to January 2014

MacroVue RPR antigen use dates	RPR antigen lot #	# boxes of antigen lot #	Valid Run Yes or No Expect Yes	Nonreactive external control/each box Expect nonreactive	Minimally to Moderately reactive external control/each box Expect reactive	Highly reactive external control/each box Expect reactive
12-20-11 to 1-18-12	RPR 1	2	Yes	nonreactive	reactive	reactive
1-19-12 to 5-29-12	RPR 2	10	Yes	nonreactive	reactive	reactive
5-30-12 to 8-27-12	RPR 3	7	Yes	nonreactive	reactive	reactive
8-28-12 to 1-11-13	RPR 4	10	Yes	nonreactive	reactive	reactive
1-14-13 to 2-9-13	RPR 5	2	Yes	nonreactive	reactive	reactive
2-11-13 to 7-12-13	RPR 6	10	Yes	nonreactive	reactive	reactive
7-15-13 to 1-7-14	RPR 7	10	Yes	nonreactive	reactive	reactive
Total # antigen boxes		51				

pre-diluted. They had to be placed in a specific location on the 96 well plate in order for the instrument to calculate a cutoff value and perform test plate validation. The pre-diluted kit controls from a prior lot number could not be used with a different kit. If the previous lot controls were placed in a patient sample location for parallel testing, they would have been diluted again with potentially incorrect test results. However, Table 16 shows that the nonreactive and high reactive kit controls exhibited 100% expected results of nonreactive (57/57) and high reactive (57/57) for all reagent lot number boxes tested.

Reliability of the Serodia TP-PA test method was determined using the kit controls from the old lot tested with the first box of kit controls from the new lot (parallel testing). There was no external control utilized with the TP-PA kit, and only qualitative results were reported. Kit controls also were performed each day of use. Table 17 shows that from December 2011 to January 2014, nine different lot numbers of Serodia TP-PA kits were used with 100% (18/18) of the kit controls in each new lot number testing as expected (nonreactive and reactive) across the different lot numbers.

Reliability of the MacroVue RPR qualitative test method was determined using an external three member QC card test (Becton Dickinson RPR Control card) when each test kit box was opened. The RPR kit did not come with kit controls. The QC card test was also performed each day of use. Table 18 shows that from December 2011 to January 2014, seven different lot numbers of MacroVue RPR kits were used for a total of 51 boxes. Each box was tested with three external controls with 100% (153/153) of the external controls testing as expected (nonreactive, reactive and reactive) across the seven different lot numbers,

Lot to lot variability. The syphilis tests used in this dissertation consistently detected the presence or absence of syphilis antibody (100% for all methods) and provided expected qualitative results as listed in the QC studies shown in Tables 16-18. There have been reports of variations in different lots of immunoassay reagents that are not detected by routine qualitative QC testing of reactive and nonreactive samples. These variations could involve manufacturing changes such as in the raw test materials, in the concentration of antigen or antibody bound to the solid phase, or in the release panel used by the manufacturer to determine the minimum level of reactivity in the new lot. Other variations could be due to reagent deterioration during transportation or storage. Variations between lots may result in shifts or trends of patient test results that are not readily apparent in presence/absence qualitative QC testing. Suggestions for reducing the amount of lot-to-lot variability include method standardization and automation of the procedure. All three syphilis test methods used by the western region laboratory were FDA approved and standardized by the manufacturer. The Captia^{TM} Syphilis IgG test method was performed on an automated analyzer. Other suggested practices include use of only QC materials for a historically stable test method or use of QC materials and previously analyzed patient samples for a method that historically shows shifts in patient values with lot number

changes. Each laboratory director must determine the best method for their laboratory environment and test menu to ensure that the method performs as initially verified (Alegeciras-Schimnich et al., 2013; Alegeciras-Schimnich, 2014; Dufour, 2004; Ezzelle et al, 2008; Holzel, 1991; Karras, 1997; Kim et al, 2012; Kitchen & Newham, 2010; Martindale et al, 2006). In September 2013, CLSI published a guideline EP26-A User Evaluation of Between Reagent Lot Variation to provide a tool for laboratories to screen new reagent lots for performance differences. According to the guideline, whatever reagent lot assessment is used by the laboratory, it should be able to be readily performed when needed (lot changes can happen at any time); use a simple protocol with simple arithmetic (multiple assessment criteria may be needed); and utilize clear, previously established acceptance criterion (Addison, 2017).

The Bio-Rad Virotrol Syphilis total external control utilized with each daily run of the Captia™ Syphilis IgG testing performed at the western region public health laboratory was an unassayed precision control for the qualitative determination of IgG and IgM antibodies to *T. pallidum* and reagin. The control was treated exactly like a patient sample. Bio-Rad did not provide concentration values or expected ranges for this control; however, it was optimized to provide a low to moderate reactive qualitative test result. Each laboratory must establish its own target ranges for each control lot number if used as a quantitative test (Bio-Rad, 2010). The Bio-Rad Virotrol product provided a consistent, optimized control material that could be utilized to detect lot-to-lot Syphilis IgG reagent variation. As seen in Table 16, the product consistently provided qualitative reactive results which were calculated by the EVOLIS analyzer from the S/CO value with each plate analysis. The EVOLIS instrument provided OD or absorbance numerical readings for each Virotrol control sample that can be used for statistical analysis. Various authors have utilized the patient S/CO value to track lot-to-lot variation; however, there

is little available information on the use of control OD values for the same function (Dufour, 2003; Kitchen & Needham, 2010). For this dissertation, the external Virotrol control OD values were utilized rather than S/CO values to provide a stable value that could be tracked over multiple reagent lot numbers. The S/CO value was calculated from the single point calibrator cut off value which changed with each test plate and lot number, thus adding another level of variability to the statistical analysis. This additional variability would make it difficult to identify a meaningful variance in kit lots (Dufour, 2003). Because the Virotrol product was optimized by the vendor as a precision control, the OD value reflected stable product activity following completion of all analytical steps up to calculating the qualitative test result interpretation using the S/CO value.

According to CLSI, the laboratory should choose a lot-to-lot assessment that can be easily performed and uses simple math with multiple statistical measures (Addison, 2017). At present, there is no universally agreed upon assessment measurement or acceptance/rejection criteria for any of the immunoassay lot-to-lot variation quantitative statistical measures that could be performed; therefore, a combination of measures were used for this dissertation (Algeciras-Schimnich et al, 2013; Dufour, 2004; Kim et al, 2012; Martindale et al, 2006). The BioRad Virotrol control OD mean and SD were easily calculated for each reagent lot number utilizing Microsoft Excel. While a range can be calculated from the mean and SD, it may not be helpful for comparing lot numbers due to the lot variation that is expected with immunological assays. The recommended statistic to use when comparing a measured value to a measured value is the percent mean difference (Kim et al, 2012; Kitchen & Newham, 2010). The percent mean difference was calculated by Microsoft Excel using the formula:

$$\% \text{ mean difference} = (\text{value 1-value 2})/([\text{value 1} + \text{value 2}]/2)*100$$

Multiple authors (Addison, 2017; Algeciras-Schimnich et al, 2013; Kim et al, 2012) have suggested acceptance criteria for the percent mean difference between reagent lots should be less than 10% and that criteria was utilized in this dissertation.

A second statistic, the F-test can be used to compare the variance of one test method (or lot number) with the variance of a comparative method (or lot number). The variance is the square of the SD. The F-test is used for random error or imprecision analysis and calculated as shown below where the larger SD^2 is the less precise method and the smaller SD^2 is the more precise method.

$$\text{Calculated F-value} = (\text{Larger SD})^2/(\text{Smaller SD})^2$$

The F-test is interpreted by comparing the calculated F-value with a critical F-value that is obtained from a statistical table using n-1 degrees of freedom for the numerator and denominator measures. The calculated F-value was determined using Microsoft Excel, and a statistical table was obtained from an online statistical calculator found at www.danielsoper.com/statcalc to obtain the critical F-value at 95% confidence level. The null hypothesis being tested at the 95% confidence level was that there was no difference between the variances of the two lot numbers. If the calculated F-value was greater than the critical F-value, the difference in variances was considered to be statistically significant and the null hypothesis was rejected. The F-test interpretation does not imply that the statistically significant value was incorrect, only that the variance of the two lot numbers was different. For this lot-to-lot variance assessment, it was expected that the calculated F-value would be less than the critical F-value (Klick, 1997; Soper, 2017; Texas Department of State Health Services Laboratory, 2014).

The third statistical test in the lot-to-lot variation assessment was the CV which expresses variation as a percentage of the mean. Due to the known variability of immunoassay methods,

the recommended measure for comparing overall precision is the CV (Reed, Lynn & Meade, 2002). The CV was calculated as CV = (SD/mean)100. The SDs of assays that produce continuous type values, such as the OD, increase or decrease in proportion to concentration. In the CV calculation, dividing by the mean removes the SD as a variability factor and the calculation is, therefore, considered to be a statistical equalizer. There is no universally agreed upon acceptance or rejection criteria for a CV value with various authors suggesting a CV of 5% as good and CV of 10% or higher as poor. However, the mean value must be taken into consideration when evaluating CVs. At very low concentrations, such as a bilirubin standard with a mean of 0.5 mg/dL and SD of 0.1 mg/dL, the CV may be 20% while at a high concentration, such as a bilirubin standard of 20 mg/dL and SD of 1.0 mg/dL, the CV may be 5% (Zady, 1999). There was only one published study, Kitchen and Newham (2010), that performed a lot to lot performance assessment of syphilis EIA tests using a 10-member panel (nine reactive and one reactive) created by the authors from previously tested patient samples with varying stages of syphilis infection. This study utilized mean, SD and CV calculations to compare the S/CO values of the 10-member panel across more than 30 lot numbers of two different syphilis EIA tests and the following results were reported:

- Abbott ICE syphilis: nine reactive samples had a mean S/CO range of 0.95-11.7; SD range 0.17 -1.78 and CV range 9.4-23.2 with an average CV of 19%. The one nonreactive sample had a mean S/CO of 0.41, SD of 0.13 and CV of 32.1%.
- bioMerieux syphilis TP recombinant: nine reactive samples had a mean S/CO range of 0.7-6.98; SD range 0.17-1.68 and CV range 18.8-40 with an average CV of 24%. The one nonreactive sample had a mean S/CO of 0.38; SD of 0.12 and CV of 30.9%.

As seen in the syphilis test comparisons above, the nonreactive sample had a CV greater than 30% with either assay. For the reactive samples, the lowest CV value was 9.4% with the average CV of the two EIA test methods at 19% and 24%. Based on this published study and the use of a low reactive control sample, an expected CV of less than 20% was used for this dissertation.

During the study period from January 2012 to December 2013, the western region laboratory used the laboratory director approved lot-to-lot QC protocol requiring acceptable performance of qualitative control results (reactive or nonreactive) with each new lot number and during day to day use along with the manufacturer's test plate validation requirements for the Syphilis IgG test. As shown in Table 16, all Captia^{TM} Syphilis IgG qualitative QC results were acceptable for all eight lot numbers of reagent; and all runs were valid. For this dissertation, a retrospective analysis was performed of the eight Captia^{TM} Syphilis IgG lot numbers received during the study period, using the OD values for the Bio-Rad Virotrol external control to identify if there was any lot number variance. The external control was required to be performed with each run; two lot numbers of control were used during the study period. Based on review of published studies, the expected values for this assessment were defined as % mean difference of less than 10%; calculated F value less than critical F value; and CV less than 20%. The expectation was that at least two of the three assessment criteria would meet the expected values thus showing consistent test results and no variation between the old and new lot. Review of Table 19 data shows that seven of eight lot numbers were within expected values for all three assessment criteria including acceptable results with the control lot change in August 2012 with the old control lot number tested against the new control lot number using the same reagent lot number IgG D. Captia^{TM} Syphilis IgG lot number IgG H met the expected % CV value at 11%; however, it did not meet the expected values for the other two assessment statistics with % mean difference of 22% and

Table 19

Captia™ Syphilis IgG Lot-to-Lot External Control OD Variation Assessment

Captia Syphilis IgG reagent use dates	Syphilis reagent lot #	Bio-Rad Virotrol Syphilis Total External Control Lot #	Number measures	Mean S/CO value external control	Mean optical density (OD) external control	Mean Difference lot # OD[a]	% Mean Difference (expect <10%)[b]	Standard deviation (SD)	SD squared	F value calculated[c] (expect F value calculated < critical F value)	Critical F value[d] (p=0.05)	Coefficient of variation (CV)[e] (expect <20%)	Syphilis lot # comparison for % Difference and F value calculated
1-3-12 to 2-10-12	IgG A	Syph Total 1	34	2.004	1.418	-0.123	-6%	0.146	0.021	1.52	1.72	10%	A vs B
2-13-12 to 4-20-12	IgG B	Syph Total 1	53	1.968	1.541	0.166	7%	0.180	0.032	1.19	1.70	12%	B vs C
4-23-12 to 6-6-12	IgG C	Syph Total 1	36	2.728	1.375	-0.054	-3%	0.163	0.027	1.08	1.68	12%	C vs D
6-7-12 to 8-14-12	IgG D	Syph Total 1	46	2.747	1.429	0.079	4%	0.158	0.025	1.47	1.76	11%	D (old ct lot) vs D (new ct lot)
8-15-12 to 9-25-12	IgG D	Syph Total 2	32	2.661	1.350	0.103	5%	0.131	0.017	1.21	1.68	10%	D (new ct lot) vs E
9-26-12 to 12-6-12	IgG E	Syph Total 2	52	2.297	1.247	0.103	6%	0.118	0.014	1.00	1.55	9%	E vs F
12-7-12 to 3-21-13	IgG F	Syph Total 2	71	2.309	1.144	-0.045	-3%	0.120	0.014	1.29	1.43	10%	F vs G
3-22-13 to 5-22-13 and 7-23-13 to 12-31-13	IgG G	Syph Total 2	136	2.434	1.189	0.357	**22%**	0.133	0.018	**4.50**	1.62	11%	G vs H
5-23-13 to 7-22-13	IgG H	Syph Total 2	36	1.663	0.832	NA	NA	0.064	0.004	NA	NA	8%	**lot #H returned to vendor and replaced with lot #G**
Total Measures			496										

Note: ct = control; S/CO = signal to cutoff

[a] mean difference = (lot 1-lot 2);

[b] % mean difference = (lot 1-lot 2)/([lot 1 + lot 2]/2)*100;

[c] Calculated F-value = $(\text{Larger SD})^2/(\text{Smaller SD})^2$;

[d] www.danielsoper.com/statcalc;

[e] CV = (SD/mean)10

calculated F value of 4.32 which is higher than critical F value of 1.62. According to the qualitative QC values seen in Table 16, lot number IgG H was acceptable for use based on valid run performance and acceptable qualitative QC results. The mean external control S/CO values for all eight reagent lot numbers were above the manufacturer's cutoff value of 1.1 for a reactive test result. While lot number IgG H had a low OD value, the mean S/CO value of 1.663 provided acceptable reactive test results.

Syphilis IgG lot number H was utilized for patient testing from May 23, 2013 through July 22, 2013. While the external control S/CO value was within acceptable limits for this lot number, further review of the samples analyzed during that time period was performed to determine if the lower OD value may have resulted in higher numbers of false negative Syphilis IgG test results. There were 429 samples analyzed during that time period with 358 (84%) of the samples from patients diagnosed with no syphilis infection and a nonreactive Syphilis IgG test result; 44 (10%) from patients diagnosed with syphilis infection and a reactive Syphilis IgG test result; and 27 (6%) from patients diagnosed with syphilis infection and nonreactive Syphilis IgG test result. Further analysis of these 27 nonreactive samples revealed that the mean S/CO value was 0.684, which was well below the manufacturer's cutoff value of 0.9 for an equivocal test result. There were 67% (18/27) concordant (RPR nonreactive) samples within this group of nonreactive Syphilis IgG test results. Further review of the syphilis stages within these 27 samples revealed that eight patients were diagnosed with primary syphilis, one with secondary syphilis, 13 with latent syphilis, and five with a history of syphilis. All 18 of the concordant RPR nonreactive samples were diagnosed with either latent syphilis infection or a history of syphilis infection. The nine patients diagnosed with either primary or secondary syphilis had a reactive RPR and TP-PA test result. Based on this data, there were 2% (9/429) of the samples

during this time period that were syphilis IgG nonreactive with RPR and TP-PA reactive test results and a diagnosis of syphilis infection.

For comparison purposes, the same data analysis was performed on Syphilis IgG lot number IgG F which had the next lowest external control OD value (1.144) of the eight reagent lot numbers used for analysis. Lot number IgG F was utilized from December 7, 2012 through March 21, 2013 and used the same external control lot number 2. There were 542 samples analyzed during that time period with 412 (76%) of the samples from patients diagnosed with no syphilis infection and a nonreactive Syphilis IgG test result, 101 (19%) from patients diagnosed with syphilis infection and a reactive Syphilis IgG test result, and 29 (5%) from patients diagnosed with syphilis infection and nonreactive Syphilis IgG test result. Further analysis of these 29 nonreactive samples revealed that the mean S/CO value was 0.621 which was well below the FDA approved cutoff value of 0.9 for an equivocal test result. There were 52% (15/29) concordant (RPR nonreactive) samples within this group of nonreactive Syphilis IgG test results. Further review of the syphilis stages within these 29 samples revealed that one patient was diagnosed with primary syphilis; one with secondary syphilis; 22 with latent syphilis and five with a history of syphilis. The 15 concordant RPR nonreactive samples were diagnosed with either latent syphilis infection or a history of syphilis infection. The remaining 14 patients were diagnosed with syphilis with a reactive RPR and TP-PA test result. Based on this data, there were 3% (14/542) of the samples during this time period that were syphilis IgG nonreactive with RPR and TP-PA reactive test results and a diagnosis of syphilis infection.

The comparison of the two different Syphilis IgG lot numbers indicates both lot numbers had similar S/CO values for nonreactive patient samples which were well below the 0.9 manufacturer equivocal cutoff value. The Syphilis IgG lot IgG H returned to the manufacturer had a slightly

higher frequency (6%) of samples that had a nonreactive Syphilis IgG with a diagnosis of syphilis versus lot IgG F with a frequency of 5%. Based on this data, there did not appear to be a higher than expected frequency of false negative Syphilis IgG samples with lot number IgG H. Additional retrospective review of the EVOLIS analyzer was performed to identify if there were any instrument reliability issues during the time period that lot number IgG H was in use.

EVOLIS instrument reliability. According to an FDA Inspection Technical Guide (1980), all manufactured products have a measure of reliability called the "failure rate" which is the number of malfunctions after a number of uses for non-continuously operating products. The failure rate occurs in three predictable stages over the life of the product or equipment. First, there is a period of initial failure when the equipment is installed. Failures are high during this phase and are usually due to weak components or manufacturing flaws. This period typically lasts one year; and failures rapidly decrease as repairs are successfully completed, and the equipment settles in to routine use. The second phase occurs during the useful life of the equipment with random failures occurring at low rates for reliable, stable equipment. This phase typically lasts one to six years, but can be extended to 10 years or more with appropriate preventive maintenance. The last phase is the wear out stage when the failure rate begins rapidly climbing again. This can be due to general physical deterioration or unreplaceable parts resulting in functioning of the equipment at an unacceptable level. Once in the wear out stage, the equipment is typically decommissioned and replaced with a newer model (FDA, 1980; Japan International Cooperation Agency [JICA], 2017). While the equipment is in operation, an analysis is typically performed when a failure occurs to identify the cause. This analysis includes identifying the symptom, locating the fault, checking the operating environment, reviewing equipment history, and analyzing operator training and experience. CLIA (2017)

requires that a record of laboratory equipment failures be maintained; however, there is no specified equipment failure rate. There is little published information on acceptable equipment failure rates; however, JICA (2017) reviewed statistics on medical equipment failure rates and identified that a 5-6% failure rate was associated with random instrument errors.

The EVOLIS analyzer was installed at the western region laboratory in 2008 and was within useful equipment life (one to six years) during the study period of 2012-2013. During this time period, the failure rate would be expected to be low, with random errors occurring at a rate of 5-6%. In addition to the routine preventive maintenance performed by laboratory staff, the vendor, Bio-Rad, also performed scheduled six-month service which would contribute to extending the life of the instrument to 10 years. The EVOLIS service record during the study period is shown in Table 20. According to Table 20, there were 18 instances of instrument failure noted during the study time period. Patient sample test results from failed instrument runs were not reported. Once the cause of the instrument failure was identified and corrected, the samples were placed on the analyzer for repeat testing and, provided all QC requirements were met, the repeat results were reported.

The EVOLIS instrument was used daily (Monday through Friday). Annual test days were calculated at 52 weeks x 5 days = 260 days/year excluding 12 government holidays/year for 248 total annual testing days or 496 (248 x 2) testing days for the study two-year time period. Therefore, the instrument failure rate for the two-year study time period would be 3.6% (18/496 x 100), which is well below the 5-6% random error rate identified by JICA (2017) indicating that the instrument was performing reliably on repeated testing days.

Three instances were noted in the instrument service record in Table 20 that involved invalid test results within valid assay runs. As described in Figure 9, an invalid test result occurred

Table 20

EVOLIS Analyzer Service Record January 2012 to December 2013

Date	EVOLIS analyzer issue	Corrective action
1/3/12	Plate transport error-failed run	Cleaned and realigned - corrected and re-run
4/16/12	Semiannual Preventive Maintenance by BioRad	Completed and QC passed
4/17/12	Failed HIV and Syphilis run	Contaminated wash bottle - cleaned and replaced-corrected and re-run
4/30/12	Invalid HIV and Syphilis results	Dirty manifold and crimped tubing - cleaned and replaced - corrected and re-run
5/4/12	Erratic washer dispense-failed run	Cleaned and realigned - corrected and re-run
6/1/12	Invalid HIV and Syphilis results	Contaminated wash bottle- cleaned and replaced - corrected and re-run
6/18/12	Open door error-failed run	Cleaned door - corrected and re-run
7/12/12	Transport error-failed run	Dirty rails-broken clasp - cleaned and replaced - corrected and re-run
7/17/12	Transport error-failed run	Realigned-corrected and re-run
8/29/12	Open door error-failed run	Replaced door assembly-corrected and re-run
8/30/12	Reagent rack not registering-failed run	Cleaned rack rail - corrected and re-run
11/1/12	Semiannual Preventive Maintenance by BioRad	Completed and QC passed
11/20/12	Tip eject failure-failed run	Replaced transfer arm and sensor-corrected and re-run
3/1/13	Transport error-failed run	Replaced carrier-corrected and re-run
3/18/13	Invalid Syph results	Contacted manufacturer who will investigate-possible instrument issue
3/22/13	Plate transport error-failed run	Replaced broken spring-corrected and re-run
3/26/13	Ambient temp failure-failed run	Replaced broken and defective plates-corrected and re-run
3/28/13	Pump leak error-failed run	Replaced pump-corrected and re-run
4/16/13	Plate transport error-failed run	Realignment-corrected and re-run
5/3/13	Semiannual Preventive Maintenance by BioRad	Completed and QC passed
5/31/13	Plate transport and carrier error-failed run	Tightened drawer-corrected and re-run
6/5/13	Multiple Syphilis and HIV invalid	Probe adjustment and recalibration-corrected and re-run
7/8/13	Washer dispense pump error-failed run	Tightened pump - corrected and re-run
7/12/13	Failed run and short samples	Realignment-corrected and re-run
7/15/13	Increase equivocal Syphilis results noted	Contaminated reagent and washer fluid - clean and replace-corrected and re-run
7/16/13	Increase invalid and equivocal syphilis results noted in June and July with lot #IgG H	Contacted vendor, investigated and Lot #IgG H returned to vendor, and replaced with lot #IgG G
12/11/13	Semiannual Preventive Maintenance by BioRad	Completed and QC passed
12/13/13	Miscommunication with LIMS-run passed manual data entry	Replace coax cable - run passed - manual data entry in LIMS
12/20/13	Power failure-failed run	Replaced APC unit-corrected and re-run

Table 20. *Continued*

EVOLIS failed run comments	During March 2013 there were a number of EVOLIS mechanical issues resulting in failed runs. Also during March, six invalid syphilis results were seen within 1 week time period using Syphilis lot #IgG G with all plate runs testing as valid. Vendor contacted and they investigated. Invalids possibly due to instrument errors which were corrected and invalid samples decreased. Per laboratory policy, invalid samples were confirmed with RPR and TPPA, additional recommendation sent to submitter to recollect sample if RPR and TPPA nonreactive. The entire valid plate runs containing the invalid results were also repeated to ensure reproducible results. All results confirmed as initially reported. During June and July 2013, additional invalid and increased equivocal syphilis results seen with Syphilis lot #IgG H. Invalid samples reported from valid plate runs based on FDA approved manufacturer's instructions. Request investigation by vendor. The entire valid plates containing the invalid results were repeated to ensure reproducible results. All results confirmed as initially reported. In July 2013, vendor requested return of lot #IgG H due to potential for low OD readings with some of the microtiter reagent strips which could cause invalid or equivocal test results within a valid plate run. Reagent lot #IgG H returned to vendor and replaced with lot #IgG G. No further problems with invalid or equivocal Syphilis results with lot #IgG G

when duplicate repeat syphilis test results performed automatically for an initial reactive or equivocal were not repeatable. If an initial reactive syphilis test result had nonreactive, nonreactive or nonreactive, equivocal repeat test results, then the final result was reported as invalid. For an initial equivocal test result, an invalid test result was reported if the duplicate repeat results were nonreactive and reactive. With all runs that had invalid test results during the study period, the QC was within expected limits and each assay plate run was valid. Therefore, according to the manufacturer's instructions, a patient result was reported as invalid if the repeat results correspond to the algorithm shown in Figure 9. Invalid results could be due to instrument, reagent, and/or operator error and cannot be interpreted as either reactive or nonreactive. Invalid test results should occur infrequently. At the western region laboratory, the laboratory director approved reflex testing of invalid Syphilis IgG EIA test results to RPR, RPR titer, and TP-PA in order to prevent delay in patient diagnosis. In that scenario, the patient sample was tested following the traditional algorithm because an invalid Syphilis IgG test result

has no clinically diagnostic value for the healthcare provider. Additionally, because of concerns about instrument reproducibility, all samples on the 96 well plate with invalid results were repeated to ensure all test results were reproducible, even though the plate met all QC and valid run requirements. Only the results of the repeat analysis were reported and included in the dissertation dataset.

In March 2012, laboratory staff noted invalid results with lot number IgG B, prompting additional instrument operational review by laboratory staff resulting in the identification of crimped tubing. The instrument issue was corrected, and all samples on the invalid plate were repeated with reproducible results. Only the repeat results were reported for all patients; however, the invalid sample still had to be reported as invalid due to the Western Region Laboratory Captia^{TM} interpretation algorithm (Figure 9). In March 2013, laboratory technologists noted six invalid results in a one week time period with lot number IgG G, which resulted in identification and correction of multiple instrument errors. Starting in June 2013, laboratory technologists identified invalid test results with lot number IgG H, which led to equipment adjustments; however, the invalid test results continued and were accompanied by an increase in equivocal results leading to further investigation by the vendor.

The vendor identified potential deterioration of lot number IgG H and requested return of all reagent kit lot number boxes. It was suspected that the deterioration may have occurred during reagent shipping as there were no reports of excessive invalid test results from other laboratories using the same lot number. The returned boxes were replaced with lot number IgG G with a corresponding decrease in invalid and equivocal test results. All runs with lot number IgG H met qualitative QC and valid assay plate run requirements, which, according to the vendor, demonstrated that all patient test results performed with lot number IgG H were valid and

reportable. Additionally, the follow-up study performed for this dissertation identified that lot number IgG H performed similar to prior lot number IgG F when used for syphilis testing within the public health clinic study population. It was the astute daily review of patient results by laboratory staff along with knowledge of historical performance of the assay that alerted staff to the potential for performance issues with this specific lot number, even though patient result reporting was not compromised. According to Algeciras-Schimnich, et al (2013), it is often the healthcare provider that raises concern about immunoassay test results when correlating test results to the patient's clinical condition.

Based on the information presented in Table 20 regarding the vendor's investigation of the invalid test results and subsequent request for return of remaining lot number IgG H, it appears that the lot-to-lot variation assessment measures shown in Table 19 would have identified a potential performance issue with lot number IgG H due to the failure of two of three measures (% mean difference 22% and F value 4.50) to meet acceptance criteria. The assessment measures are easily calculated and may provide a useful tool for detecting lot-to-lot variations in immunoassay reagents. Using only the qualitative QC values, the potential performance issue would not have been discovered unless laboratory staff had noticed the OD decrease in the external control with lot number IgG H. The healthcare providers did not notify the laboratory of any inconsistency with laboratory results and clinical assessment during the time period that lot number IgG H was in use. According to the manufacturer, because the QC was valid and each assay plate run passed validation requirements, all lot number IgG H results were acceptable for reporting and were included in the study data.

RPR titer reliability. The RPR titer was performed using an undiluted sample and serial two-fold dilutions of the initial RPR qualitative reactive sample. The titer dilutions were

repeated for the RPR test procedure until the titer dilution reached a nonreactive endpoint. The final test result was reported as the highest dilution value showing a reactive test result. The RPR titer is a numerical value which doubles in value as it increases and was reported as 1:1 (undiluted), 1:2, 1:4, 1:8, 1:16, 1:32, etc. According to CLIA (2017) regulations CFR 493.1256(d)(3)(iii), the RPR titer test procedure must include a control material with graded or titrated reactivity. The RPR titer uses the same antigen as the qualitative RPR test; therefore, it provides a numerical result for determining the reliability of the RPR test.

The western region laboratory utilized an external commercial control product, Blackhawk BioSystems Virotrol RPR Panel, to meet the CLIA regulations regarding inclusion of a control with titrated reactivity. The Virotrol RPR panel is an unassayed QC panel for the qualitative determination of nontreponemal (reagin) antibodies (Blackhawk BioSystems, Inc., 2003). The panel consisted of three members with the following expected reactivity profile with the MacroVue RPR reagent: panel member 01 nonreactive, panel member 02 minimally reactive, and panel member 03 moderate to high reactive. Only panel member 03 which provided the highest reactivity was used for RPR titer QC at the western region laboratory. According to the Virotrol package insert, each laboratory must establish its own target range for quantitative RPR testing by performing replicate measurements over multiple days with multiple operators. Results from replicate measurements were used to calculate a mean and standard deviation to establish a target range. The western region laboratory determined that the Virotrol RPR control was extremely stable with a long outdate of 18-24 months; therefore, it purchased large batches of the same lot number sufficient to last for 6-12 months. When a new lot number of control was received, parallel testing with the old control lot number was performed by multiple laboratory staff for 30 days. A mean and SD were calculated and a target range of mean +/- 2SD was

determined for each control lot number. The RPR titer was reported in two-fold dilutions; therefore, to calculate a mean and SD, the titer value was converted to a whole number with a titer value of 1:1 changed to a numerical value of 1; 1:2 changed to 2; 1:4 changed to 4; and 1:8 changed to 8, etc. After each parallel test value was converted to a whole number designation, the mean and SD were calculated for the new lot number using Microsoft Excel. Once the new lot number target range was established, the whole numbers were converted back to a two-fold ratio so that the range could be used for routine QC testing. For example, Table 21 shows that Virotol lot number 1 had a mean of 7.4 and SD of 1.9 which calculates to a +/- 2 SD target range of 3.6-11.2. This represents an RPR titer equivalent of 1:4 -1:8. Within the target range, 3.6 represents the reactive two-fold titer of 1:4 and 11.2 represents the last two-fold reactive titer of 1:8. In routine testing, the reactive two-fold titer after 1:8 would be 1:16. As 16 is higher than the upper end of the target range (11.2), a 1:16 titer would give a nonreactive test result; therefore, the highest reactive titer result would be 1:8. Based on historical data, the laboratory determined that an acceptable control range for each new lot number was a titer result of 1:4 to 1:8. Three lot numbers of Virotrol control were utilized during the two-year study period with seven different MacroVue RPR reagent lot numbers included in the study (Table 17). The range for all three lots (Table 21) was within the laboratory expected historical 1:4 to 1:8 titer range with 100% reliability (90/90).

Equivalence. According to Polit and Beck (2012), equivalence is another reliability aspect that could be assessed. Equivalence refers to how well two or more independent observers agree on an instrument score or rating with a high level of agreement associated with minimal measurement error. The manual RPR, RPR titer, and TP-PA tests require visual reading of the final test result which could be a source of measurement error. One way to reduce this error is to

Table 21

Virotrol RPR Panel Control Lot-to-Lot Variation Assessment

Virotrol RPR Panel Member 03 Use Dates	Virotrol lot #	# parallel testing measures	Mean	Standard Deviation (SD)	Range (Mean +/- 2 SD)	RPR titer range equivalent	Acceptable Yes/No
12-20-11 to 5-29-12	Virotrol 1	30	7.4	1.9	3.6 to 11.2	1:4 to 1:8	Yes
5-30-12 to 1-11-13	Virotrol 2	30	7.2	1.6	4.0 to 10.4	1:4 to 1:8	Yes
1-14-13 to 1-7-14	Virotrol 3	30	7.8	0.8	6.1 to 9.5	1:4 to 1:8	Yes

provide training and clearly defined nonoverlapping measurement categories. The three manual syphilis tests used in this study (RPR, RPR titer, and TP-PA) could only be performed if laboratory staff were trained appropriately and each person passed annual competency to ensure continued acceptable test performance. Additionally, the final test result for all three measures was clearly defined in the manufacturer's package insert with nonoverlapping categories.

To further enhance test result reproducibility, in 2011, the public health laboratory instituted dual observation of each step of the manual syphilis test process. For each RPR, RPR titer, and TP-PA test performed, a second technologist reviewed each step of the manual test process to ensure that it was completed properly before proceeding to the next step. The second technologist visually confirmed the manual RPR, RPR titer, and TP-PA final test results before reporting. There must be a consensus (exact agreement) by both technologists before the final test result was reported. If the technologists did not agree on the final results, the test was repeated until there was agreement. All steps were documented on the RPR (Figure 13), RPR titer (Figure 14) or TP-PA (Figure 11) Day of Use Worksheet.

Polit and Beck (2012) state that another assessment of equivalence is interrater reliability, which is the level of agreement between two independent observations for the same event. During the two-year study period, the laboratory maintained a continuous spreadsheet as a

Public Health Laboratory **Qualitative RPR (non Syphilis Monitoring Panel) Day of Use Worksheet**

1 Result: N R	**2 Result: N R**	**3 Result: N R**	**4 Result: N R**	**5 Result: N R**
6 Result: N R	**7 Result: N R**	**8 Result: N R**	**9 Result: N R**	**10 Result: N R**

RPR Tasks: Only 10 samples are to be tested per run. ALL TUBES ARE TO REMAIN CAPPED WHILE SITTING IN TESTING RACKS
Only 1 tube is to be opened at a time.
IF AT ANY TIME during testing there are questions regarding placement of the sample or possible sample contamination, the results must be discarded and the testing repeated or the sample redrawn.

Performed by: (**initialing** each line indicates that task was performed)

1. Worksheet labels checked against tube names _____
2. Tube placed in second row after dispensing on card _____
3. Tube capped after dispensing _____
4. Post testing - card read and results circled _____
5. Results posted in LIMS _____
6. Page number written at bottom of page _____
7. Reactive samples set aside for quant RPR or TPPA _____

Date:_____ Time:_____

Reviewed by: (**initialing** each line indicates that task was performed:

1. QC acceptable including room temp during run _____
2. Test card results checked against worksheet results _____
3. Worksheet labels checked against tube names _____
4. Check tubes are placed in second row and recapped _____
5. Discard test card _____
6. Results verified in LIMS _____
7. No problems identified with run _____
   Problems identified with run – list corrective action taken _____

Date: _____ Time: _____

Qualitative RPR Result Legend
N = nonreactive
R = REACTIVE

Room temp _____°C (23 – 29 °C) **Note with each run**
This worksheet is used after documentation of Quality Control activities on the RPR Day of Use/QA Worksheet.

Figure 13. Western Region Public Health Laboratory Qualitative RPR Day of Use Worksheet.
P. A. Armour, Qualitative RPR Day of Use worksheet, personal communication, 7-15-2011

Public Health Laboratory **Quantitative RPR (STAT RPR, RPR titer and Syphilis Monitoring Panel) Day of Use Worksheet**

Place sample label in this box	**Quantitative RPR Results legend:** N = non reactive R = reactive Rm = reactive minimally Circle last reactive dilution result and enter this results in LIMS			Reactive results at 1:16 require further testing with 1:50 dilution of non reactive serum.
1 Result: ____Titer: 1:1	**2 Result: ____Titer: 1:2**	**3 Result: ____Titer: 1:4**	**4 Result: ____Titer: 1:8**	**5 Result: ____Titer:1:16**
Place sample label in this box	**Quantitative RPR Results legend:** N = non reactive R = reactive Rm = reactive minimally Circle last reactive dilution result and enter this results in LIMS			Reactive results at 1:16 require further testing with 1:50 dilution of non reactive serum.
6 Result: ____Titer: 1:1	**7 Result: ____Titer: 1:2**	**8 Result: ____Titer: 1:4**	**9 Result: ____Titer: 1:8**	**10 Result: ____Titer: 1:16**
Place sample label in this box	**Quantitative RPR Results legend:** N = non reactive R = reactive Rm = reactive minimally Circle last reactive dilution result and enter this results in LIMS			Reactive results at 1:16 require further testing with 1:50 dilution of non reactive serum.
11 Result: ____Titer:1:1	**12 Result: ____Titer: 1:2**	**13 Result: ____Titer: 1:4**	**14 Result: ____Titer: 1:8**	**15 Result: ____Titer: 1:16**

Room temp _____°C (23 – 29 °C) **Note with each run. Worksheet is used after documentation of QC activities on the RPR Day of Use/QA Worksheet.**

Quantitative RPR Tasks: Only 3 samples are to be tested in a run. ALL TUBES ARE TO REMAIN CAPPED WHILE SITTING IN TESTING RACKS
IF AT ANY TIME during testing there are questions regarding placement of the sample or possible sample contamination, the results must be discarded and the testing repeated or the sample redrawn.

Performed by: (**initialing** each line indicates that task was performed)

1. Worksheet labels checked against tube names _____
2. Tube placed in second row after dispensing on card _____
3. Tube capped after dispensing _____
4. Post testing - card read and results circled _____
5. Results posted in LIMS _____
6. Page number written at bottom of page _____
7. Reactive samples set aside for TPPA _____
   (Note: No TPPA performed for Syph monitoring panel)

Date:_____ Time:_____

Reviewed by: (**initialing** each line indicates that task was performed:

1. QC acceptable including room temp during run _____
2. Test card results checked against worksheet results _____
3. Worksheet labels checked against tube names _____
4. Check tubes are placed in second row and recapped _____
5. Discard test card _____
6. Results verified in LIMS _____
7. No problems identified with run _____
   Problems identified with run – list corrective action taken _____

Date: _____ Time: _____

Page # _____

Figure 14. Western Region Public Health Laboratory Quantitative RPR Day of Use Worksheet.
P. A. Armour, Quantitative RPR Day of Use worksheet, personal communication, 7-15-2011

component of an expanded QA project with the public health clinic. Multiple laboratory staff entered data onto the spreadsheet that was queried from the LIMS (analysis date, unique medical record number, unique laboratory accession number, gender, date of birth, and laboratory test results including Syphilis IgG S/CO value) for those samples tested with the three different syphilis tests. Additionally, the laboratory staff and the student researcher were trained to access the public health clinic electronic patient chart and add the patient diagnosis and reason for clinic visit to the spreadsheet after testing was completed. The laboratory spreadsheet was completed by multiple laboratory staff and/or the student researcher in real time during the study period. The patient's chart was completed by multiple clinic staff in real time during the study period with data entered directly into the computer by clinic staff. The laboratory test results were entered manually into the patient's chart by clinic staff; however, the clinic used a two-person verification to ensure the data was entered correctly into the electronic medical record.

Because syphilis diagnosis was the gold standard for statistical analysis of the study hypotheses and the syphilis test results are hypotheses variables, the reliability of the data on the laboratory spreadsheet was assessed by the student researcher prior to deidentification of the retrospective study data. The student researcher determined that there were 4,077 patients who had all three syphilis tests performed that met the inclusion criteria during the retrospective study time period. The researcher received training from the public health clinic and IT staff for accessing data from the patient electronic chart. The clinic used standardized coding for electronic charting and the same coding was used in the standardized laboratory spreadsheet. Based on the medical record number in the laboratory spreadsheet, the student researcher accessed each patient's electronic chart (4,077 charts) and verified that the study variable data in the chart matched the study variable data in the spreadsheet. During the data assessment, if the

medical record number on the spreadsheet did not match the number in the electronic chart or if the test results for a medical record did not match the spreadsheet results, the student researcher used the LIMS to correctly identify the patient who had all three syphilis tests performed and verified the laboratory test results. All instances of nonmatching samples (191) shown in Table 22 were corrected on the laboratory spreadsheet by the student researcher and a nonmatching record log was maintained during the chart review.

Table 22

Nonmatching Study Variables Identified During Electronic Chart Review

Variable category on lab spreadsheet	**Number nonmatching samples on spreadsheet**
Date of Birth	2
Gender	3
Medical record number	32
Syphilis stage	104
Syphilis lab test result	50
Total nonmatching samples	**191**

The highest instance of disagreement (104) was incorrect or missing syphilis stage followed by incorrect syphilis lab results (50) entered on the laboratory spreadsheet. Interrater reliability of the two independent observations (laboratory spreadsheet and electronic chart) for agreement or disagreement of patient variable documentation can be determined by calculating the proportion of agreement using the number of agreements divided by the sum of the number of agreements plus disagreements. Based on the data in Table 22, the proportion of agreement for the chart review was 0.95 (3886/4077), with 1.0 being perfect agreement. Another measure of interrater reliability is Cohen's kappa (κ) which eliminates chance agreement. Statistical Package for the Social Science (SPSS) statistical software version 24 was utilized to calculate $\kappa=1.0$ ($p<0.001$), which is either a very good or excellent strength of agreement depending on

the reference text thus, enhancing the reliability of the spreadsheet data (Kaji, Schriger and Green, 2014; Polit and Beck, 2012; Xie, 2017).

Measurement validity. Polit and Beck (2012) describe measurement validity as how well the instrument measures the target construct. For this study, measurement face validity was enhanced by FDA approval of all three syphilis tests as well as the CDC recommendations for the syphilis algorithm serologic test interpretations for determining syphilis infection. Content validity refers to how well the measurement covers all aspects of the construct domain (Karras, 1997; Messick, 1997). Syphilis infection is a progressive disease that occurs in stages. Multiple prior studies shown in Table 4 demonstrate that all three syphilis tests (RPR, TP-PA, and EIA) will detect syphilis infection in different stages, although nontreponemal tests are not as sensitive as treponemal tests for detection of syphilis infection at later stages. The Captia™ Syphilis IgG package insert (Trinity Biotech, 2003) shows a sensitivity of 98.2% for the Syphilis IgG EIA tested against a known panel of serum from patients with untreated syphilis infection in all stages of illness. The nontreponemal VDRL tested against the same panel had a sensitivity of 80.2% with the lowest level of reaction with serum from patients diagnosed with late latent syphilis. The Serodia TP-PA package insert shows a sensitivity of 90% within a normal population and 100% within a population of suspected positive syphilis patients (Fujirebio Inc., 2006). Criterion related validity refers to how well the measurement scores against an external reference sample known to contain the construct analyte (Karras, 1997). The criterion validity of all three syphilis tests used in the study as shown in Tables 16-18 depicts acceptable (100%) qualitative QC results (syphilis reactive and nonreactive) with a known syphilis control across multiple lot numbers for all three tests during the study time period.

According to Messick (1995), "both convergent and discriminant evidence are basic to construct validation" (p. 746). Convergence refers to evidence that other test methods will provide similar test results for the construct. For this study, all three syphilis test methods were performed independently and had different manufacturers; yet each measurement approach converged on the construct, which was detection of syphilis infection. Tables 16-18 show that all three test methods used at the western region laboratory could to detect syphilis antibody in different known control samples. This enhanced convergent validity of the syphilis measurements. Discriminability refers to how distinctly the measurement identifies the construct when mixed with similar constructs. Syphilis is an STD that can occur concurrently with other STDs. Current syphilis test methods rely on detection of antibodies developed in response to the infection. These antibodies could be specific for the syphilis infection or a nonspecific inflammatory response. The manufacturer's package insert for each test method was reviewed to determine the level of cross-reactivity that occurs with each method, which can affect the level of discriminability.

Captia™ Syphilis IgG package insert (Trinity Biotech, 2003) lists results of testing a panel of 1,690 specimens containing samples from patients with no known history or serological evidence of syphilis infection. The panel included 1,002 patients with no known disease or condition and 688 patients with known non-syphilis disease, but whose sera was reactive for other disease states and/or had characteristics known to cause false reactive test results with other syphilis serologic methods. The 688 known disease patients included those infected with HIV, Hepatitis B and C, Human T-lymphotropic virus (HTLV), genital herpes, lyme disease, and leptospirosis. Also included were samples from patients with characteristics known to cause false reactive syphilis test results. These samples included patients with rheumatoid factor antibody, myeloma,

hypergammaglobulinemia, autoimmune disease, systemic lupus erythematosus (SLE), celiac disease, colitis, gout, sera from intravenous drug uses, and known false positive nontreponemal tests. All 1,002 normal nondisease sera were nonreactive with the Syphilis IgG test method. Eight samples from the 688 known disease sera were false reactive for a specificity of 99% (1682/1690*100) using the vendor's cross-reactivity panel, thus showing excellent discriminability for syphilis infection.

The Fujirebio, Inc. (2006) Serodia TP-PA package insert lists a cross-reactivity panel of 174 sera that had tested positive for different diseases or conditions including SLE, arthritis, HIV, lyme disease, *Helicobacter pylori*, *Toxoplasma*, and drug addiction. There were 24 false reactive TP-PA tests within the panel for a specificity of 86% (150/174*100). The HIV reactive sera accounted for 19 of the 24 false reactive results. While a specificity of 86% with this panel provides good discriminability for syphilis, the TP-PA test may exhibit a limitation when testing a population with a high HIV infection rate.

The MacroVue RPR package insert (Becton Dickinson and Company, 2010) did not provide results of a non-syphilis disease panel; however, it did list the following disease conditions as potentially causing biological false positive RPR test results: infectious mononucleosis, leprosy, malaria, SLE, connective tissue disorders, rheumatoid arthritis, pregnancy, drug addiction, and autoimmune disease. Prior studies shown in Table 4 provide a specificity range of 93-99% for the RPR test when tested in the general population.

Reliability and validity summary. The reliability of the three syphilis tests was enhanced by the data shown in Tables 16-18 with repeated measures providing consistent reactive and nonreactive qualitative results across multiple reagent lot numbers using known QC material. A new lot-to-lot variation assessment tool developed to utilize Syphilis IgG EIA quantitative test

results, identified a potential performance issue with one reagent lot number IgG H which was also identified by the vendor during the study and all boxes of the lot number were returned. According to the manufacturer, all qualitative QC and valid assay plate run patient measurements were acceptable; therefore, the samples tested with lot number IgG H were included in the study data. The error rate for the EVOLIS automated immunoassay analyzer (3.6%) was determined to be lower than the expected random error rate (5-6%) for diagnostic equipment; therefore, the analyzer provided consistent day to day results. The reliability of the laboratory data spreadsheet was enhanced by an extensive review of each patient's chart to determine the strength of agreement (very good to excellent) for the study data variables.

The validity of the three syphilis tests was supported by FDA clearance of each method for detection of syphilis infection in serum samples. The external consistency of all three syphilis tests was shown in Tables 16-18 with all three tests providing expected qualitative test results for an external reference material known to contain syphilis antibodies. A review of manufacturer's package inserts identified that all three tests were capable of detecting syphilis infection at multiple disease stages and discriminate syphilis infection from other diseases or conditions. The CDC recognizes the cross-reactivity (specificity) limitation of syphilis serological testing as well as the sensitivity limitation for various syphilis stages and, therefore, recommends the use of more than one syphilis test to determine syphilis infection as seen in the traditional and reverse algorithms in Figures 4 and 5.

Population and Sample Size

For STD surveillance purposes, the CDC utilizes four US census regions, West, Midwest, South and Northeast, for reporting STD infections. The West region includes the states of Alaska, Arizona, California, Colorado, Hawaii, Idaho, Montana, Nevada, New Mexico, Oregon,

Utah, Washington, and Wyoming. According to the CDC 2013 Syphilis statistics, the rate of P&S syphilis cases in the West region continued to increase during the study time period. The West region P&S rate was 4.9 cases per 100,000 population in 2011 and 5.8 cases per 100,000 population in 2012 which was an 18.4% increase in one year.

The study setting was an STD public health clinic located in a large western region metropolitan area comprising over two million residents. The STD public health clinic was operated by the public health agency within the western region metropolitan area and provided the following services:

- diagnosis and treatment of active or suspected cases of chlamydia, gonorrhea, syphilis, HIV, trichomoniasis, yeast infections, and bacterial vaginosis;
- high risk behavior counseling;
- HIV nursing case management;
- partner notification for STD infection; and
- sexual assault follow-up.

The STD clinic received federal funding through the CDC STD prevention cooperative grant, which sought to reduce syphilis infection especially among high-risk populations, such as MSM. The clinic staff followed federal guidelines for syphilis detection and treatment. In addition to patients who had signs and symptoms of syphilis infection, the STD clinic also targeted pregnant women, MSM, persons with HIV, and contacts of known syphilis cases for conducting syphilis testing. Because syphilis infection can occur concurrently with other STDs, clinic patients were often screened for multiple STDs (HIV, syphilis, gonorrhea, and chlamydia) during the same clinic visit. All syphilis testing ordered by the STD clinic staff was performed at the public health agency public health laboratory. The public health laboratory also performed syphilis

testing for various other public health agency departments, such as the family planning clinic and the HIV/AIDS clinic. The HIV/AIDS clinic provided outreach testing to high risk populations in jails, juvenile detention centers, and Lesbian, Gay, Bisexual and Transgender (LGBT) community centers. The test results from these outreach locations were provided by the clinic to the patient. The patient made the decision to either visit the STD clinic for treatment or follow up with their personal physician; therefore, not all patients tested at the public health laboratory were seen in the STD clinic.

The target population included all patients who sought medical evaluation at the STD clinic from January 1, 2012 through December 31, 2013. Purposive sampling was used to select sample members to include in the study. The STD clinic staff determined those patients who required syphilis testing based on clinic criteria. The selection criteria for inclusion in the study are males and females 18 years and older who were seen in the public health STD clinic, met the STD clinic criteria for syphilis testing and had all three syphilis tests (Syphilis IgG, RPR, and TP-PA) performed with clinically diagnostic test results on each sample collected. Exclusion criteria included males and females under 18 years of age, those who did not visit the STD clinic, those that did not have three syphilis tests performed, samples that tested invalid by the Syphilis IgG test, and samples that tested inconclusive by the TP-PA test method.

Males and females under 18 years of age are minors and were excluded because they represent a vulnerable population. A chart review was required to obtain a clinical diagnosis which was the gold standard for the statistical evaluation of the study variables. Patients who did not visit the STD clinic did not have a medical record; therefore, a diagnosis could not be obtained, which excluded them from the study. Invalid and inconclusive test results cannot be interpreted as either reactive or nonreactive; therefore, the samples did not meet the criteria for

providing clinically diagnostic test results. The laboratory and the STD clinic performed syphilis screening testing using a modification of the CDC reverse algorithm, which started with an initial Syphilis IgG test and reflexed to concurrent performance of RPR, RPR titer, and TP-PA if the initial screening test was reactive or equivocal. The laboratory also performed reflex testing to RPR and TP-PA for those samples that had a nonreactive Syphilis IgG test with a S/CO value equal to or greater than 0.450. The public health agency staff also had the ability to order a panel of Syphilis IgG, RPR, and TP-PA tests if the patient was in a high risk group or staff determined all three tests should be ordered based on clinical assessment. Other public health departments, such as the HIV/AIDS clinic, could also order all three syphilis tests. The public health laboratory also received samples with requests for one or two syphilis tests to be performed as in the case of syphilis monitoring testing, which required performance of RPR and RPR titer. Only samples that had all three syphilis tests performed on the same sample and had a chart established based on a clinic visit were included in the study as clinical correlation of the test results was necessary to determine a syphilis diagnosis based on CDC recommendations. During the two-year study period (2012-2013), the western region public health laboratory performed 34,340 Syphilis IgG tests. As a component of an ongoing QA project with the public health clinic, the laboratory maintained a spreadsheet that included demographic information and test results for patients who were tested with all three syphilis test methods. For the study time period, there were 5,061 patients on the laboratory spreadsheet. These 5,061 patients included those that tested initially reactive or equivocal with the Syphilis IgG test and were reflexed to RPR, RPR titer, and TP-PA testing as well as patients who were tested with all three syphilis test (Syphilis IgG, RPR, and TP-PA) as ordered by public health agency staff. Additionally, nonreactive Syphilis IgG samples with S/CO values greater than or equal to 0.450 were reflexed

by the laboratory to RPR and TP-PA testing. Therefore, the list of 5,061 patients included those that were nonreactive for all three syphilis tests as well as those that were reactive for one or more of the tests. Based on the study exclusion criteria, 216 patients were under the age of 18 years and were removed from the study. There were 751 patients who were not seen in the public health clinic. The patients not seen represented those patients tested through the HIV/AIDS clinic outreach centers; jails or juvenile detention centers who did not go to the public health clinic for reactive test treatment; those whose test results were nonreactive and, thus, did not require treatment; or those that had reactive test results with a history of syphilis infection, which did not require clinical follow up. These patients did not have a clinic chart and were excluded from the study as diagnoses were not available. There were six samples that were resulted as TP-PA inconclusive, and 11 samples resulted as Syphilis IgG invalid. An inconclusive or invalid test result could be due to instrument, reagent, or operator error. These samples were excluded from the study as they could not be interpreted as either reactive or nonreactive, thus they did not provide clinically diagnostic test results. Additional discussion regarding the invalid and inconclusive results is in the variables section. The final number of patients included in the study was 4,077 with 3,218 patients with a diagnosis of no syphilis and 859 with a diagnosis of syphilis in various stages (Table 23). The public health clinic ordered testing based on the reverse algorithm which started with a treponemal EIA for initial screening. The total number of samples initially tested for Syphilis IgG during the two-year study period was 34,340; therefore, the syphilis prevalence rate for this public health agency population would be 3% (859/34340*100).

Table 23

Study Exclusion Category, Number and Rationale

Exclusion category	Excluded number	Exclusion rationale
Males and females under age 18 years	216	Minor children; Vulnerable population
Patients not seen in public health clinic	751	Represents samples collected at outreach sites; No medical chart and therefore cannot obtain clinical diagnosis
Inconclusive TP-PA test results	6	Results cannot be interpreted as either reactive or nonreactive, and therefore cannot be used to clinically interpret syphilis algorithms
Invalid Captia™ Syphilis IgG test results	11	Results cannot be interpreted as either reactive or nonreactive, and therefore cannot be used to clinically interpret syphilis algorithms
Total excluded	984	

Statistical Power Analysis

There are two types of error that affect a researcher's ability to conclude that a hypothesis is probably true or probably false. A Type I error rejects a true null hypothesis and a Type II error accepts a false null hypothesis. Type I error can be controlled by selecting a level of significance or alpha (α) which represents the probability of finding statistical significance when there is none (false positive). Two commonly used α levels are 0.01 and 0.05 which represent the amount of acceptable risk of rejecting the true null hypothesis. At α of 0.01, the researcher accepts the risk that 1 sample out of 100 will reject the true null hypothesis. At α of 0.05, the researcher accepts the risk of rejecting the true null hypothesis for 5 samples out of 100. An α of 0.05 is commonly used in research, as lowering the level to 0.01 also increases the risk of Type II error. A mechanism to reduce the risk of Type II error is to increase the sample size (Lipsey & Hurley, 2009; Polit & Beck, 2012; Sullivan & Feinn, 2012). Type II error or β represents the probability of not finding statistical significance when there is an effect (false negative). The

statistical power of a study or 1-β represents the probability that a statistically significant difference will be found when a difference actually exists. If the statistical power is high, then the likelihood of deciding there is an effect when one is present is also high. Cohen (1992) proposed that setting β at 0.20 would provide a statistical power of 0.80 (1-0.20) which would be high enough to reduce the risk of Type II error. Since Cohen's recommendation was published, researchers have conventionally set α at 0.05 and β at 0.20 thus providing a 4:1 ratio (0.20 to 0.05) for both Type I and Type II errors.

Multiple authors (Cohen, 1990; Ferguson, 2009; Lipsey & Hurley, 2009; Polit & Beck, 2012; Sullivan & Feinn, 2012) discuss the relationship between sample size and power. If the sample size is large enough, almost any level of power and significance (*p* value) can be detected, even if the difference in outcomes between groups is not necessarily important or meaningful. A statistically significant *p* value indicates that the study results are unlikely to have been caused by chance; however, it does not indicate the size of the effect or association between the variables. An effect size estimate is a statistical tool that is not affected by sample size and provides the researcher with an idea of the impact of a statistically significant result. According to Sullivan and Feinn (2012), the type of comparison performed within the study influences which indices are used to estimate the effect size. The indices can be divided into two main categories: effect sizes between groups (Cohen's *d*, odds ratio[OR], relative risk[RR]) or measures of association between variables (Pearson's *r* correlation, r^2 coefficient of determination). The OR is recommended for binary outcome variables and classifies effect sizes as small (1.5), medium (2) and large (3). A small effect size is less than medium but not so small as to be considered meaningless. Ferguson (2009) does not recommend translation of OR or RR

values into *r* or *d* when the data are binomial and further states that there is no agreement on how large an effect size must be in order to be considered practically significant.

Prior to starting a research project, it is important to determine the sample size necessary to provide statistically significant results that are meaningful. A statistical power analysis will determine the sample size that ensures the study has acceptable power to support the null hypothesis. According to Polit and Beck (2012), it is difficult to estimate sample sizes when the study will be testing differences in proportions between groups, such as the dichotomous variables which were analyzed in this research study. Various authors (Bujang & Adnan, 2016; Hajian-Tilaki, 2014; Malhotra & Indrayan, 2010; Obuchowski, 1998) suggest that the statistical indices of sensitivity and specificity should be used for determining sample size for diagnostic medical tests at a desired power. There are four values that are necessary for sample size determination using sensitivity or specificity: a) pre-determined sensitivity and/or specificity which can be obtained from previously published studies or clinical judgement, b) the significance criterion (α), c) the desired power level ($1-\beta$), and d) the precision of estimates of sensitivity and/or specificity which is the maximum difference between the estimated sensitivity and/or specificity and the true value. Additionally, the disease prevalence within the population should also be considered to ensure sufficient numbers of patients with the disease are included in the study. In general, a larger sample size will be needed to obtain a higher sensitivity in a lower prevalence population. Low syphilis prevalence of one percent or less is typically seen in the general population with high syphilis prevalence of greater than 10 percent often seen in MSM populations (Snowden, et al., 2010). According to the Captia™ Syphilis IgG package insert (Trinity Biotech, 2003) a general clinical laboratory testing referred and routine specimens may have a 4.5% reactivity rate.

The syphilis tests performed in this study are used for screening the public health population for syphilis infection which requires tests of high sensitivity to detect as many true-positives as possible. However, the RPR and TP-PA tests are also used as confirmatory diagnostic tests and require sufficiently high specificity to detect true-negatives. The study variables include three different syphilis tests and interpretations of two different algorithms, which requires use of a sufficiently high degree of both sensitivity and specificity values that will include all variables. According to the syphilis test ranges listed in Table 4, various published studies provided a syphilis EIA sensitivity range of 57%-100% and specificity range of 80%-100%; an RPR sensitivity range of 37%-100% (depending on disease stage) and specificity range of 93%-98%; and a TP-PA sensitivity range of 69%-100% and specificity range of 95%-100%. There was only one published study by Tong, et al (2013) that determined the sensitivity of the traditional algorithm at 75.8% and the reverse algorithm at 99.8% based on presence or absence of disease. The manufacturer's package inserts list a sensitivity of 98.2% and specificity of 99% for the Captia^{TM} Syphilis IgG EIA assay along with a nontreponemal (VDRL) sensitivity of 80.2% (Trinity Biotech, 2003) and a sensitivity of 90% with specificity of 86% for the TP-PA (Fujirebio Diagnostics, Inc., 2006). Bujang and Adnan (2016) recommend use of a predetermined sensitivity and specificity of at least 70% within the null hypothesis and at least 80% within the alternate hypothesis in order to indicate that the test is a fairly good diagnostic tool. Based on this recommendation, the values in Table 4 and the test manufacturer's package insert, an estimated sensitivity and specificity of 80%, was used in the study power analysis. This sensitivity and specificity represents the lower end of practical importance as seen in the ranges listed in Table 4 and was a rough estimate of the minimum acceptable value for the study with the expectation that the tests will actually perform at 90% sensitivity and specificity. This

represents a precision of 0.10 which is the maximum difference between estimated sensitivity and specificity and the true value.

A statistical power analysis based on a binomial distribution was performed using Power Analysis and Sample Size (PASS) 15 software (PASS 15, 2017). According to the PASS 15 (2017) manual, a binomial variable should exhibit the following properties a) the variable is binary (can take on one of two possible values), b) the variable is observed a known number of times, c) the probability that the outcome of interest occurs is constant for each trial, and d) the trials are independent (the outcome of one trial does not influence the outcome of the other trial). Because the retrospective study sample size (4,077) and population disease prevalence (3%) was known, a two-sided test for one-sample sensitivity and specificity was designed in PASS 15 to solve for power with the following values: alpha = 0.05, sample size = 4,077, disease prevalence = 0.03, sensitivity and specificity = 0.80, and precision = 0.10. The PASS 15 software calculated that a total sample size of 4,077 (which includes at least 122 subjects with the disease) achieves 84% power to detect a change in sensitivity from 0.8 to 0.9 using a two-sided binomial test and 100% power to detect a change in specificity from 0.8 to 0.9 using a two-sided binomial test. The target significance level was 0.05. The software also calculated that the total sample size of 4,077 achieves 100% power to detect a change in sensitivity and specificity from 0.8 to 0.99 using a two-sided binomial test with target significance level at 0.05. These calculations indicate that the sample size was sufficiently large within a population with 3% prevalence to detect small (0.1) or medium (0.2) differences in sensitivity and specificity with greater than 80% power. The software calculated a 20% dropout rate to determine the number of subjects that are expected to be lost at random during the course of the study and for whom no response data will be collected. Based on the sample size, the expected number of dropouts was 1,020;

therefore, a dropout inflated enrollment sample size of 5,097 would be recommended for the study. The actual dropout-inflated enrollment sample size was 5,061 and the number of dropouts (excluded samples) was 984.

Cohen (1990), Ferguson (2012), and Sullivan and Feinn (2012) recommended use of OR to determine effect size for medical outcomes research. Using PASS 15 software, a test for two correlated proportions (McNemar test) was designed. Because the sample size was known, the test was designed to solve for OR as follows: sample size = 4,077, alpha = 0.05, power at various levels of 0.8, 0.9, 0.95 and 0.99, and proportion discordant = 0.1. The PASS 15 software calculated that a sample size of 4,077 achieved 99% power to detect an odds ratio of 1.5 using a two-sided McNemar test with a significance level of 0.05. An odds ratio of 1.5 would be considered a small effect, which according to Sullivan and Feinn (2012) is not so small as to be considered meaningless.

Variables

The public health laboratory staff gathered the study variables from January 1, 2012 through December 31, 2013 as a component of the enhanced syphilis surveillance performed by the public health agency. Laboratory staff entered the data on an Excel spreadsheet; which was stored in a secured, protected location on the public health agency firewall protected intranet. The study variables included both descriptive and observed variables. Descriptive variables included gender (operationalized as male, female and transgender), age (operationalized as years), reason for clinic visit (operationalized as STD screen with symptoms, STD screen without symptoms, or notification which included contact with an STD or receipt of a letter notifying the patient that a partner had an STD), infection diagnosis (operationalized as no STD, syphilis past or current, HIV, chlamydia and/or gonorrhea, or other), and syphilis stage

(operationalized as primary, secondary, latent, tertiary, not staged, and history). The contact STD category included contact with HIV, trichomonas, herpes, gonorrhea, chlamydia, syphilis, or nongonococcal urethritis (NGU). The other infection category included diagnosis of herpes, warts, molluscum contagiosum, condyloma, human papillomavirus (HPV), yeast, NGU, bacterial vaginosis, trichomoniasis, lupus, gout, or autoimmune disease. The age was calculated from the date of birth and the sample collection date using a formula created in Microsoft Excel by laboratory staff. Once the age was calculated, date of birth was removed from the data set and was not included in the study variables. Observed variables included the Captia™ Syphilis IgG test result and S/CO value, Becton Dickinson MacroVue RPR test result and titer value, Fujirebio Serodia TP-PA test result, syphilis traditional and reverse algorithm serologic interpretation, contingency table classification, and clinical diagnosis of syphilis based on chart review (Table 24).

The Captia™ Syphilis IgG test result was calculated by the BioRad EVOLIS automated analyzer based on the S/CO value and reported as reactive, equivocal, nonreactive, or invalid. According to the manufacturer's package insert, initially reactive or equivocal results should be retested in duplicate. Specimens which gave repeat equivocal results should be considered reactive until confirmatory testing was completed. Additionally, the package insert stated that equivocal test results were scored as reactive during performance validation studies (Trinity Biotech, 2003). Standard practice at the public health laboratory was to confirm both equivocal and reactive results with follow-up RPR, RPR titer, and TP-PA testing; thus, an equivocal result was treated as a reactive. As shown in Table 25, there were 54 equivocal (EQ) results in the study with 61% (33/54) having a diagnosis of no syphilis.

Table 24

Study Variables, Definition, Type, and Operationalized Outcome

Variable	**Variable definition**	**Variable type**	**Operationalized outcome**
Gender	Patient gender	Categorical	male, female, transgender
Age	Patient age at time of sample collection	Continuous	years
Reason for visit	Patient reason for visit to STD clinic	Categorical	STD screen with symptoms, STD screen without symptoms, notification
Infection diagnosis	STD infection reported by clinic staff in chart	Categorical	Syphilis past or current, HIV, gonorrhea and/or chlamydia, no STD or other
Syphilis IgG result	FDA approved clinically diagnostic test result	Dichotomous	nonreactive, reactive Note: equivocal calculated separately as reactive and nonreactive
Syphilis IgG S/CO value	Numerical value determined by EVOLIS analyzer	Continuous	decimal number from 0-5.50
TP-PA result	FDA approved clinically diagnostic test result	Dichotomous	nonreactive, reactive
RPR result	FDA approved clinically diagnostic test result	Dichotomous	nonreactive, reactive
RPR titer result	FDA approved reported test result	Continuous	0 for nonreactive, 1 for 1:1, 2 for 1:2, 4 for 1:4, 8 for 1:8, 16 for 1:16, 32 for 1:32, 64 for 1:64, 128 for 1:128, 256 for 1:256, 512 for 1:512, 1024 for 1:1024
Algorithm serologic interpretation	Lab interpretation based on traditional or reverse sequence algorithm and clinical correlation for discordants	Dichotomous	syphilis infection unlikely, syphilis infection likely
Algorithm contingency table	Comparison of traditional and reverse algorithm interpretation and clinical diagnosis	Categorical	true positive, true negative, false positive, false negative

Table 24. *Continued*

Variable	Variable definition	Variable Type	Operationalized outcome
Laboratory test contingency table	Comparison of Syphilis IgG and TP-PA test result and clinical diagnosis	Categorical	true positive, true negative, false positive, false negative
Clinical diagnosis of syphilis	Diagnosis determined by clinic staff obtained from patient chart	Dichotomous Gold standard	syphilis infection no, syphilis infection (past or current) yes
Syphilis stages	Stage determined by clinic staff and obtained from patient chart	Categorical	primary, secondary, latent (includes early, late and unknown duration), tertiary, not staged, history

Table 25

Equivocal Syphilis IgG Test Result Frequency and Diagnosis/Stage Frequency

Result	Result frequency	Diagnosis/stage frequency
$IgG^{EQ}/RPR^{-}/TP\text{-}PA^{-}$	30	27 No syphilis; 1 Late latent; 1 not staged; 1 history syphilis
$IgG^{EQ}/RPR^{-}/TP\text{-}PA^{+}$	12	2 No syphilis; 1 Primary; 7 Late latent; 2 history syphilis
$IgG^{EQ}/RPR^{+}/TP\text{-}PA^{+}$	7	1 Primary; 2 Early; 4 Late latent
$IgG^{EQ}/RPR^{+}/TP\text{-}PA^{-}$	5	4 No syphilis; 1 Secondary
Total	**54**	

According to a 2007 FDA guidance document for reporting results from studies evaluating diagnostic tests, equivocal results should not be discarded or ignored when performing statistical evaluation as it may introduce bias into the calculations. To address the potential bias issue, the FDA suggests reporting two different sets of performance measures: one set based on including the equivocal results with the test reactive results and one set based on including the equivocal results with the test nonreactive results (FDA, 2007). For the purposes of this study, the equivocal results were included and separate statistical calculations were performed based on this FDA recommendation.

The invalid test results were removed from the study as they cannot be interpreted as either reactive or nonreactive. Therefore, they provide uninterpretable diagnostic information which cannot be used within the reverse algorithm interpretation. At the western region laboratory, the laboratory director approved reflex testing to both RPR, RPR titer, and TP-PA for all invalid results so that patient diagnosis was not delayed. Additional recommendations were to recollect the sample if the RPR and TP-PA results were not consistent with the patient's clinical picture. There were 11 Syphilis IgG test results that were reported as invalid out of 4,077 study samples representing 0.3% (11/4077) of the tested samples. According to Begg (1978), uninterpretable diagnostic test results will not bias statistical analysis provided that the test is repeatable and the cause of the uninterpretable result is random. The Syphilis IgG test could be repeated with a new sample. To determine if the invalid results were random, they were treated as missing data. Tabachnick and Fidell (2013) state that if less than 5% of the missing data points represent a random subsample of the whole sample and are concentrated in a few variables which are highly correlated with other complete variables, then the cases can be deleted without compromising statistical analysis. The 11 invalid Syphilis IgG test results represented 0.3% of the study cases (Table 26). A review of the 11 invalid results revealed that the test analysis dates had a random pattern with only one invalid test result per analysis date. Additionally, the invalid results were highly correlated to the other syphilis test variables. There were three instances in March 2013 during a time period when there was a known instrument issue. There was also one instance in June 2013 and two instances in July 2013 when there were known instrument and reagent issues (Tables 19 and 20). Known instrument and reagent issues also provide a rationale for removal of the invalid results. Additional clinical data shown in Table 26 reveal that three of the 11 patients were diagnosed with syphilis infection either past or present while the remaining eight patients

Table 26

Invalid Syphilis IgG Test Result Analysis Dates and Clinical Data

Study case number	Analysis date	Clinical data
40	04-26-2012	Nonreactive RPR and TP-PA; No syphilis; Yeast infection; No redraw
3578	08-23-2012	Nonreactive RPR and TP-PA; No syphilis; No redraw
4875	03-07-2013	Reactive RPR with titer 1:64 and reactive TP-PA; Latent syphilis infection unknown stage; contact with HIV and syphilis infections
4876	03-18-2013	Reactive RPR with titer 1:4 and reactive TP-PA; Late latent syphilis infection
4863	03-21-2013	Nonreactive RPR and TP-PA; Late latent syphilis infection; Gonorrhea and chlamydia infections; No redraw
4864	05-23-2013	Nonreactive RPR and TP-PA; No syphilis; Yeast infection; Redraw Syphilis IgG nonreactive
4867	06-24-2013	Nonreactive RPR and TP-PA; No syphilis; Chlamydia and nongonococcal urethritis infections; No redraw
4877	07-21-2013	Nonreactive RPR; Reactive TP-PA; History of syphilis infection 2012; Chlamydia infection; No redraw
4868	07-24-2013	Nonreactive RPR and TP-PA; No syphilis; Gonorrhea infection; No redraw
4873	10-22-2013	Reactive RPR with titer 1:1; Nonreactive TP-PA; No syphilis; Redraw Syphilis IgG nonreactive
4874	11-15-2013	Reactive RPR with titer 1:2; Nonreactive TP-PA; No syphilis; Redraw Syphilis IgG nonreactive

were diagnosed with no STD infection or with another STD infection (gonorrhea, chlamydia, or yeast). Because the invalid results represent a small random subsample (0.3%) of the large data set utilized in this study and are correlated to other complete variables, the test results are not critical to the analysis and can be deleted from the study. Therefore, the Captia™ Syphilis IgG test result was operationalized as a dichotomous variable (reactive or nonreactive).

The Captia™ Syphilis IgG S/CO was operationalized as a decimal number automatically calculated by the BioRad EVOLIS automated analyzer. The S/CO value was derived by expressing the absorbance of the test specimen as a ratio of the mean absorbance of the test kit's

Low Titre Reactive Control. A numerical ratio value between 0.9 and 1.1 was considered an equivocal test result. A ratio value greater than or equal to 1.1 was considered a reactive result. A ratio value less than or equal to 0.9 was considered a nonreactive result. The S/CO values for this study ranged from 0 to 5.5. According to the FDA approved manufacturer's instructions, the S/CO value was not reported and was used only for internal calculations to determine the Syphilis IgG test result. The S/CO value is a continuous variable because the value reflects the amount of treponemal antibody present in the patient's serum and can take on any value within the range of the instrument scale (Tabachnick & Fidell, 2013).

The Becton Dickinson MacroVue RPR was operationalized as reactive or nonreactive based on manual visual interpretation of the test result. The qualitative RPR result is a dichotomous variable. Following CDC recommendations, all reactive qualitative RPR tests were reflexed to a quantitative RPR titer. The RPR titer was performed by making serial two-fold dilutions of the initial reactive sample (undiluted) and repeating the RPR test until the titer dilution reached a nonreactive endpoint. The final test result was reported as the highest dilution value showing a reactive result. According to Tabachnick and Fidell (2013), the RPR titer result can be treated as a continuous variable because there are numerous result categories; and each category (titer result) is a numerical value that falls along a quantitative continuum. The RPR titer value was operationalized as a numerical value which doubled in value as it increased and was reported as 1:1(undiluted), 1:2, 1:4, 1:8, 1:16, 1:32, 1:64, 1:128, 1:256, 1:512, 1:1024, with 1:1024 representing the highest dilution in the data set.

The Fujirebio Serodia TP-PA was reported as nonreactive, reactive, inconclusive, or invalid. There were zero invalid TP-PA test results in the study data and six inconclusive TP-PA results out of 4,077 study samples representing 0.1% (6/4077) of the study data. Inconclusive results

cannot be interpreted as either reactive or nonreactive. Therefore they provide uninterpretable diagnostic information which cannot be used within either the reverse or traditional algorithm interpretations. Inconclusive results require collection of a new sample for repeat analysis. A review of the six inconclusive results (Table 27) revealed that the test analysis dates had a random pattern with only one inconclusive test result per analysis date.

Table 27

Inconclusive TP-PA Test Result Analysis Dates and Clinical Data

Study case number	Analysis date	Clinical data
1021	11-06-2012	Nonreactive Syphilis IgG and RPR; History of syphilis infection 2012; No redraw
4838	05-09-2013	Nonreactive Syphilis IgG and RPR; No syphilis; Redraw TP-PA nonreactive
4834	06-10-2013	Nonreactive Syphilis IgG; Reactive RPR with titer 1:1; Early latent syphilis infection; Contact with syphilis infection; No redraw
4841	07-09-2013	Nonreactive Syphilis IgG and RPR; No syphilis infection; Gonorrhea infection; Redraw TP-PA inconclusive
4849	10-02-2013	Reactive Syphilis IgG; Nonreactive RPR; History of syphilis infection unknown date; No redraw
4852	10-26-2013	Nonreactive Syphilis IgG and RPR; Primary syphilis infection; Contact early syphilis: Redraw Syphilis IgG equivocal, RPR reactive, TP-PA reactive

Additional clinical data revealed that four of the six patients were diagnosed with syphilis past (2) or present (2). Of interest is study case number 4852, who had a diagnosis of primary syphilis based on clinical assessment and symptoms with initial nonreactive Syphilis IgG and RPR with inconclusive TP-PA testing. Repeat testing two weeks after the initial test revealed equivocal Syphilis IgG, reactive RPR, and reactive TP-PA tests. This case shows the difficulty of diagnosing syphilis due to varying antibody levels, especially in the early stages of infection. Due to the randomness of the missing data and correlation with other syphilis variables, these six

inconclusive results are not critical to the analysis making up a small subsample (0.1%) of the large data set. Therefore, the results were deleted from the study, and the TP-PA test result was operationalized as a dichotomous variable (nonreactive or reactive).

The CDC and APHL provide suggested laboratory serologic interpretation language for both the traditional (Figure 4) and reverse sequence (Figure 5) syphilis algorithms (CDC, 2011; Morshed & Singh, 2015). Both algorithms require reflex testing for initially reactive RPR or Syphilis IgG tests. Reflex testing using other test methods of varying sensitivity and specificity may result in discordant or nonmatching results.

For the traditional algorithm, a nonreactive initial RPR test was interpreted as syphilis infection unlikely with no further testing performed unless the healthcare provider requests it. If the RPR was reactive, then a TP-PA test was performed. If the TP-PA test was reactive, the traditional algorithm was interpreted as syphilis infection likely. If the TP-PA test was nonreactive, then the results were discordant (nonmatching) and required clinical correlation to determine if the discordance was due to a biological false positive (often seen with RPR tests) or indication of syphilis infection (either past or present). The traditional algorithm laboratory serologic interpretation was operationalized as a dichotomous variable (syphilis infection likely or unlikely); however, for the discordant results, clinical correlation was necessary to complete the dichotomous algorithm interpretation (Figure 4). For this study, syphilis diagnosis was captured for each sample which allowed the student researcher to clinically correlate the discordant results to determine if the traditional algorithm interpretation was a true or false positive. Discordant RPR positive results with a diagnosis of syphilis were categorized as likely true positive and discordant RPR positive results with no diagnosis of syphilis were categorized as likely false positive. Because all samples in this study were tested with all three syphilis tests,

instances arose in which the RPR test was nonreactive and the TP-PA test was reactive. In those instances, only the nonreactive RPR test result was used in algorithm interpretation. The TP-PA results were ignored, since, if the traditional algorithm was followed, the TP-PA test would not have been performed if the initial RPR was nonreactive. The Syphilis IgG test results were also ignored because the test would not be performed within the traditional algorithm (Table 28).

Table 28

Traditional Algorithm Interpretation Contingency Table Measurement Outcomes

Traditional algorithm interpretation	**Disease (syphilis infection)**	
	Yes	**No**
Syphilis infection likely	True positive (TP) $RPR^{+}/TP\text{-}PA^{+}$ [a]$RPR^{+}/TP\text{-}PA^{-}$	False positive (FP) $RPR^{+}/TP\text{-}PA^{+}$ [a]$RPR^{+}/TP\text{-}PA^{-}$
Syphilis infection unlikely	False negative (FN) $RPR^{-}/TP\text{-}PA^{-}$ [b]$RPR^{-}/TP\text{-}PA^{+}$	True negative (TN) $RPR^{-}/TP\text{-}PA^{-}$ [b]$RPR^{-}/TP\text{-}PA^{+}$

[a]Discordant results: according to the traditional algorithm interpretation, a reactive RPR with nonreactive TP-PA could be indicative of either biological false positive or early syphilis infection and required clinical correlation to determine contingency table category.
[b]According to traditional algorithm, TP-PA test would not be performed if RPR is nonreactive; therefore, only RPR result used to determine contingency table category.

The reverse sequence algorithm laboratory serologic interpretations included syphilis infection likely, syphilis infection unlikely, or syphilis infection inconclusive (discordant results). The reverse algorithm interpretation was operationalized as a dichotomous variable (likely or unlikely) with the inconclusive (discordant) results resolved as either likely or unlikely using clinical correlation. As with the traditional algorithm, discordant results occur when a sample that tested initially Syphilis IgG reactive or equivocal is nonreactive when the RPR test was performed. Syphilis IgG equivocal results were counted separately as reactive and nonreactive, with two different sets of statistical calculations performed. According to the reverse sequence algorithm, a second treponemal test, TP-PA, was performed as the tie-breaker.

If the TP-PA test was nonreactive, the algorithm interpretation was inconclusive as this could indicate a false positive Syphilis IgG or detection of an early syphilis infection. Clinical correlation was required to determine the reverse algorithm interpretation (Figure 5). For those inconclusive algorithm interpretations, the student researcher clinically correlated the discordant results with the syphilis diagnosis to determine if the reverse algorithm interpretation was a true or false positive. Reactive Syphilis IgG, nonreactive RPR, and nonreactive TP-PA results with a diagnosis of syphilis were categorized as likely true positive while reactive Syphilis IgG, nonreactive RPR, and nonreactive TP-PA results with no diagnosis of syphilis were categorized as likely false positive (Table 29).

Table 29

Reverse Algorithm Interpretation Contingency Table Measurement Outcomes

	Disease (syphilis infection)	
Reverse algorithm interpretation	**Yes**	**No**
Syphilis infection likely	True positive (TP)	False positive (FP)
	$IgG^+/RPR^+/TP\text{-}PA^+$	$IgG^+/RPR^+/TP\text{-}PA^+$
	$IgG^+/RPR^+/TP\text{-}PA^-$	$IgG^+/RPR^+/TP\text{-}PA^-$
	$IgG^+/RPR^-/TP\text{-}PA^+$	$IgG^+/RPR^-/TP\text{-}PA^+$
	[a]$IgG^+/RPR^-/TP\text{-}PA^-$	[a]$IgG^+/RPR^-/TP\text{-}PA^-$
Syphilis infection unlikely	False negative (FN)	True negative (TN)
	$IgG^-/RPR^-/TP\text{-}PA^-$	$IgG^-/RPR^-/TP\text{-}PA^-$
	[b]$IgG^-/RPR^-/TP\text{-}PA^+$	[b]$IgG^-/RPR^-/TP\text{-}PA^+$
	[b]$IgG^-/RPR^+/TP\text{-}PA^+$	[b]$IgG^-/RPR^+/TP\text{-}PA^+$
	[b]$IgG^-/RPR^+/TP\text{-}PA^-$	[b]$IgG^-/RPR^+/TP\text{-}PA^-$

[a]Discordant results: according to the reverse algorithm interpretation, a reactive Syphilis IgG with nonreactive TP-PA and RPR could be indicative of either false positive or early syphilis infection and required clinical correlation to determine contingency table category.
[b]According to the reverse algorithm, RPR and TP-PA tests would not be performed if Syphilis IgG is nonreactive; therefore, only Syphilis IgG result used to determine contingency table category.
Note: Equivocal Syphilis IgG results are counted as two different sets of measures: one set with equivocal counted as reactive and one set with equivocal counted as nonreactive

A sample that initially tested Syphilis IgG nonreactive was interpreted as syphilis infection unlikely and no additional testing was performed unless ordered by the healthcare provider or

reflexed following laboratory protocol. Each of the study samples were tested with all three syphilis tests which could result in reactive RPR and/or TP-PA tests with a nonreactive Syphilis IgG test. In those instances, only the nonreactive Syphilis IgG test result was used for reverse algorithm interpretation with the RPR and TP-PA results ignored, because, according to the reverse algorithm, the RPR and TP-PA tests would not have been performed if the initial Syphilis IgG was nonreactive.

Syphilis clinical diagnosis was operationalized as syphilis infection yes or no which is a dichotomous variable. The diagnosis was based on review, by the trained student researcher, of the electronic charts of all 4,077 STD clinic patients included in the study. The electronic chart contained a diagnosis that the patient was either currently infected with syphilis or had a history of syphilis infection. Both past and current syphilis infection were categorized as syphilis infection yes. A patient with no current syphilis infection or with no history in the chart was categorized as syphilis infection no. The diagnosis of syphilis infection was used as the "gold standard" or recognized reference method for the statistical calculations performed in the study. According to the FDA (2007), a reference standard is "considered to be the best available method for establishing the presence or absence of the target condition" (p. 6). The reference standard can be a single test or a combination of methods including clinical follow-up using nationally recognized guidelines. The clinical diagnosis in the electronic chart was based on the 2012 CDC syphilis case definition (Table 14).

The dichotomous variables operationalized for the laboratory tests (Syphilis IgG and TP-PA) and algorithm interpretations were compared to each patient's diagnosis of syphilis infection yes or no to obtain four possible outcomes (categories): true positive, true negative, false positive, or false negative. The frequencies of the outcomes were placed in a 2x2 contingency table for

calculation of the multiple statistical measures shown in Table 6. Table 28 shows the criteria for each category of the traditional algorithm outcome measurements with the discordant results categorized as either true or false positive based on the diagnosis of syphilis infection. Table 29 shows the criteria for each category of the reverse algorithm outcome measurements with the discordant results categorized as either true or false positive based on the diagnosis of syphilis infection. Table 30 shows the criteria for each category of the Syphilis IgG and TP-PA test based on the diagnosis of syphilis infection with the invalid Syphilis IgG and inconclusive TP-PA results excluded from the study. Based on the criteria shown in Tables 28-30, the student researcher categorized all 4,077 samples in the data set into one of the four categories in the 2x2 contingency table.

For those patients diagnosed with syphilis, the electronic medical record contained the syphilis stage determined by trained STD clinic staff based on the 2012 CDC syphilis case definitions listed in Table 14. The stage information was collected by the student researcher during chart review. The chart contained the following stage categories: primary, secondary, early latent, late latent, latent of unknown duration, tertiary and syphilis not staged. The early latent, late latent, and latent of unknown duration stages were compressed into one latent category. If the patient had a history of syphilis infection, the information was included in the chart. Because the syphilis infection study variable included syphilis infection either current or past as the outcome of syphilis infection yes, a history of syphilis infection was operationalized as a syphilis stage. If the patient had both a history of syphilis infection and a current infection, only the current infection was used for determining the syphilis infection stage. During chart review, it was noted that there were some charts that listed a syphilis diagnosis; however, the stage was not included in the chart. This could be due to a data entry error by clinic staff or the

Table 30

Syphilis IgG and TP-PA Contingency Table Measurement Outcomes

Syphilis infection (past or current)?	**Syphilis IgG reactive**	**Syphilis IgG nonreactive**	**Syphilis IgG invalid (not clinically diagnostic)**	**TP-PA reactive**	**TP-PA nonreactive**	**TP-PA inconclusive (not clinically diagnostic)**
Yes	True positive (TP)	False negative (FN)	Excluded	True positive (TP)	False negative (FN)	Excluded
No	False positive (FP)	True negative (TN)	Excluded	False positive (FP)	True negative (TN)	Excluded

patient was presumptively diagnosed with syphilis based on symptoms and/or outside laboratory test results. The stage section in the chart was left incomplete with the expectation that the patient would return for staging once the syphilis test results were completed. Presumptive syphilis cases often received prophylactic antibiotic treatment and the patient may have decided to either not return for the test results or to go to their primary healthcare provider for follow up care. The samples with a diagnosis of syphilis with no staging were included in the study and operationalized as syphilis not staged. Based on the information listed in this paragraph, the syphilis stages were operationalized as primary, secondary, latent, tertiary, not staged and history.

Data Collection Procedures

Data was collected from various sources by public health laboratory staff during the study time period of January 1, 2012 to December 31, 2013. The Syphilis IgG, RPR, RPR titer, and TP-PA test results were gathered using an electronic query from the public health laboratory LIMS. The LIMS data provided the patient unique accession number, unique medical record

number, gender, date of birth, date of collection, and all three syphilis test results. This data was entered by laboratory staff onto an Excel spreadsheet, which was saved in a secured and protected location on the public health agency firewall protected intranet. The BioRad EVOLIS analyzer was used to perform the Syphilis IgG tests and the S/CO value of each sample was printed on a daily instrument log, which also contained each tested person's unique accession number. The unique accession number obtained from the LIMS query was used to identify the S/CO value listed on the instrument printout. The patient S/CO value was added to the Excel spreadsheet by laboratory staff. The unique medical record number for each patient was used by public health laboratory staff and the student researcher to query the public health agency electronic medical record, which contained patient demographic and infection diagnosis information. Patient information collected from the chart included the following study variables: reason for the clinic visit, STD infection diagnosis, and syphilis stage if syphilis infection present. The variables were recorded and added to the Excel spreadsheet for each patient by either the laboratory staff or the student researcher. The traditional and reverse sequence algorithm serologic interpretations were performed by the student researcher based on the cascade of test results in each algorithm (Figures 4 and 5). The algorithm serologic interpretations of syphilis likely or unlikely were added to the Excel spreadsheet by the student researcher along with the contingency table outcome results determined by the student researcher as shown in Tables 28-30. The student researcher reviewed all 4,077 charts prior to deidentification of study data to verify that the variable results listed on the spreadsheet matched the variables in the chart.

Human Subjects

All syphilis testing performed for this retrospective study was utilized for patient clinical care at a public health STD clinic. The test results were all reported to the patient at the time of treatment. The only testing performed on the samples collected during the study time period were the tests ordered by the public health clinic staff or reflexed as approved by the public health laboratory director. All testing was performed to protect the health of the population within the public health agency's jurisdiction. The data was not collected for the proposed research project and was collected as part of an ongoing public health agency QA project during the study time period. The study data included in the Excel spreadsheet was deidentified by an independent third party within the public health agency with removal of all Protected Health Information (PHI) such as name, date of birth, date of collection, medical record number, and laboratory accession number prior to the start of data analysis. The data set did not include any of the 18 Health Insurance Portability and Accountability Act (HIPAA) identifiers. No minor children under the age of 18 years were included in the study data. Only aggregate deidentified non-PHI information was accessible by the student researcher for the data analysis. The student researcher did not have the ability to link the test results back to a specific individual.

The Council of State and Territorial Epidemiologists (CSTE) published a report in 2004 that provided guidance on the distinction between public health practice and public health research. According to the report, public health practice is about protecting the public's health by performing epidemiological investigations, surveillance, programmatic evaluation and clinical care of the population. Identifiable health data is collected and analyzed by the public health authority while conducting these activities. The data is used to protect the health of the population within the health authority's jurisdiction (CSTE, 2004). The western region public

health agency did not require Institutional Review Board (IRB) approval to utilize the deidentified data set collected as a component of its public health practice. Letters of approval to utilize the deidentified data set for this dissertation study were obtained from the public health agency directors prior to data analysis.

An IRB proposal for the study was submitted to Virginia Commonwealth University (VCU) prior to the start of data analysis. The blood samples were collected for clinical testing performed as a component of clinical care activities within the public health agency STD clinic. Because the research involved retrospective data that was collected solely for medical treatment or diagnosis; existed prior to the research was proposed; and was not collected for the proposed research project, an Exempt Category 4 review was requested from the VCU IRB. The VCU IRB returned a decision that the research study was not subject to the Health and Human Services (HHS) Regulations for the Protection of Human Subjects and did not require IRB review or approval. According to the VCU IRB determination (T. Nadder, personal communication, June 7, 2016), Section 45 CFR 46.102(d) defines research and Section 45 CFR 46.102(f) defines a human subject. The VCU IRB determined that the research study did not meet both human subject and research definitions; therefore, it was not subject to the federal regulations.

Data Analysis

All deidentified syphilis variable outcome data in the Microsoft Excel spreadsheet was exported to Number Cruncher Statistical Software (NCSS) version 11 (NCSS, 2016). Frequency statistics along with mean and range were calculated using NCSS 11 for age to provide demographic background of the study population. Frequency counts for the reason for clinic visit, infection diagnosis, and syphilis stage variables were also calculated using NCSS 11 to

provide additional background information of the study population. Statistical analysis of the observed data was performed using NCSS 11 as listed in Table 24 and detailed in the following paragraphs. A summary of the hypotheses, theoretical domains, variables and analyses used in the study is provided in Table 31.

The study research question was: What is the usefulness of the Captia™ Syphilis IgG EIA test method and the reverse algorithm for detection of syphilis infection in a public health population? As was discussed in the literature review section, laboratory test usefulness is composed of multiple diagnostic test measures as calculated in Table 6. Statistical measures for determining syphilis test usefulness included sensitivity, specificity, predictive values, likelihood ratios, accuracy and/or ROC curves. The three study hypotheses were:

- $H1_o$: Among patients seeking STD services at a large western region metropolitan public health clinic, there will be no difference between the Captia™ Syphilis IgG EIA and Fujirebio Serodia TP-PA test results.
- $H2_o$: Among patients seeking STD services at a large western region metropolitan public health clinic, there will be no difference in diagnostic interpretation of the Captia™ Syphilis IgG EIA S/CO value and Becton Dickinson MacroVue RPR titer result.
- $H3_o$: Among patients seeking STD services at a large western region metropolitan public health clinic, there will be no difference between the syphilis traditional and reverse algorithm interpretations.

Statistical analysis using NCSS 11 for the first null hypothesis "Among patients seeking STD services at a large western region metropolitan public health clinic, there will be no difference

Table 31

Study Hypothesis, Theoretical Domains, Variables, and Analyses

Null hypothesis	**Theoretical domain**	**Variable**	**Analyses**
Among patients seeking STD services at a large western region metropolitan public health clinic, there will be no difference between the Captia™ Syphilis IgG EIA and Fujirebio Serodia TP-PA test results	Analytical (Process)	Syphilis IgG and TP-PA result (dichotomous) and contingency table value (categorical) Clinical diagnosis (dichotomous) Gold standard	Sensitivity, specificity, predictive value, accuracy and likelihood ratio were calculated to determine usefulness of each treponemal test to predict syphilis infection using clinical diagnosis as the gold standard. McNemar test was performed to determine differences between the two tests and test the hypothesis.
Among patients seeking STD services at a large western region metropolitan public health clinic, there will be no difference in diagnostic interpretation of the Captia™ Syphilis IgG EIA S/CO value and Becton Dickinson MacroVue RPR titer result	Analytical (Process)	Syphilis IgG S/CO value and RPR titer (continuous) Clinical diagnosis (dichotomous) Gold standard	ROC AUC values were calculated to determine accuracy of the Syphilis IgG test and RPR titer for predicting syphilis infection and test the hypothesis. Optimum cutoff values were determined for Syphilis IgG.
Among patients seeking STD services at a large western region metropolitan public health clinic, there will be no difference between the syphilis traditional and reverse algorithm interpretations	Post analytical (Outcome)	Traditional and reverse algorithm serologic interpretation (dichotomous) and contingency table value (categorical) Clinical diagnosis (dichotomous) Gold standard	Sensitivity, specificity, predictive value, accuracy, and likelihood ratio were calculated to determine usefulness of each algorithm for predicting syphilis infection using clinical diagnosis as the gold standard. McNemar test was performed to determine differences between algorithms and test the hypothesis.

between the Captia™ Syphilis IgG EIA test and the Fujirebio Serodia TP-PA test" included calculation of sensitivity, specificity, predictive value, accuracy, and likelihood ratio using syphilis infection yes or no as the gold standard. The Syphilis IgG and TP-PA are both treponemal tests and, therefore, will detect treponemal antibodies, IgG for the Syphilis IgG and both IgG and IgM for the TP-PA test. The CDC (2011) recommends use of the TP-PA test as a confirmatory test for discordant results within the reverse algorithm and as the confirmatory treponemal test within the traditional algorithm. CDC also recommends that the confirmatory test should have a sensitivity that is equal to or greater than the screening treponemal test. Determining the sensitivity of the Syphilis IgG and TP-PA test will add to the existing body of knowledge regarding both tests. The Syphilis IgG and TP-PA tests were performed on the same sample and were likely to be substantially correlated. According to Gray and Kinnear (2012), the statistical test to determine differences between two related test methods with nominal data is the McNemar test which was utilized to test the first null hypothesis.

Statistical analysis using NCSS 11 for the second null hypothesis "Among patients seeking STD services at a large western region metropolitan public health clinic, there will be no difference in diagnostic interpretation of the Captia™ Syphilis IgG EIA S/CO value and Becton Dickinson MacroVue RPR titer result" included use of the empirical ROC method of Delong, Delong and Clarke-Pearson (1988) to compare the two AUC values of the S/CO and titer continuous variables. Each sample had a known condition value (syphilis infection yes or no) and numerical value for each test. The total range of cutoff values from 0 to 5.5 was included to determine if the FDA approved Syphilis IgG cutoff values of 0.9 and 1.0 for equivocal results and 1.1 for reactive results represents the optimum cutoff. The total range of 0, 1, 2, 4, 8, 16, 32, 64, 128, 256, 512, and 1024 was used to determine the RPR titer cutoff values. The two ROC

curves were compared to determine differences and confidence intervals for the AUC to test the second null hypothesis. The Syphilis IgG was the initial test in the reverse algorithm and the RPR was the initial test in the traditional algorithm; therefore, an understanding of the diagnostic accuracy of these first initial tests will expand the body of knowledge about the usefulness of both algorithms in a public health setting.

Statistical analysis using NCSS 11 for the third null hypothesis "Among patients seeking STD services at a large western region metropolitan public health clinic, there will be no difference between the syphilis traditional and reverse algorithm interpretations" included calculation of sensitivity, specificity, predictive value, accuracy, likelihood ratio, and performance of McNemar test for each serologic algorithm interpretation (syphilis likely or unlikely) using the clinical diagnosis of syphilis infection yes or no as the gold standard to test the third null hypothesis. There is little published information detailing the usefulness measures listed for either algorithm and few published studies with statistical analysis of a direct comparison of the algorithms. Most studies have only compared lab test to lab test and not to the clinical diagnosis. This statistical analysis will expand the body of knowledge regarding usefulness of either algorithm in a public health setting, using diagnosis as the gold standard.

Summary

This retrospective, nonexperimental descriptive correlational study assessed the usefulness of the Captia™ Syphilis IgG EIA test and the syphilis reverse algorithm serologic interpretation for detection of syphilis infection among patients seeking care at a large western region metropolitan public health STD clinic. The theoretical framework for the study was based on the Donabedian Quality Framework with a target construct of syphilis infection and a proposition that useful syphilis testing will improve syphilis infection detection. The process or analytic domain

activities determined the usefulness of the syphilis test results. The outcome or post-analytic domain activities determined the usefulness of the syphilis reverse and traditional serologic interpretations. The three null hypotheses were:

- $H1_o$: Among patients seeking STD services at a large western region metropolitan public health clinic, there will be no difference between the Captia™ Syphilis IgG EIA and Fujirebio Serodia TP-PA test results.
- $H2_o$: Among patients seeking STD services at a large western region metropolitan public health clinic, there will be no difference in diagnostic interpretation of the Captia™ Syphilis IgG EIA S/CO value and Becton Dickinson MacroVue RPR titer result.
- $H3_o$: Among patients seeking STD services at a large western region metropolitan public health clinic, there will be no difference between the syphilis traditional and reverse algorithm interpretations.

Design validity was enhanced by use of standardized measurement methods, providing staff training, and use of automation. Selection bias was controlled by use of a homogeneous public health population with each sample in the study tested by all three syphilis test methods. Each sample acted as its own matching control thus controlling temporal ambiguity. External design validity was enhanced by using samples collected in a real-world public health setting.

Measurement reliability was enhanced through use of repeated measures that provided consistent reactive and nonreactive qualitative results across multiple reagent lot numbers using known QC material. The automated EVOLIS analyzer had a lower than expected random error rate and thus provided consistent day to day test results. The data spreadsheet was extensively reviewed prior to deidentification to ensure all data values were reliably entered.

Measurement validity was supported by FDA approval of all three syphilis tests for detection of syphilis infection. All three tests provided expected qualitative test results for external reference material known to contain syphilis antibodies. According to the manufacturer's package inserts all three syphilis tests were capable of discriminating syphilis infection from other diseases or conditions.

The study setting was a public health STD clinic located in a large western metropolitan area with a population of over two million residents. The clinic was operated by a public health agency and participated in federally approved syphilis detection and surveillance programs. The STD clinic provided services to persons who had signs and symptoms of syphilis infection as well as those who were contacts to infected individuals. Based on CDC surveillance recommendations, the clinic also targeted pregnant women, MSM, and persons with HIV infection for testing and treatment. All syphilis tests ordered by the STD clinic staff were performed at the public health laboratory, which was also owned by the public health agency. All clinic staff were trained to perform syphilis infection clinical assessment following standard CDC STD guidelines.

The target population included all patients who sought clinical services at the STD clinic from January 1, 2012 through December 31, 2013. Purposive sampling was used to select sample members to include in the study. The selection criteria for inclusion in the study was males and females 18 years and older who were seen in the public health STD clinic, met the STD clinic criteria for syphilis testing, and had all three syphilis tests (Syphilis IgG, RPR, and TP-PA) performed with clinically diagnostic results on each sample collected. Exclusion criteria included males and females under 18 years of age, those who did not visit the STD clinic, those that did not have three syphilis tests performed, samples that tested invalid by the Syphilis IgG

test, and samples that tested inconclusive by the TP-PA test method. Statistical power analysis indicated that the sample study size of 4,077 achieved 84% power to detect a change in sensitivity from 0.8 to 0.9 and 100% power to detect a change in specificity from 0.8 to 0.9.

The public health laboratory was owned by the public health agency and coordinated testing methods with the clinic staff. All laboratory reflex testing was approved by the laboratory director. Trained public health laboratory staff performed all testing and gathered the study variables during the study time period as a component of the enhanced syphilis surveillance performed by the public health agency. Laboratory staff entered the data on an Excel spreadsheet; which was stored in a secured, protected location on the public health agency firewall protected intranet. Descriptive variables included gender, age, reason for clinic visit, infection diagnosis, and syphilis stage. Observed variables included the Captia^{TM} Syphilis IgG test result and S/CO value, Becton Dickinson MacroVue RPR test result and titer value, Fujirebio Serodia TP-PA test result, syphilis traditional and reverse algorithm serologic interpretation, algorithm and contingency table values, and clinical diagnosis of syphilis based on chart review. All observed variables were dichotomous, except for the S/CO and RPR titer value which was continuous and contingency table values which was categorical. The data set was deidentified of all PHI prior to data analysis and consent to use the data set was obtained from the public health agency directors. According to VCU, IRB approval was not required as the study did not meet both human subject and research definitions and, therefore, was not subject to the federal regulations.

The approach for determining laboratory test usefulness is not clearly defined; however, numerous statistical tests can be combined to provide a comprehensive determination. For the dissertation study, usefulness was measured by sensitivity, specificity, accuracy, predictive

values, and likelihood ratios for dichotomous variables using a 2x2 contingency table. McNemar's test was performed to determine differences in proportions when comparing two test methods and test the hypotheses. For continuous quantitative values, such as the Syphilis IgG S/CO and the RPR titer, a ROC curve analysis was performed to determine the ability of each measurement to correctly classify syphilis infection. This analysis was also used to determine an optimum cutoff value. The AUC was calculated to provide a determination of the ability of each test to discriminate between true- and false-positive test results. When two diagnostic tests are performed on the same sample, the AUC can be used to determine which test is better when the two test results are interpreted independently against the same gold standard and therefore test the study hypothesis.

Chapter Four: Results

This chapter will provide details of the statistical analysis of all study variables and hypotheses testing results. The result data will be presented in tabular form with a brief explanation of the results for both the descriptive and observed variables. Statistical analysis and hypothesis testing for each null hypothesis is presented separately with a brief summary concluding the chapter.

Descriptive Variables

Frequency statistics that describe the study population were calculated using NCSS 11 software for gender, age, reason for clinic visit, infection diagnosis, and syphilis stage. Males comprised 68.56% (2795/4077) of the study population with 31.30% (1276/4077) females and 0.15% (6/4007) transgender patients (Table 32). The larger proportion of males in the study was expected as the CDC recommends targeted testing within male populations due to their high syphilis rates. According to CDC surveillance data (2017), men accounted for almost 90% of all P&S syphilis cases reported in 2016.

Table 32

Gender Frequency Count and Percentage (N=4,077)

Gender	Frequency Count	Percentage
Female	1276	31.30
Male	2795	68.55
Transgender	6	0.15

The 4,077 study samples were grouped by age as shown in Table 33. The 18-20 and the 21-30 age groups combined totaled 50.1% (2043/4077) of the study sample. The 21-30 age group alone comprised 44.3% (1807/4077) of the study sample. The mean age of the study population is 33 years with a standard deviation of 11.7. According to CDC surveillance data (2017), in 2016, men aged 24-29 years had the highest rate of reported P&S syphilis cases (48.5/100,000 population) compared with any other age group for either gender. Based on CDC surveillance data and targeted testing recommendations, the under 30 age group would be expected to be the most prevalent within this study data.

Table 33.

Age Group Frequency Count and Percentage (N=4,077)

Study age group (years)	Frequency Count	Percentage
18-20	236	5.8
21-30	1807	44.3
31-40	942	23.1
41-50	672	16.5
51-60	305	7.5
61-70	94	2.3
71-80	13	0.3
81-90	8	0.2
Total 4077		100.0
Mean 33		
SD 11.7		

Frequency statistics were calculated for the reason for the clinic visit using NCSS 11 software with each sample placed into one of three categories (Table 34). Those patients who visited the STD clinic and had symptoms of any STD were categorized as STD screen with symptoms. They comprised 22.4% (912/4077) of the study population. The largest number of patients, 42.6% (1736/4077) visited the STD clinic for an STD screen and did not exhibit symptoms. Standard practice at the STD clinic was to perform HIV, syphilis, gonorrhea, and

Table 34

Reason for Clinic Visit Frequency Count and Percentage (N=4,077)

Reason for clinic visit	Frequency count	Percentage
STD screen with symptoms	912	22.4
STD screen without symptoms	1736	42.6
Notification	1429	35.0

chlamydia testing on patients who presented to the clinic for STD screening with or without symptoms. A person can be infected with more than one STD at the same time and syphilis diagnosis is especially problematic because, after the secondary phase, the symptoms disappear. The next highest reason for clinic visit representing 35.0% (1429/4077) of the study were patients who had received notification in the form of a letter or call that they had been exposed to an STD, had received notification from an infected partner, were referred by a private physician for testing, or had a positive syphilis screening test with no confirmation from an outside facility, e.g. plasma center or blood bank. Depending on the person's risk and the clinical assessment, STD clinic staff could order a panel of all three syphilis tests, Syphilis IgG, RPR, and TP-PA, on any of these patients.

Each patient seen in the STD clinic received a clinical diagnosis of STD infection or no STD infection. The infection diagnosis was operationalized as no STD, syphilis, HIV, chlamydia and/or gonorrhea, or other. The other infection category included diagnosis of herpes, warts, molluscum contagiosum, condyloma, HPV, yeast, NGU, bacterial vaginosis, trichomoniasis, lupus, gout, or autoimmune disease. As shown in Table 35, no STD infection was diagnosed in 61.7% (2517/4077) of the patients included in the study. Many of the patients visiting the STD clinic were there for STD screening to ensure that they were not infected with an STD; however, they may have exhibited a high risk behavior or had contact with a person

Table 35

Infection Diagnosis Frequency Count and Percentage (N=4,077)

Infection diagnosis	Frequency count	Percentage
Syphilis	859	21.1
Gonorrhea and/or chlamydia	360	8.8
HIV	20	0.5
Other infection	321	7.9
No STD infection	2517	61.7

infected with an STD, which prompted the STD clinic staff to order all three syphilis tests (Syphilis IgG, RPR, and TP-PA). Patients with no STD infection could also be within groups targeted by the CDC for additional screening, specifically males under 30 years of age. The no STD infection group also included patients with contact to HIV, trichomonas, herpes, gonorrhea, chlamydia, syphilis, or NGU infections. Within the no STD infection category, there were 712 patients who were identified as having contact with an STD. Of those 712 patients, 72% (511/712) visited the clinic due to contact with a person infected with syphilis. Syphilis infection either past or current was diagnosed in 21.1% (859/4077) of the study patients.

The syphilis stage was charted for each patient diagnosed with a syphilis infection. All syphilis stages except tertiary were represented in the study (Table 36). Due to antibiotic use in the US and aggressive public health outreach, tertiary cases are seldom seen in patients visiting a public health STD clinic and would most likely be seen in a hospital setting due to the serious disease complications which can occur in this late syphilis stage. Latent syphilis cases, which included early, late, and unknown latent stages comprised 60.2% (517/859) of the syphilis study cases. Of the latent syphilis cases, 86.8% (449/517) were staged as late latent syphilis, which is an asymptomatic stage. Patients with a history of syphilis infection were included with syphilis stage frequency counts and represent 19.7% (169/859) of the syphilis infection cases in the

Table 36

Syphilis Stage Frequency Count and Percentage (N = 859)

Syphilis stage	Frequency count	Percentage
Primary	49	5.7
Secondary	109	12.7
Latent (includes early, late and unknown)	517	60.2
Tertiary	0	0
Not staged	15	1.7
History of syphilis infection	169	19.7

study. The study variable "syphilis infection present" includes patients with a prior history of syphilis infection. There was no chance of overlap in frequency counts among those with a history of infection with the other stages. If a study patient had a history of syphilis infection and a current diagnosis of syphilis infection, the history of infection was not used as the sample represented a current infection and the current infection stage was assigned to the sample.

Observed variables

Statistical analysis of the observed variables was calculated using NCSS 11 statistical software, and results are presented in the explanation of each null hypothesis and usefulness measures. The 2 x 2 contingency table calculations are detailed in Table 6.

First null hypothesis. Among patients seeking STD services at a large western region metropolitan public health clinic, there will be no difference between the Captia^{TM} Syphilis IgG EIA and Fujirebio Serodia TP-PA test results.

Before performing statistical analysis of the first null hypothesis, it was necessary to determine if there was a bias for including the equivocal Syphilis IgG test results as reactive. As recommended in the FDA (2007) guidance document, two sets of usefulness performance measures were calculated: one with equivocal results counted as reactive (Table 37) and one with equivocal results counted as nonreactive (Table 38) using syphilis infection (yes or no) as the

Table 37

Captia™ Syphilis IgG Contingency Table (Equivocal Counted as Reactive) (N = 4,077)

	Syphilis infection Yes	**Syphilis infection No**	**Total**	**Measures**
Reactive Syphilis IgG test	TP 707	FP 42	749	94% PPV
Nonreactive Syphilis IgG test	FN 152	TN 3176	3328	95% NPV
Total	859	3218	4077	
Measures	82% sensitivity 63.06 +LR	99% specificity 0.18 -LR		95% accuracy 21% prevalence

Note: TP = true positive; FP = false positive; FN = false negative; TN = true negative; +LR = positive likelihood ratio; -LR = negative likelihood ratio; PPV = positive predictive value; NPV = negative predictive value

Table 38

Captia™ Syphilis IgG Contingency Table (Equivocal Counted as Nonreactive) (N = 4,077)

	Syphilis infection Yes	**Syphilis infection No**	**Total**	**Measures**
Reactive Syphilis IgG test	TP 686	FP 9	695	99% PPV
Nonreactive Syphilis IgG test	FN 173	TN 3209	3382	95% NPV
Total	859	3218	4077	
Measures	80% sensitivity 285.54 +LR	100% specificity 0.20 -LR		96% accuracy 21% prevalence

Note: TP = true positive; FP = false positive; FN = false negative; TN = true negative; +LR = positive likelihood ratio; -LR = negative likelihood ratio; PPV = positive predictive value; NPV = negative predictive value

gold standard. The results listed in Tables 37 and 38 show slight differences in all performance measures except the positive likelihood ratio which was higher (285.54) when the equivocal results were included as nonreactive. To determine if there was a significant difference in performance measures between including Syphilis IgG equivocal results as either reactive or nonreactive, a McNemar test was performed using NCSS 11 software. The McNemar test was not statistically significant, $p = 1.025$ ($\alpha = 0.05$), indicating that there was no difference between

including the equivocal results as either reactive or nonreactive; therefore, the null hypothesis was not rejected (Table 39). Based on the McNemar test that there was no bias present, the manufacturer's package insert recommendation that repeat equivocal results should be considered reactive, and standard practice at the western regional laboratory, the Syphilis IgG equivocal results were considered to be reactive and the results shown in Table 37 were used for all study analyses.

Table 39

McNemar Two-sided Hypothesis Test of the Difference Between Captia™ Syphilis IgG Equivocal Counted as Reactive or Nonreactive (N = 4,077)

H_0: $P_1-P_2= 0$ vs H_a: $P_1-P_2 \neq 0$

Statistical Test	**Proportion Difference** P_1-P_2	**95% Confidence interval (CI)**	**Confidence Interval Width**	**Test Statistic Value**	**Probability level** $\alpha = 0.05$
McNemar	-0.0029	-0.0067 to -0.0006	0.0.0073	2.667	1.025

Note: Proportion (risk) difference ($\delta = P_1 – P_2$) is a direct method of comparison between two event probabilities.

Using the data presented in Table 40, the sensitivity, specificity, predictive value, likelihood ratio, and accuracy measures were calculated for the Fujirebio Serodia TP-PA using the diagnosis of syphilis infection yes or no as the gold standard. Syphilis infection yes included both past and present syphilis infection. These measures provide supporting statistical data for determining the usefulness of the Captia™ Syphilis IgG test for detection of syphilis infection and were calculated using NCSS 11 statistical software. Calculation formula details for each measure are shown in Table 6.

The column contingency table calculations shown in Table 41 are based on the true condition (syphilis infection). These calculations include sensitivity and specificity which are reported with 95% confidence intervals (CI) for each treponemal test. Sensitivity and specificity calculations are not influenced by the population disease prevalence. The Captia™ Syphilis IgG

Table 40

Fujirebio Serodia TP-PA Contingency Table (N = 4,077)

	Syphilis infection Yes	Syphilis infection No	Total	Measures
Reactive TP-PA test	TP 794	FP 9	803	99% PPV
Nonreactive TP-PA test	FN 65	TN 3209	3274	98% NPV
Total	859	3218	4077	
Measures	92% sensitivity 330.50 +LR	100% specificity .08 -LR		98% accuracy 21% prevalence

Note: TP = true positive; FP = false positive; FN = false negative; TN = true negative; +LR = positive likelihood ratio; -LR = negative likelihood ratio; PPV = positive predictive value; NPV = negative predictive value

Table 41

Captia™ Syphilis IgG and Serodia TP-PA Sensitivity, Specificity, 95% CI

Test	Sensitivity	95% CI	Specificity	95% CI
Captia™ Syphilis IgG[a]	0.82	0.80-0.85	0.99	0.98-0.99
Serodia TP-PA	0.92	0.90-0.94	1.00	0.99-1.00

Note: CI = Confidence interval; Proportion Confidence Interval Method: Exact Binomial
[a]Equivocal results included as reactive

test exhibited a lower sensitivity (82%) and specificity (99%) than the Serodia TP-PA test which had a sensitivity of 92% and specificity of 100%. An ideal test would have 100% sensitivity and specificity.

The row contingency table calculations are based on the predicted condition of syphilis infection, and the results are influenced by the disease prevalence within the population. These calculations include PPV and NPV which are reported with 95% CI for each treponemal test in Table 42. The Captia™ Syphilis IgG test exhibited lower PPV (94%) and NPV (95%) than the Serodia TP-PA test which had a PPV of 99% and NPV of 98%. The PPV and NPV calculations with 95% CI in Table 42 were based on the estimated prevalence of 21% calculated within the study population of 4,077 patients. The PPV and NPV were adjusted based on the 3%

Table 42

Captia™ Syphilis IgG and Serodia TP-PA PPV, NPV, 95% CI, Adjusted PPV and NPV

Test	PPV	95% CI	NPV	95% CI	Adj PPV 3% prevalence	Adj NPV 3% prevalence
Captia™ Syphilis IgG[a]	0.94	0.92-0.96	0.95	0.95-0.96	0.66	0.99
Serodia TP-PA	0.99	0.98-0.99	0.98	0.97-0.98	0.91	1.00

Note: PPV = positive predictive value; NPV = negative predictive value; CI = Confidence interval; Adj PPV = known prevalence (3%) adjusted positive predictive value; Adj NPV = known prevalence (3%) adjusted negative predictive value; Proportion Confidence Interval Method: Exact Binomial
[a]Equivocal results included as reactive

prevalence calculated for the STD clinic population (859 syphilis infections/34340 patients with syphilis IgG tests*100). The adjusted value calculations do not provide 95% CI. The Syphilis IgG adjusted PPV is much lower (66%) than the unadjusted PPV and the adjusted NPV is higher (99%) than the unadjusted NPV. The TP-PA adjusted PPV dropped slightly (91% vs 99%) and the adjusted NPV increased to 100%. An ideal test would have 100% PPV and NPV.

Other calculations are based on values present in the whole table and include likelihood ratios, accuracy, and prevalence which are reported with 95% CI for each treponemal test in Table 43. Likelihood ratios and accuracy are not influenced by the disease prevalence in the population. The 21% prevalence reported in the table was based on the study population of 4,077 patients, and does not reflect the general STD clinic population. The Captia™ Syphilis IgG test exhibited a lower positive LR (63.06) than the Serodia TP-PA (330.50). According to Table 5, a positive LR greater than 10 provides strong evidence for ruling in a disease diagnosis. The Captia™ Syphilis IgG exhibited a higher negative LR (0.18) than the TP-PA (0.08). According to Table 5, a negative LR less than 0.1 provides strong evidence for ruling out a disease diagnosis. A negative LR between 0.1 and 0.2 provides moderate evidence for ruling out a diagnosis. The accuracy of the Syphilis IgG test (95%) is less than the TP-PA accuracy (98%).

Table 43

Captia™ Syphilis IgG and Serodia TP-PA Likelihood Ratios, 95% CI, Accuracy, and Study Prevalence

Test	Positive LR	95% CI	Negative LR	95% CI	Accuracy	Prevalence (Study)
Captia™ Syphilis IgG[a]	63.06	46.62-85.30	0.18	0.16-0.21	0.95	0.21
Serodia TP-PA	330.50	172.07-634.79	0.08	0.06-0.10	0.98	0.21

Note: LR = likelihood ratio; CI = Confidence interval; Ratio Confidence Interval Method: Katz Logarithm

[a]Equivocal results included as reactive

An ideal test would exhibit 100% accuracy.

Hypothesis testing of the first null hypothesis was performed using the McNemar test to determine if there was a difference between the Captia™ Syphilis IgG and Serodia TP-PA treponemal tests for detecting syphilis infection. The frequency table used to input data into NCSS 11 is shown in Table 44. The McNemar test was statistically significant, $p < 0.0002$ ($\alpha = 0.05$), indicating that there was a difference between the two treponemal test results and therefore the null hypothesis was rejected (Table 45).

Table 44

McNemar Test Captia™ Syphilis IgG and Serodia TP-PA Frequency Table (N = 4,077)

	TPPA Reactive	TPPA Nonreactive	Total	Marginal Proportion $P_1=(A+B)/N$
Syphilis IgG Reactive[a]	674 A	75 B	749	$P_1=0.1837$ (749/4077)
Syphilis IgG Nonreactive	129 C	3199 D	3328	
Total	803	3274	4077	
Marginal Proportion $P_2=(A+C)/N$	$P_2=0.1970$ (803/4077)			

Note: Marginal Proportions P_1 and P_2 represent marginal homogeneity or probability that the row and column frequencies are equal.

[a]Equivocal results included as reactive

Table 45

McNemar Two-sided Hypothesis Test of the Difference Between Captia™ Syphilis IgG and Serodia TP-PA (N = 4,077)

H_0: P_1-P_2 = 0 vs H_a: P_1-P_2 ≠ 0

Statistical Test	Proportion Difference P_1-P_2	95% Confidence interval (CI)	Confidence Interval Width	Test Statistic Value	Probability level $\alpha = 0.05$
McNemar	-0.0132	-0.0202 to -0.0064	0.0138	14.294	0.0002

Note: Proportion (risk) difference ($\delta = P_1 - P_2$) is a direct method of comparison between two event probabilities.

Second null hypothesis: Among patients seeking STD services at a large western region metropolitan public health clinic, there will be no difference in diagnostic interpretation of the Captia™ Syphilis IgG EIA S/CO value and Becton Dickinson MacroVue RPR titer result.

NCSS 11 software was used to first determine the AUC for the Syphilis IgG and RPR titer separately to ensure that each test could distinguish the presence or absence of syphilis infection. The presence of syphilis infection included both past and present infection. The NCSS 11 software utilized the DeLong et al. (1988) empirical (nonparametric) method for AUC computation. The AUC was determined using the S/CO and RPR titer values for all 4,077 samples in the study compared to the clinical determination of presence or absence of syphilis infection. According to Zwieg and Campbell (1993), the AUC provides a single number which expresses a test's diagnostic accuracy. The possible AUC values range from 0.5 for a test with no diagnostic ability to distinguish disease from nondisease to 1.0 for a test with perfect diagnostic ability. As shown in Table 43, the AUC value for the Captia™ Syphilis IgG S/CO was 0.9500 and the MacroVue RPR titer was 0.8155. A *z* test was performed to determine the probability that the observed sample AUC for either the Syphilis IgG or RPR titer can be found when the true (population) AUC is 0.5 which represents a test with no syphilis diagnostic ability.

As shown in Table 46, both Syphilis IgG and RPR titer values were statistically significant, $p <$ 0.0000 ($\alpha = 0.05$); therefore, the null hypothesis that the AUC = 0.5 was rejected, indicating both tests have the ability to distinguish the presence or absence of syphilis infection.

Table 46

Hypothesis Test for AUC Comparison of Ability of Syphilis IgG S/CO Value and RPR Titer Value to Detect Syphilis Infection (N = 4,077)

H_0: AUC = 0.5 vs H_a: AUC ≠ 0.5 (0.5 represents a test with no diagnostic ability)
Known condition value: syphilis infection no or yes (yes includes current and past infection)

Test criterion	**AUC**	**95% CI**	**Standard Error**	**Z-Value to test AUC >0.5**	***p* value** $\alpha = 0.05$
Syphilis IgG S/CO value	0.9500	0.9382-0.9596	0.0054	83.316	0.0000
RPR titer value	0.8155	0.7986-0.8311	0.0083	38.098	0.0000

Note: S/CO = Signal to cutoff; RPR = Rapid plasma reagin; AUC = Area under curve; CI = Confidence interval

Hypothesis testing of the second null hypothesis was performed using NCSS 11 software to compare the Syphilis IgG and RPR titer AUCs using a two-sided *z* test. The procedure compared the ROC curves for paired sample cases with each subject having a known condition (syphilis present or absent) and continuous values from two diagnostic tests (Syphilis IgG and RPR titer) performed on each sample. As shown in Table 47, the difference between the interpretation of the Syphilis IgG S/CO and RPR titer AUCs calculated in Table 46 was compared using a *z* test and the *p* value was statistically significant ($p < 0.0000$, $\alpha = 0.05$). Therefore the second null hypothesis that there was no difference between the Syphilis IgG S/CO value and RPR titer value was rejected.

According to Zweig and Campbell (1993), a ROC curve provides a graphical representation of a test's ability to classify a condition (syphilis infection present or absent) "over all decision thresholds" (pg.564). They recommend that accuracy assessment using a ROC curve should not be performed at a designated single point as it may provide erroneous information when

Table 47

Hypothesis Test Comparing Two AUCs (Empirical Estimation) for Syphilis IgG S/CO Value and RPR Titer Value (N = 4,077)

H_0: AUC1 = AUC2 vs H_a: AUC1 ≠ AUC2
Paired Criterion Variables: Syphilis IgG S/CO value = AUC1; RPR titer value = AUC2

Syphilis IgG S/CO AUC1	**RPR Titer AUC2**	**Difference AUC1-AUC2**	**95% CI**	**Difference Standard Error**	**Z value**	***p* value** $\alpha = 0.05$
0.9500	0.8155	0.1345	0.1168-0.1523	0.0091	14.849	0.0000

Note: S/CO = signal to cutoff; AUC = area under curve; CI = confidence interval

comparing two different tests. The use of an optimized cut off value is necessary when reporting a test result that will be used for patient treatment. The ROC curve is not influenced by disease prevalence and plots the true positive rate (sensitivity) against the false positive rate (1-specificity) for all possible decision points. In viewing the ROC curve, the more diagnostically accurate a test is, the more quickly the curve nears the upper left corner which represents perfect sensitivity (100%). The diagonal line on the ROC plot is a reference line that represents a test with no discriminatory ability (random classification). For visual comparison purposes, the ROC curves for the Syphilis IgG S/CO and RPR titer were plotted together in Figure 15. The Syphilis IgG S/CO ROC curve lies above and to the left of the RPR titer ROC curve, indicating that the Syphilis IgG S/CO test value has greater observed diagnostic accuracy.

In order to obtain FDA in vitro diagnostic device (IVD) approval, manufacturers must include a description of how positive, negative, equivocal (if applicable), and invalid test results are determined. They must indicate the cut off values for all assay outputs in the package insert (FDA, 1994; FDA, 2011). The clinical laboratory must follow the assay cut off values listed in the FDA approved manufacturer's package insert. Any deviation from the manufacturer's

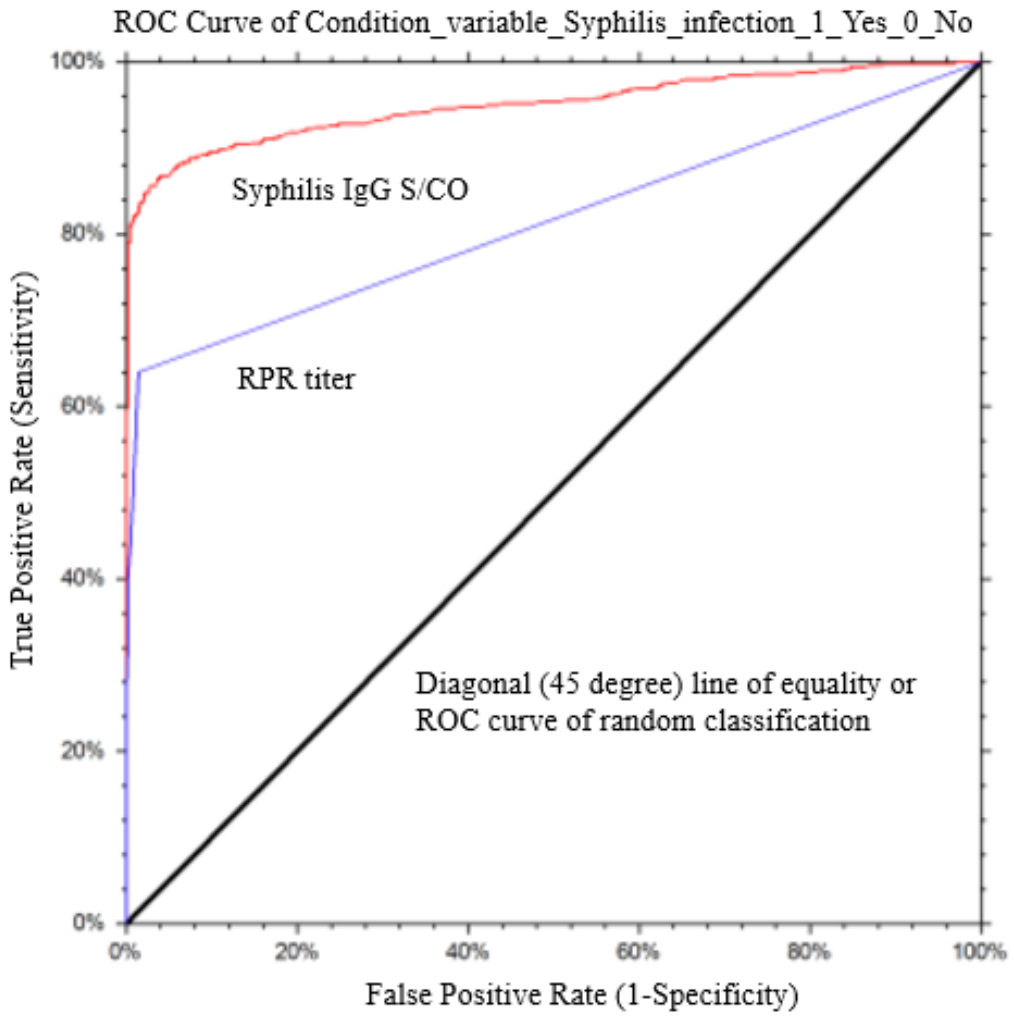

Figure 15. Graphical Comparison of Syphilis IgG S/CO and RPR Titer ROC Curves

package insert cut off values requires that the clinical laboratory must then perform a complete extensive method validation for the off-label use.

According to the FDA approved manufacturer's package insert (Trinity Biotech, 2003), the cut off values for the Captia™ Syphilis IgG were "derived from clinical trials as the value giving optimum discrimination between specimens which are reactive or nonreactive for antibodies to *T. pallidum* as characterized by a range of standard serological techniques" (p. 3). The Captia™ Syphilis IgG test results based on the S/CO cut off values are as follows:

Nonreactive: S/CO value less than or equal to 0.9

Equivocal: S/CO value between 0.9 and 1.1

Reactive: S/CO value greater than or equal to 1.1

The ROC curve analysis provides data for determining if the manufacturer's cut off values represent the optimal decision threshold. Perkins and Schisterman (2006) recommend the use of the Youden Index, which is the point on the ROC curve that is farthest from the diagonal line (line of equality) with a higher Youden Index indicating an optimal cut off value. The Youden Index is equivalent to the difference between true positive rate (TPR) and true negative rate (TNR), thus maximizing the correct classification rate. An alternate optimal decision threshold measure is the distance from the top left corner of the ROC curve to the point on the ROC curve, also called the "point closest to". With this measure, a lower value would indicate an optimal cut off value. With two different measurements, it would be possible to get two different optimal cut off values. According to Hajian-Tilaki (2013) and Perkins and Schisterman (2006), the Youden Index is the preferred method for determining an optimal cutoff. While it is important to choose a point that classifies the most number of patients correctly and thus classifies the least number incorrectly, the cost and social consequences of misclassification as well as the disease prevalence in the tested population must also be considered (Ridge and Vizard, 1993). The various indices shown in Table 48 can assist with determining if the cut off values provided in the FDA approved manufacturer's package insert are the optimal values for classifying a sample from a patient with syphilis infection.

The cutoff value $\geq$0.7 had the lowest distance to corner value (0.1385) of the range of cutoff values listed. However, the Youden Index is the recommended index for determining an optimal cutoff and the cutoff value $\geq$0.8 had the highest Youden Index (0.8265) of all values listed. This

Table 48

Syphilis IgG S/CO Cutoff Values, TPR, TNR, PPV, Accuracy, Youden Index and Distance to Corner

Cutoff Value	TPR (Sens)	TNR (Spec)	PPV	Accuracy	TPR+TNR	Youden Index	Distance to Corner
≥0.0	1.0000	0.0000	0.2107	0.2107	1.0000	0.0000	1.0000
≥0.1	1.0000	0.0031	0.2112	0.2131	1.0031	0.0031	0.9969
≥0.2	0.9919	0.1594	0.2395	0.3348	1.1513	0.1513	0.8406
≥0.3	0.9790	0.3409	0.2839	0.4753	1.3199	0.3199	0.6594
≥0.4	0.9639	0.4226	0.3083	0.5367	1.3865	0.3865	0.5785
≥0.5	0.9464	0.6280	0.4045	0.6951	1.5745	0.5745	0.3758
≥0.6	0.9104	0.8294	0.5875	0.8465	1.7398	0.7398	0.1927
≥0.7	0.8882	0.9183	0.7437	0.9119	1.8065	0.8065	0.1385
≥0.8	0.8638	0.9627	0.8608	0.9419	1.8265	0.8265	0.1412
≥0.9	0.8231	0.9869	0.9439	0.9524	1.8100	0.8100	0.1774
≥1.0	0.8149	0.9922	0.9650	0.9549	1.8071	0.8071	0.1853
≥1.1	0.7986	0.9969	0.9856	0.9551	1.7955	0.7955	0.2014
≥1.2	0.7579	0.9981	0.9909	0.9475	1.7560	0.7560	0.2421
≥1.3	0.7090	0.9988	0.9935	0.9377	1.7077	0.7077	0.2910
≥1.4	0.6729	0.9988	0.9931	0.9301	1.6716	0.6716	0.3271
≥1.5	0.6321	0.9988	0.9927	0.9215	1.6309	0.6309	0.3679
≥1.6	0.6007	0.9988	0.9923	0.9149	1.5995	0.5995	0.3993
≥1.7	0.5611	0.9991	0.9938	0.9068	1.5602	0.5602	0.4389
≥1.8	0.5320	0.9991	0.9935	0.9007	1.5311	0.5311	0.4680
≥1.9	0.4889	0.9994	0.9953	0.8918	1.4883	0.4883	0.5111
≥2.0	0.4459	0.9994	0.9948	0.8828	1.4452	0.4452	0.5541
≥2.1	0.4109	0.9994	0.9944	0.8754	1.4103	0.4103	0.5891
≥2.2	0.3830	0.9994	0.9940	0.8695	1.3824	0.3824	0.6170
≥2.3	0.3609	0.9994	0.9936	0.8649	1.3603	0.3603	0.6391
≥2.4	0.3353	0.9994	0.9931	0.8595	1.3347	0.3347	0.6647
≥2.5	0.3050	1.0000	1.0000	0.8536	1.3050	0.3050	0.6950
≥3.0	0.2002	1.0000	1.0000	0.8315	1.2002	0.2002	0.7998
≥3.5	0.1106	1.0000	1.0000	0.8126	1.1106	0.1106	0.8894
≥4.0	0.0536	1.0000	1.0000	0.8006	1.0536	0.0536	0.9464
≥4.5	0.0175	1.0000	1.0000	0.7930	1.0175	0.0175	0.9825
≥5.0	0.0058	1.0000	1.0000	0.7905	1.0058	0.0058	0.9942
≥5.4	0.0012	1.0000	1.0000	0.7896	1.0012	0.0012	0.9988
≥5.5	0.0000	1.0000		0.7893	1.0000	1.0000	1.0000

Note: TPR = true positive rate; TNR = true negative rate; PPV = positive predictive value

cutoff value was lower than the 0.9 S/CO value recommended by the manufacturer's package insert for an equivocal result; and would require an extensive off-label method validation if put into clinical use. The Youden Index for cutoff value ≥0.9 was 0.8100 which was the next highest value and was the manufacturer's cut off value for an equivocal test result. The manufacturer's cutoff value for a reactive test result was ≥1.1; and it provided the highest accuracy (proportion properly classified) value at 0.9551 with a Youden Index of 0.7955. Cutoff values equal to and above 2.5 showed 100% specificity (true negative rate) and PPV for syphilis infection.

The RPR titer cutoff values shown in Table 49 ranged from 0 (nonreactive) to 1024 (highest reactive titer in study samples). The ≥1 cutoff value represented a 1:1 titer dilution which was equivalent to the reactive qualitative RPR test result (undiluted). This cutoff value had the highest Youden Index (0.6266), the lowest Distance to Corner (0.3600) and the highest accuracy (0.9134) of all values listed. Cutoff values equal to and above 64 showed 100% specificity (true negative rate) and PPV for syphilis infection.

Table 49

RPR Titer Cutoff Values, TPR, TNR, PPV, Accuracy, Youden Index and Distance to Corner

Cutoff Value	TPR (Sens)	TNR (Spec)	PPV	Accuracy	TPR+TNR	Youden Index	Distance to Corner
≥0	1.0000	0.0000	0.2107	0.2107	1.0000	0.0000	1.0000
≥1	0.6403	0.9863	0.9259	0.9134	1.6266	0.6266	0.3600
≥2	0.5495	0.9907	0.9402	0.8977	1.5402	0.5402	0.4506
≥4	0.4633	0.9935	0.9499	0.8818	1.4568	0.4568	0.5367
≥8	0.4016	0.9972	0.9746	0.8717	1.3988	0.3988	0.5984
≥16	0.3341	0.9988	0.9863	0.8587	1.3329	0.3329	0.6659
≥32	0.2701	0.9997	0.9957	0.8460	1.2698	0.2698	0.7299
≥64	0.1851	1.0000	1.0000	0.8283	1.1851	0.1851	0.8149
≥128	0.0978	1.0000	1.0000	0.8099	1.0978	0.0978	0.9022
≥256	0.0373	1.0000	1.0000	0.7972	1.0373	0.0373	0.9627
≥512	0.0105	1.0000	1.0000	0.7915	1.0105	0.0105	0.9895
≥1024	0.0058	1.0000	1.0000	0.7905	1.0058	0.0058	0.9942

Note: TPR = true positive rate; TNR = true negative rate; PPV = positive predictive value

Third null hypothesis. Among patients seeking STD services at a large western region metropolitan public health clinic, there will be no difference between the syphilis traditional and reverse algorithm interpretations.

NCSS11 software was used to calculate the sensitivity, specificity, predictive value, likelihood ratio, and accuracy measures for each syphilis algorithm using the diagnosis of syphilis infection yes or no as the gold standard. A diagnosis of syphilis infection included both past and present infection. Contingency Tables 50 and 51 provide supporting statistical data for determining the usefulness of the reverse algorithm for detecting a syphilis infection.

The column contingency table calculations are based on the true condition (syphilis infection). These calculations include sensitivity and specificity which are reported with 95% confidence intervals (CI) for each algorithm in Table 52. Sensitivity and specificity calculations are not influenced by population disease prevalence. The traditional algorithm exhibited a lower sensitivity (64%) than the reverse algorithm which had a sensitivity of 82%. The specificity for both algorithms was 99%. An ideal test would have 100% sensitivity and specificity.

Table 50

Traditional Algorithm Contingency Table (N = 4,077)

	Syphilis infection Yes	**Syphilis infection No**	**Total**	**Measures**
Likely syphilis Traditional algorithm	TP 550	FP 44	594	93% PPV
Unlikely syphilis Traditional algorithm	FN 309	TN 3174	3483	91% NPV
Total	859	3218	4077	
Measures	64% sensitivity 46.83 +LR	99% specificity 0.36 -LR		91% accuracy 21% prevalence

Note: TP = true positive; FP = false positive; FN = false negative; TN = true negative; +LR = positive likelihood ratio; -LR = negative likelihood ratio; PPV = positive predictive value; NPV = negative predictive value

Table 51

Reverse Algorithm Contingency Table (N = 4,077)

	Syphilis infection Yes	Syphilis infection No	Total	Measures
Likely syphilis Reverse algorithm	TP 707	FP 42	749	94% PPV
Unlikely syphilis Reverse algorithm	FN 152	TN 3176	3328	95% NPV
Total	859	3218	4077	
Measures	82% sensitivity 63.06 +LR	99% specificity 0.18 -LR		95% accuracy 21% prevalence

Note: TP = true positive; FP = false positive; FN = false negative; TN = true negative; +LR = positive likelihood ratio; -LR = negative likelihood ratio; PPV = positive predictive value; NPV = negative predictive value

Table 52

Traditional and Reverse Algorithm Sensitivity, Specificity, 95% CI

Algorithm	Sensitivity	95% CI	Specificity	95% CI
Traditional	0.64	0.61-0.67	0.99	0.98-0.99
Reverse	0.82	0.80-0.85	0.99	0.98-0.99

Note: CI = Confidence interval; Proportion Confidence Interval Method: Exact Binomial

The row contingency table calculations are based on the predicted condition of syphilis infection and the results are influenced by the disease prevalence within the population. Table 53 lists the row calculations of PPV and NPV along with 95% CI for each algorithm.

Table 53

Traditional and Reverse Algorithm PPV, NPV, 95% CI, Adjusted PPV and NPV

Algorithm	PPV	95% CI	NPV	95% CI	Adj PPV 3% prevalence	Adj NPV 3% prevalence
Traditional	0.93	0.90-0.95	0.91	0.90-0.92	0.59	0.99
Reverse	0.94	0.92-0.96	0.95	0.95-0.96	0.66	0.99

Note: PPV = positive predictive value; NPV = negative predictive value; CI = Confidence interval; Adj PPV = known prevalence (3%) adjusted positive predictive value; Adj NPV = known prevalence (3%) adjusted negative predictive value; Proportion Confidence Interval Method: Exact Binomial

The traditional algorithm exhibited slightly lower PPV (93%) and NPV (91%) than the reverse algorithm which had a PPV of 94% and NPV of 95%. The PPV and NPV calculations with 95% CI were based on the estimated prevalence of 21% calculated from the study population. The PPV and NPV were adjusted based on the 3% prevalence calculated for the STD clinic population. The adjusted value calculations do not provide 95% CI range. The traditional algorithm adjusted PPV is much lower (59%) than the unadjusted PPV (93%) and the adjusted NPV is higher (99%) than the unadjusted NPV (91%). The reverse algorithm adjusted PPV was also much lower (66%) than the unadjusted PPV (94%); however, it is still higher than the traditional algorithm adjusted PPV (59%). The reverse algorithm adjusted NPV increased to 99% and is the same as the traditional algorithm. An ideal test would have 100% PPV and NPV.

Other calculations were based on values present in the whole table and included likelihood ratios, accuracy, and prevalence which are reported with 95% CI for each algorithm in Table 54. Likelihood ratios and accuracy are not influenced by the disease prevalence in the population and therefore the values are not adjusted for the STD clinic prevalence. The 21% prevalence reported in the Table 54 was based on the study population of 4,077 patients. The traditional algorithm exhibited a lower positive LR (46.83) than the reverse algorithm (63.06); however, they are both greater than 10 which provides strong evidence for ruling in a disease diagnosis (Table 5). The traditional algorithm exhibited a higher negative LR (0.36) than the reverse algorithm (0.18). According to Table 5, a negative LR between 0.1 and 0.2 provides moderate evidence for ruling out a diagnosis and a negative LR between 0.2 and 0.5 provides weak evidence for ruling out a diagnosis. The accuracy of the traditional algorithm (91%) was less than the reverse algorithm accuracy (95%). An ideal test would exhibit 100% accuracy.

Table 54

Traditional and Reverse Algorithm Likelihood Ratios, 95% CI, Accuracy, and Study Prevalence

Test	Positive Likelihood Ratio	95% CI	Negative Likelihood Ratio	95% CI	Accuracy	Prevalence (Study)
Traditional algorithm	46.83	34.77-63.07	0.36	0.33-0.40	0.91	0.21
Reverse algorithm	63.06	46.62-85.30	0.18	0.16-0.21	0.95	0.21

Note: CI = Confidence interval; Ratio Confidence Interval Method: Katz Logarithm

Hypothesis testing of the third null hypothesis was performed using the McNemar test to determine if there was a difference between the traditional and reverse algorithms for interpreting a diagnosis of syphilis. Frequency Table 55 was used to input data into NCSS 11. The McNemar test was statistically significant, $p < 0.0000$ ($\alpha = 0.05$), indicating that there was a difference between the two algorithms; therefore, the null hypothesis was rejected (Table 56).

Table 55

McNemar Test Traditional and Reverse Algorithm Frequency Table (N = 4,077)

	Likely syphilis Traditional algorithm	Unlikely syphilis Traditional algorithm	Total	Marginal Proportion $P_1 = (A+B)/N$
Likely syphilis Reverse algorithm[a]	499 A	217 B	716	P_1=0.1756 (716/4077)
Unlikely syphilis Reverse algorithm	52 C	3309 D	3361	
Total	551	3526	4077	
Marginal Proportion $P_2 = (A+C)/N$	P_2=0.1351 (551/4077)			

Note: Marginal Proportions P_1 and P_2 represent marginal homogeneity or probability that the row and column frequencies are equal.

[a]Equivocal results included as reactive

Table 56

McNemar Two-sided Hypothesis Test of the Difference Between Traditional and Reverse Algorithms (N = 4,077)

H_0: P_1-P_2 = 0 vs H_a: P_1-$P_2 \neq 0$

Statistical Test	Proportion Difference P_1-P_2	95% Confidence interval (CI)	Confidence Interval Width	Test Statistic Value	Probability level $\alpha = 0.05$
McNemar	0.0405	0.0327 to 0.0484	0.0156	101.208	0.0000

Note: Proportion (risk) difference ($\delta = P_1 - P_2$) is a direct method of comparison between two event probabilities.

Summary

Descriptive study variables revealed that males represented the highest proportion (68.56%) of the study population of 4,077 patients and the population mean age was 33 years old. The majority of patients (42.6%) were asymptomatic when they visited the STD clinic; and most patients (61.7%) were not diagnosed with an STD. Of the patients diagnosed with an STD the majority (21.1%) were diagnosed with syphilis. The 859 patients diagnosed with syphilis were also staged by the STD clinic staff with the majority of patients (60.2%) staged as latent (early, late or unknown) syphilis.

Statistical analysis of the observed variables was performed with NCSS 11 software. All three null hypotheses were rejected with statistically significant *p* values, thus providing statistical support that there was a difference between the Captia™ Syphilis IgG and Fujirebio Serodia TP-PA test result interpretations; between the Captia™ Syphilis IgG S/CO and MacroVue RPR titer test results; and between the reverse and traditional syphilis interpretation algorithms (Table 57).

As shown in Table 58, analysis of the 2x2 contingency table for the Captia™ Syphilis IgG and Serodia TP-PA test results revealed that the TP-PA had a higher sensitivity (92%) and specificity (100%) than the Syphilis IgG (82% and 99%, respectively). The TP-PA test also had

Table 57

Summary Null Hypotheses Statistical Tests

Null Hypothesis Number	There will be no difference	Statistical test	Probability $\alpha = 0.05$	Statistically significant?	Null hypothesis?
$H1_o$	between Captia™ Syphilis IgG and Fujirebio Serodia TP-PA test results	McNemar Two sided	0.0002	Yes	Rejected
$H2_o$	in diagnostic interpretation of Captia™ Syphilis IgG S/CO value and Becton Dickinson MacroVue RPR titer result	*z* test Two sided	0.0000	Yes	Rejected
$H3_o$	between Syphilis traditional and reverse algorithm interpretations	McNemar Two sided	0.0000	Yes	Rejected

Table 58

Summary Usefulness Statistics for Syphilis IgG, TP-PA, and Algorithms

Test	Sens	Spec	Adj PPV[b]	Adj NPV[b]	+LR	-LR	Accuracy
Captia™ Syphilis IgG	0.82[a]	0.99	0.66	0.99	63.06	0.18	0.95
Serodia TP-PA	0.92	1.00	0.91	1.00	330.50	0.08	0.98
Traditional algorithm	0.64	0.99	0.59	0.99	46.83	0.36	0.91
Reverse algorithm	0.82[a]	0.99	0.66	0.99	63.06	0.18	0.95

Note: Sens = sensitivity; Spec = specificity; Adj PPV = Adjusted positive predictive value; Adj NPV = Adjusted negative predictive value; +LR = positive likelihood ratio; -LR = negative likelihood ratio
[a]Captia™ Syphilis IgG sensitivity based on equivocal test results counted as reactive
[b]Adjusted PPV and NPV based on 3% syphilis prevalence within STD clinic population

a higher positive LR (330.50) and lower negative LR (0.08) than the Syphilis IgG (63.06 and 0.18, respectively). The population based adjusted PPV and NPV for TP-PA was higher than

Syphilis IgG (0.91 and 1.00 versus 0.66 and 0.99, respectively). Finally, based on the 2x2 contingency table, the TP-PA had higher accuracy (98%) than the Syphilis IgG (95%).

Analysis of the 2x2 contingency table for the traditional and reverse algorithms (Table 58) revealed that the reverse algorithm had a higher sensitivity (82%) than the traditional (64%), while both algorithms had the same 99% specificity. The reverse algorithm also had a higher positive LR (63.06) and lower negative LR (0.18) than the traditional (46.83 and 0.36, respectively). The adjusted PPV for the reverse algorithm (0.66) was higher than the traditional (0.59), while both algorithms had the same 0.99 adjusted NPV. Based on the 2x2 contingency table, the reverse algorithm had higher accuracy (95%) than the traditional algorithm (91%).

A statistical comparison of AUC values for the continuous variables, Syphilis IgG S/CO and RPR titer revealed that the Syphilis IgG AUC (0.9500) was higher than the RPR titer (0.8155) indicating greater diagnostic accuracy. Statistical analysis of the Syphilis IgG S/CO values to determine an optimal cutoff value showed that the cutoff value of $\geq$0.8 had the highest Youden Index (0.8265); however, that cutoff value was not listed in the manufacturer's package insert for an equivocal test result. The manufacturer's cutoff value for an equivocal test result was between 0.9 and 1.0 and these values had the next highest Youden Indexes along with high diagnostic accuracy (0.9524 and 0.9549, respectively). The manufacturer's cutoff value for a reactive test result ($\geq$1.1) had the highest diagnostic accuracy (0.9551) indicating that this value along with the equivocal cutoff values properly classified a higher proportion of syphilis infected patients. These results indicate that the cutoff values listed in the manufacturer's package insert are the optimal values for reporting patient test results.

Chapter Five: Discussion

This chapter will discuss the study results, limitations, conclusions, and implications for future studies. The purpose of this study was to determine the usefulness of a specific automated test method, the Captia™ Syphilis IgG EIA, and the syphilis reverse algorithm interpretation for detecting syphilis infection among patients seeking care at public health clinics. The target construct was syphilis infection and the proposition for the research question was that more useful syphilis testing will improve syphilis detection. The process (syphilis test results and S/CO value) and outcome (traditional and reverse algorithm interpretations) domains of the Donabedian Quality Framework provided the study theoretical framework. The specific research question was: What is the usefulness of the Captia™ Syphilis IgG EIA test method and the reverse algorithm for detection of syphilis infection in a public health population? There were three null hypotheses associated with the research question which were all rejected by statistical analysis of the study data.

Syphilis can be a difficult disease to diagnose due to its progression to an asymptomatic stage when left untreated. Additionally, current syphilis diagnostic tests are serologically based and require an adequate immune response from the patient, which may not be sufficient in HIV infected or immunocompromised patients. The Becton Dickinson MacroVue RPR test is a nontreponemal serological test used for initial syphilis screening within the traditional algorithm. Because the test can cross react with many different non syphilis conditions giving a false positive test result, the CDC recommendation for screening with the RPR test is to verify an

initial reactive test result with a treponemal test. If the follow up treponemal test is discordant (nonreactive), then clinical correlation and/or additional testing at a later date is recommended. The Captia™ Syphilis IgG EIA test is a treponemal serological test used for initial syphilis screening within the reverse algorithm. Because a treponemal test cannot distinguish between past and current syphilis infection due to the lifelong presence of treponemal IgG antibodies, the CDC recommendation for screening with a treponemal test is to follow up a reactive test result with a nontreponemal test. If the nontreponemal test is discordant (nonreactive), then CDC recommends performance of a second more sensitive treponemal test to act as a tie breaker. If the second treponemal test is discordant (nonreactive), then clinical correlation and/or testing at a later date is recommended.

Results

Public health population. This study was the first to determine that the Captia™ Syphilis IgG test result, the S/CO value, and reverse algorithm interpretation are useful diagnostic predictors of syphilis infection among public health clinic patients. Syphilis has serious public health consequences and is considered a sentinel public health event by CDC. Public health programs at the local and state level usually include both the clinical and laboratory services necessary to perform syphilis detection, treatment, and prevention. Public health responsibilities related to STD prevention include appropriate laboratory testing, correct disease staging, appropriate treatment, public health disease reporting, and participation in effective partner services (CDC, 2006). Public health clinics provide a two phase disease approach which includes detection and treatment of multiple STD infections combined with effective partner notification and contact treatment.

The study setting for this dissertation was a public health STD clinic located in a large western metropolitan area with a population of over two million residents. The clinic was operated by a public health agency and participated in federally approved syphilis detection and surveillance programs. The STD clinic provided services to persons who had signs and symptoms of syphilis infection as well as those who were contacts to infected individuals. All syphilis tests ordered by the STD clinic staff were performed at the public health laboratory, which was also owned by the public health agency. All clinic staff were trained to perform syphilis infection clinical assessment following standard CDC STD guidelines. The STD clinic received federal funding through the CDC STD prevention cooperative grant, which seeks to reduce syphilis infection especially among high-risk populations, such as MSM. The clinic staff followed federal guidelines for syphilis detection and treatment. In addition to patients who had signs and symptoms of syphilis infection, the STD clinic also targeted pregnant women, MSM, persons with HIV, and contacts of known syphilis cases for conducting syphilis testing. Because syphilis infection can occur concurrently with other STDs, clinic patients were often screened for multiple STDs (HIV, syphilis, gonorrhea, and chlamydia) during the same clinic visit. All syphilis testing ordered by the STD clinic staff was performed at the public health agency public health laboratory. All syphilis reflex testing performed by the laboratory staff was approved by the laboratory director.

The dissertation study demographic variables closely matched those identified by Pathela, et al (2015) who performed a retrospective study of 40 geographically diverse US public health STD clinics. They determined that 61.9% of STD clinic patients were male; 47.1% were between 20-29 years old; and 35% were diagnosed with an STD during their clinic visit. The dissertation study demographic variables were very similar with males representing 68.6% of the

study population; 44.3% of the study population were between 21-30 years old; and 38.3% were diagnosed with an STD; thus, reflecting a "typical" public health STD population and enhancing the generalizability of the study results to other public health clinics.

The higher rates of STD infections in public health clinics provide an excellent environment for performing STD diagnostic test evaluation. Linnet et al. (2012) suggest that diagnostic test evaluation should be carried out in a population suspected of having the target disease, preferably with study participants chosen based on predefined symptoms and/or signs of the disease. All participants should be independently tested with both the gold standard reference and the index test with laboratory personnel blinded to the results of either test. Because laboratory test results can vary due to disease progression, it is important to correctly classify each participant as diseased or nondiseased and include the disease stage if possible. The dissertation study patients included 859 (21.1%) who were diagnosed with syphilis infection and all cases were staged according to the 2012 CDC syphilis case definitions. All syphilis stages, except tertiary, were represented within the study population.

The public health laboratories associated with STD clinics are uniquely positioned to provide assistance with STD screening algorithms; information on test reliability and validity; suggestions for additional testing that could be used to resolve discordant results; and to act as a conduit to subject matter expertise at the CDC. The western region public health laboratory where the study was performed focused its clinical testing on STDs that were of particular interest to the public health clinic. This STD focus contrasts with the broad testing menu established at hospital or reference clinical laboratories. The focus of public health laboratory testing is disease detection, surveillance, and response. Public health laboratory professionals

have the expertise to determine which STD tests will provide useful test results when ordered in a public health setting (APHL, 2017).

Recently the perception of laboratory services as a "commodity" has been discussed by various authors. This study provides justification that laboratory services are not just a commodity which is defined by Plebani and Lippi (2010) as "a mass produced unspecialized products that should therefore be evaluated on the simple basis of cost, taking quality for granted" (p. 940). The laboratory is often thought of as a "black box" with samples entering and results coming out. There is limited understanding or recognition among healthcare professionals outside the laboratory of the complexities of ensuring that each test method was selected appropriately, each sample was collected and handled properly, each result was accurately reported and interpreted before being properly communicated to the ordering entity. Plebani (1999) discusses the role of laboratory medicine as a medical discipline rather than a technical discipline and states that "the ultimate value of a laboratory test depends not only on its chemical or clinical performance characteristics, but on its impact on patient management, the only true assessment of the quality of testing being quality of patient outcomes" (p. 36). The American Society for Clinical Laboratory Science (ASCLS) published a position paper in 2005 on the value of clinical laboratory services in health care. The paper emphasized the impact of laboratory services on improving the quality of patient care. Laboratory professionals have the skills and knowledge to provide guidance on maintaining quality within not only the analytical, but the preanalytical and postanalytical phases of the QMS path of laboratory workflow.

In order to keep up with increasing workload and a shrinking workforce, clinical laboratories and some public health laboratories have become extensively automated. While automation can maximize productivity, and reduce cost per test in high volume settings, quality must not be

taken for granted. Focusing only on reduction in testing cost ignores the value that laboratory professionals can contribute toward improving patient care by using their knowledge of test performance and laboratory workflow. As seen in the cost analyses performed in Tables 10-12, labor is a significant portion of any test cost calculation. Automation may free up additional time for laboratory professionals to participate on multidisciplinary teams to ensure that laboratory tests are ordered, performed, and interpreted appropriately. Laboratory professional guidance will ultimately impact the quality of laboratory testing and, therefore, patient outcomes (Beasley, 2016; Plebani, 1999; Plebani, 2002; Plebani, Laposata, & Lundberg, 2011).

Public health laboratories are an example of the importance of communication with the clinical entities ordering laboratory tests. The western region public health laboratory where the dissertation study was performed communicated on a regular basis with public health clinic staff. This communication was especially necessary when discordant syphilis test results were identified by clinic staff. Discussions with clinic and laboratory staff resulted in implementation of the modified testing algorithms and reflex testing in use during the study. Continued dialogue ensured that clinic staff recognized the quality of the laboratory testing performed and the concern of laboratory staff for providing accurate test results for each patient; thus, the laboratory staff acted as liaisons with the clinic staff. Plebani (2002) recognized that clinical liaisons may provide a new role for laboratorians. These liaisons can: a) ensure laboratory test quality whether in the central laboratory or at the patient's bedside; b) improve pre- and post-analytical quality; c) evaluate interpretation and utilization of laboratory data for improved patient management; and d) participate in multidisciplinary teams for joint research projects along with providing clinical consulation in the patient management process.

Expanding the role of laboratory professionals will allow for a mechanism to assess not only laboratory efficiency, but also effectiveness and clinical outcomes. Plebani (2002) suggests that surrogate markers for measuring effectiveness could include: a) inclusion of interpretation aids on test reports, b) reporting decision limits along with the type of action to be taken at various limits and the source of each limit, and c) allowable levels of imprecision and reproducibility of laboratory results. This dissertation study provided suggestions for including syphilis test sensitivity, specificity, and likelihood ratios on laboratory test reports to allow healthcare providers to utilize Bayes nomogram for result interpretation. Additionally, the Syphilis IgG S/CO values or a "trigger" value at a specified limit could be reported for nonreactive test results to alert healthcare providers that additional follow-up testing in a few weeks may be warranted. Implementation of these suggestions will require that laboratory professionals proactively ensure that they have a "seat at the table" in multidisciplinary discussions focused on improving patient care.

Usefulness measures. As discussed in Chapter Two Literature Review, multiple authors have suggested a variety of statistical measure combinations to evaluate the usefulness of a laboratory diagnostic test. Based on the published research discussed in Chapter Two, the following statistical measures were used in this study to define test usefulness: sensitivity, specificity, predictive values, likelihood ratios, accuracy, and/or ROC analysis. Because test feasibility is also a usefulness measure, test cost was included in the discussion (Aamir & Hamilton, 2014; Galen & Gambino, 1975; Hulley et al., 2013; Vihinen, 2012).

Galen and Gambino (1975) discuss the attributes for determining the preferable statistical measure of a diagnostic test (Table 9). They recommend a test with high sensitivity for detecting a treatable, serious disease that should not be missed and for which false positive results do not

lead to serious financial or emotional consequences. A test with high accuracy is recommended for detecting a treatable, serious disease with serious consequences for both false negative and false positive test results. For public health purposes, a highly sensitive test that can determine syphilis infection in an early stage of disease has treatment implications not only for the infected patient but also for stopping the spread of infection within the community; however, there can be psychological and financial consequences of misdiagnosis due to either false positive or false negative test results. Therefore, use of a highly sensitive and highly accurate test for initial syphilis screening would be desirable for those patients seeking care at STD public health clinics. Additionally, because the likelihood ratio statistic is not affected by disease prevalence within the tested population, a test that provides a high positive LR and a low negative LR would also be preferable for a screening test.

Determining the financial or emotional consequences of false positive or false negative test results is difficult. Chesson, Collins, and Koski (2008) developed a series of formulas that could be used to estimate direct medical costs and indirect costs (lost productivity) averted by sexually transmitted infection (STI) programs within the US. The authors focused on the benefits of treating people with STIs, including P&S syphilis. "These benefits included the sequelae costs averted by treatment of people with STIs, the prevention of congenital syphilis in infants born to mothers treated for P&S syphilis, the interruption of STI transmission in the population, the reduction of STI-attributable HIV infections (HIV infections that would not have occurred without the facilitative effects of STIs on HIV transmission and acquisition)" (p. 2). The authors based all medical costs on 2006 dollars, with the average direct medical cost per case of P&S syphilis averted calculated at $572 and $112 for indirect (lost productivity) costs. Included in the averted cost per case was the possibility of neurosyphilis and cardiovascular syphilis in untreated

cases, and also the possibility of antibiotic treatment prior to symptom onset. For congenital syphilis, the authors assumed that 50% of pregnant women with untreated P&S would deliver a child with congenital syphilis. They did not include a cost for premature births and stillbirths associated with congenital syphilis and they state that the averted cost is most likely understated. The average direct medical cost per case of congenital syphilis averted was calculated as $6,738. The indirect cost for congenital syphilis was calculated at $3,369 for the first year. The authors assumed that there would be lifelong impacts of the infection and that the indirect cost would be incurred every year for 25 years for a total indirect cost of $60,421. These formulas were developed for STI program managers to use to determine the economic impact of their program activities on detection and prevention of disease. The formulas also provide an estimate of the impact of a false negative syphilis test result. A missed syphilis case could result in a total direct and indirect cost of $684; however, that total cost would increase 100-fold to $70,528 if the false negative was reported on a pregnant woman who gave birth to a child with congenital syphilis. In those cases, a false negative has serious consequences, thus reinforcing the need for use of highly accurate syphilis tests especially among pregnant females.

The Syphilis IgG sensitivity (82%) was lower than the manufacturer's package insert sensitivity value of 98.2% (165/168). The manufacturer's package insert sensitivity was calculated based on analysis of 168 samples taken from patients at various stages of syphilis infection. Diagnosis was based on clinical history combined with serological data from VDRL and FTA-ABS tests. The package insert also lists a sensitivity of 98.4% and specificity of 99.3% for Captia™ Syphilis IgG testing performed on a panel of 1,321 serum specimens that had tested reactive or nonreactive by two different syphilis serological tests (Trinity Biotech, 2003). There are few published studies that compare the Captia™ Syphilis IgG EIA test method to either

syphilis diagnosis or to another treponemal test. Castro et al (2013) compared the Captia™ Syphilis IgG and the Serodia TP-PA using samples known to be TP-PA reactive or nonreactive. In that study, the Syphilis IgG EIA method showed 100% sensitivity (109/109) and 97% specificity (176/181) when compared to TP-PA. There are numerous published studies listed in Table 4 that show a wide variation in different Syphilis EIA test method sensitivities depending on the stage of infection, disease prevalence in the population tested, and the standard used for comparison in the published study (Loeffelholz & Binnicker. 2012, Manavi, et al, 2006, Peeling & Ye, 2003, Sena, et al, 2010, Smith et al, 2013, Wicher, et al, 1999). The Syphilis EIA sensitivity for primary syphilis as listed in Table 4 ranged from 57% to 100%, secondary syphilis ranged from 97% to 100% and latent syphilis ranged from 75% to 100%. The implication for the study data is that early and late stages of syphilis infection can produce widely variable results. In this study, 60% of the samples tested were diagnosed with latent syphilis which is an asymptomatic stage and difficult to diagnose. The large proportion of latent samples may be a contributing factor to the lower sensitivity results observed with the Captia™ Syphilis IgG.

This study was the first to determine the sensitivity of the Captia™ Syphilis IgG EIA in a public health population using the presence or absence of syphilis infection as the gold standard. Prior studies had used another treponemal or nontreponemal test as the gold standard rather than diagnosis. While the TP-PA showed a higher sensitivity (92%) than the Syphilis IgG (82%) in this study, the CDC (2011) does not recommend the use of TP-PA as an initial screening test within the reverse screening algorithm. The TP-PA test is not a clinically practical screening test choice as it cannot be automated and, as shown in Tables 12 and 13, has a cost (direct = $4.41; fully loaded = $15.43) which is almost triple that of either the RPR (direct = $1.76; fully loaded = $6.16) or Syphilis IgG (direct =$1.85; fully loaded = $6.47) screening tests. The CDC (2011)

and various authors (Castro et al, 2013; Zhang et al, 2012) recommend that a more sensitive treponemal test, such as the TP-PA, should be used as the confirmatory test for discordant syphilis results within the reverse algorithm. This study confirmed that the western regional laboratory where the study was performed followed CDC reverse algorithm recommendations by using a more sensitive confirmatory test (TP-PA) for discordant test results.

The Captia™ Syphilis IgG test does not detect IgM antibodies, and the lower than expected sensitivity results in the study population could be due to the presence of syphilis IgM antibodies in the samples tested. The Serodia TP-PA method detects both syphilis IgG and IgM antibodies. Sensitivity was calculated as the number of true positive samples divided by the number of true positive and false negative samples. As shown in Table 59, there were 152 false negative Syphilis IgG results reported (S/CO values 0.0 to 0.8) and 124 of the samples were confirmed reactive by TP-PA testing indicating the presence of either IgG and/or IgM syphilis antibodies at a level undetectable by the Captia™ Syphilis IgG test method.

A review of the clinical data for the 124 samples revealed that 100% (124/124) of the samples had a diagnosis of syphilis infection past or current. Of the 124 TP-PA reactive samples, 56% (69/124) were diagnosed as late latent; 18% (22/124) were diagnosed with other syphilis stages; 15% (19/124) had a history of syphilis infection; and 11% (14/124) had a positive syphilis test from an outside laboratory. Possible explanations for the false negative Syphilis IgG test results include: a) diagnosis of late latent syphilis can be difficult due to the lack of symptoms; b) patients with a history of syphilis infection may have IgG levels below the detectable range of the Captia™ test method; or c) use of a positive syphilis test from an outside laboratory may have influenced the clinic staff in their staging decision.

Table 59

Syphilis IgG S/CO Frequency Counts, FN, FP, Discordants, and TP-PA Confirmation

S/CO	Syphilis IgG Result	# samples	# FN Syphilis IgG	#FP Syphilis IgG	Discordant +Syphilis IgG[a]	Discordant +RPR[b]	TP-PA confirmed
≤0.1	Nonreactive	520	7			17	0
0.2	Nonreactive	595	11			14	6
0.3	Nonreactive	276	13			11	10
0.4	Nonreactive	676	15			8	13
0.5	Nonreactive	679	31			13	27
0.6	Nonreactive	305	19			10	17
0.7	Nonreactive	164	21			8	20
0.8	Nonreactive	113	35			9	31
0.9	Equivocal	24		17	21		5
1.0	Equivocal	30		16	21		7
1.1	Reactive	39		4	29		22
1.2	Reactive	44		2	29		21
1.3	Reactive	31		0	18		14
1.4	Reactive	35		0	20		17
1.5	Reactive	27		0	8		8
1.6	Reactive	35		1	16		14
1.7	Reactive	25		0	8		6
1.8	Reactive	38		1	17		15
1.9	Reactive	37		0	15		13
2.0	Reactive	30		0	5		4
2.1	Reactive	24		0	8		7
2.2	Reactive	19		0	2		2
2.3	Reactive	22		0	5		5
2.4	Reactive	28		2	9		8
2.5	Reactive	19		0	3		3
2.6	Reactive	16		0	3		3
2.7	Reactive	22		0	0		0
2.8	Reactive	21		0	2		2
2.9	Reactive	12		0	0		0
3.0	Reactive	11		0	0		0
3.1	Reactive	20		0	0		0
3.2	Reactive	24		0	1		1
3.3	Reactive	13		0	1		0
3.4	Reactive	8		0	1		1
3.5	Reactive	19		0	2		2
3.6	Reactive	5		0	0		0
3.7	Reactive	11		0	0		0
3.8	Reactive	5		0	1		1
3.9	Reactive	9		0	0		0
4.0	Reactive	7		0	1		1
≥4.1[c]	Reactive	39		0	0		0
Totals		4077	152	43	246	90	306

[a]Discordant + Syphilis IgG = Reactive Syphilis IgG and nonreactive RPR; [b]Discordant + RPR = Reactive RPR and nonreactive Syphilis IgG; [c]S/CO counts between 4.1 and 5.5 were combined in one line to conserve space.

The 124 false negative Syphilis IgG samples that were confirmed reactive by TP-PA testing contributed to the lower sensitivity observed with the Captia^{TM} Syphilis IgG test. This brings up the question of determining the optimum cutoff value for an EIA test value. What would be the impact of decreasing the Syphilis IgG cutoff value from the FDA approved value on the Syphilis IgG test result interpretation for the 124 false negative samples?

Based on the study data, it appears that the manufacturer's cut off values for both reactive and equivocal results provided a highly accurate test result interpretation. As discussed in Chapter Four results section, the manufacturer's package insert S/CO cutoff for an equivocal Syphilis IgG test result was between 0.9 and 1.1, and any change in this cutoff value would require an extensive off label method verification.

According to Perkins and Schisterman (2006), the Youden Index is the preferred method for determining an optimal cutoff value. As shown in Table 48, the S/CO cutoff value $\geq$0.8 had the highest Youden Index value (0.8265). In order to determine if the S/CO cutoff value $\geq$0.8 was an optimal cutoff, the number of samples at each cutoff value were counted and entered into Table 59. There were 113 samples at the $\geq$0.8 cutoff value, and 35 of the samples were classified as nonreactive according to the manufacturer's cutoff value and had a diagnosis of syphilis infection past or current; therefore, they would be considered a false negative. The remaining 78 samples in this cutoff category were all RPR and TP-PA nonreactive and were not diagnosed with syphilis; therefore, they would be considered to be a true negative. The diagnostic characteristics of the 35 false negative samples are shown in Table 60.

While all 35 samples had a diagnosis of syphilis infection, 74% (26/35) had a nonreactive RPR test and would not have been detected using the traditional algorithm. Discordant confirmatory testing was also observed in the 35 samples, with 89% (31/35) having a reactive

Table 60

Captia™ Syphilis IgG S/CO 0.8 Cutoff False Negative Diagnostic Characteristics

RPR and TP-PA results	Frequency	Diagnosis frequency	
RPR nonreactive and TP-PA reactive	23	History syphilis infection	6
		Late latent syphilis	17
RPR nonreactive and TP-PA nonreactive	3	History syphilis infection	1
		Secondary syphilis	1
		Late latent syphilis	1
RPR reactive and TP-PA reactive	8	History syphilis infection	1
		Primary syphilis	2
		Secondary syphilis	2
		Late latent syphilis	3
RPR reactive and TP-PA nonreactive	1	History syphilis infection	1

TP-PA test and 11% (4/35) with a nonreactive TP-PA. The majority (60%) of the 35 samples (21/35) were diagnosed with late latent syphilis which is an asymptomatic, noninfectious stage and may be difficult to diagnose. If the equivocal cutoff value was lowered to include those samples with a ≥0.8 cutoff value, then the 35 samples would be reclassified as true positive which will increase the Syphilis IgG sensitivity to 86% (707/824). However, the 78 samples previously classified as true negative will now be reclassified as false positive, which will lower the Syphilis IgG specificity to 96% (3176/3296). None of the 78 samples had a syphilis diagnosis, and the other syphilis tests were nonreactive. While lowering the S/CO cutoff value would increase the test sensitivity, it also lowers the accuracy for that cutoff to 94%. There is a human and financial impact related to reporting a false positive STD. The emotional stress to the patient and their sexual contacts could be high. There is also an impact to public health with added staff time required for investigation and reporting of false positive test results. Based on the values calculated in Tables 12 and 13, the laboratory cost of confirming one false positive

Syphilis IgG test with a nonreactive or discordant RPR would be $80.79:

Syphilis IgG test performed in duplicate	Fully loaded cost $ 6.47 x 2	=	$12.94
RPR	Fully loaded cost $ 5.98	=	$ 5.98
RPR titer (calculations not shown in table)	Fully loaded cost $46.44	=	$46.44
TP-PA	Fully loaded cost $15.43	=	$15.43
Total additional cost of one false positive test			$80.79

The total laboratory cost of performing 78 false positive follow-up tests was calculated as $6,301.62 (78 x $80.79). This cost did not include the public health staff time for follow up and contact investigation nor the impact of an adverse reaction to antibiotics administered to combat the false infection.

Conversely, there is also a cost related to lack of treatment for the 35 false negative test results. In this study, because the 35 patients were diagnosed in the public health clinic based on risk standards, they would have been appropriately treated following CDC recommendations. If a high risk patient had an initial negative test result and the healthcare provider was concerned about a false negative, then additional testing could be ordered at a later date or the patient could be treated based on symptoms. All three syphilis tests were ordered on the 113 patients within the $\geq$0.8 cutoff value because the healthcare provider had a concern about a potential for syphilis infection or the nonreactive sample was reflexed by the laboratory because the S/CO value was greater than 0.450. The 35 patients were diagnosed with syphilis because the clinician had that option or the laboratory used an alert range to perform additional testing. Healthcare providers should have the option to determine which syphilis tests should be ordered based on their clinical assessment and not be restricted to a specific reflex panel. Alternatively, laboratories could determine which elevated S/CO values on nonreactive samples could be flagged with a comment on the test report to alert health care providers that additional clinical follow up should be performed. Wong, et al (2011) suggested that including a flag on the test report may indicate

those nonreactive samples that were close to the optimal cut off values but may not have antibody levels high enough to provide a reactive result. A repeat test 2-3 weeks later might provide a reactive test result using the manufacturer's cutoff value. There is precedent for reporting S/CO values for immunology tests. The CDC (2003) recommended reporting S/CO values for immunoassay tests for antibodies to hepatitis C. The CDC guidance document provides implementation recommendations for reporting S/CO values to minimize the requirement for supplemental testing of reactive anti-hepatitis C tests. Laboratory professionals can provide this type of information to healthcare providers either in written reports or verbal consultation thus potentially improving patient outcomes.

Because syphilis is a disease that can present in multiple stages and be present at different prevalence levels within a population, a test manufacturer must determine optimal cutoff values that cover a range of disease possibilities. The frequency and misclassification counts shown in Table 59 provide some insight into the basis for the manufacturer's optimal cut off values. The false negative Syphilis IgG column frequency reflects the number of samples that were misclassified as nonreactive; however, the patient was diagnosed with syphilis. Subtracting this number from the total number of samples for the cutoff value provides a frequency of false positives if the cutoff value was used for determining a reactive or equivocal sample. Dividing the false positive by the total will provide a false positive percentage if the optimum cutoff value is changed. False positive calculations for cutoff values of interest are: $\geq$0.7 FP at 87% (143/164); $\geq$0.8 FP at 69% (78/113); $\geq$0.9 FP at 71% (17/24); $\geq$1.0 FP at 50% (15/30); $\geq$1.1 FP at 10% (4/39) and $\geq$1.2 FP at 0.4% (2/44). The false positive percentages generally decrease as the S/CO value increases with the largest decrease at the 1.1 reactive cutoff value.

A test kit manufacturer must utilize a cutoff value that maximizes the benefit from testing a population and that includes the financial and psychological consequences of misdiagnosis (Ridge and Vizard, 1993). While the $\geq$ 0.8 cutoff value had the highest Youden Index as shown in Table 48, it also had a 69% false positive rate which could lead to financial and psychological consequences for the patient and for public health. Additionally, the healthcare provider may develop the perception that the laboratory was not providing quality test results due to the high number of false positive tests. Using a cutoff point that reflects high accuracy would decrease those consequences. According to the data in Table 48, the highest accuracy (95%) was obtained at the $\geq$1.1 S/CO cutoff, which is also the manufacturer's cutoff value for a reactive test result. The equivocal S/CO cutoff values of 0.9 and 1.0 also had 95% accuracy, while the $\geq$ 0.8 cutoff had 94% accuracy. According to the manufacturer's package insert, the cutoff values were determined based on multiple clinical trials using a range of serological tests to provide an optimum cut off value for reactive and equivocal values. Because the syphilis antibody level within an infected patient can vary over time, the equivocal cutoff values represent a "gray zone" which provide a healthcare provider with an alert that additional testing should be performed on the patient at a later date at which time the antibody levels may rise up to the reactive cutoff value. Based on the study data, it appears that the manufacturer's cut off values for both reactive and equivocal results provided a highly accurate test result interpretation.

The reverse algorithm which started with the Captia™ Syphilis IgG EIA test had a higher sensitivity (82%) than the traditional algorithm (64%) which started with the nontreponemal MacroVue RPR test (Table 58). The lower sensitivity for the traditional algorithm was observed previously in the study by Tong, et al (2013) with the traditional algorithm showing a sensitivity of 75.81% and the reverse algorithm showing a sensitivity of 99.85%. The Syphilis CIA test

used in the Tong, et al study detected both IgG and IgM antibodies which may explain why that study observed a higher sensitivity than seen in the dissertation study using an initial treponemal method that detected only IgG antibodies. Additionally, the MacroVue RPR test is a nontreponemal test which is not specific for syphilis antibodies and is not usually reactive in patients with a history of syphilis infection due to the decline of nontreponemal antibodies over time. The gold standard for the dissertation study was syphilis infection past or present which could explain the lower sensitivity seen for the MacroVue RPR. There were 83 patients with a history of syphilis infection in the study that were categorized as false negative in the traditional algorithm statistical analysis. If those 83 patients with a history of syphilis were removed and the statistical analysis was repeated, there was only a slight improvement in the traditional algorithm sensitivity (71%). The sensitivity of the reverse algorithm was higher than the traditional algorithm regardless of whether syphilis history was included in the syphilis diagnosis.

The accuracy of the reverse algorithm (95%) was greater than the traditional algorithm (91%). These accuracy results were further corroborated by the AUC values computed by ROC curve analysis. The ROC curve analysis was performed on the Syphilis IgG (initial test for the reverse algorithm) and the RPR titer (initial test for the traditional algorithm) to determine their ability to distinguish the presence or absence of syphilis. While both the Syphilis IgG and RPR titer results were able to detect a syphilis infection, the Syphilis IgG AUC value (0.9500) was higher than the RPR titer value (0.8155). The comparison of the ROC curve for each test's ability to classify a condition (syphilis present or absent) over all possible decision points displayed graphically that the Syphilis IgG test demonstrated a greater observed diagnostic accuracy than the RPR titer (Figure 15).

The positive likelihood ratios for both the traditional (46.83) and reverse algorithms (63.06) were above 10 indicating strong evidence for ruling in a diagnosis. According to Jaeschke et al. (1994) and Kent and Hancock (2016), a positive likelihood ratio above 10 provides strong evidence for ruling in a diagnosis and a negative likelihood ratio less than 0.1 provides strong evidence for ruling out a diagnosis (Table 5). A likelihood ratio of one provides no information about the probability of disease presence. The negative likelihood ratio for the traditional algorithm (0.36) provides weak evidence (-LR between 0.2 and 0.5) that a negative test result excludes a diagnosis of syphilis, whereas, the reverse algorithm negative likelihood ratio (0.18) provides moderate evidence for ruling out a diagnosis (-LR between 0.1 and 0.2). A false negative syphilis test has serious clinical implications for the patient as non-treatment could lead to debilitating illness later in life and is especially of concern in pregnancy because of the impact on the fetus. The public health impact of a false negative test is the continued transmission of syphilis infection within the population. The likelihood ratio is not usually reported by the clinical laboratory, but could be calculated from data determined during initial method verification and reported if requested. According to multiple authors, the likelihood ratio provides the most powerful method for measuring a test's usefulness. When combined with pretest probability of disease and Bayes' nomogram, the likelihood ratio provides a valuable individual patient diagnostic tool for the healthcare provider (Grimes and Schulz, 2005; Guyatt et al. 1986; Hulley et al., 2013; Jaeschke, Guyatt and Sackett, 1994; Kent and Hancock, 2016; Plebani, 1999).

Limitations

The limitations of the study included use of purposive convenience retrospective data. The disadvantage of nonprobability samples is that they may not be representative of the population;

however, the population receiving treatment in any STD public health clinic are more likely to be infected with an STD. In this study, the target construct was syphilis infection. A diverse group of infected individuals was needed to effectively determine the usefulness of the Captia™ Syphilis IgG test. According to Polit and Beck, (2013), a purposeful sampling approach may provide more representative samples for generalizing to a population.

Additional limitations include the potential for misclassified syphilis clinical diagnosis and staging information in the electronic medical record. The STD clinic staff were trained to recognize signs and symptoms of syphilis infection as well as to follow the CDC case definition guidelines (Table 14); however, human error may result in a misdiagnosis or typographical error. Syphilis infection is difficult to diagnose due to the lack of symptoms in later noninfectious stages. The majority of study patients (60.2%) were staged as latent syphilis which is asymptomatic. Misclassification may increase selection bias and threaten the validity of the syphilis clinical diagnosis and staging variable. There were six out of 4,077 patients (0.1%) who had a diagnosis of no syphilis infection with two reactive syphilis tests and 17 out of 4,077 patients (0.4%) who had a diagnosis of syphilis infection with no reactive syphilis tests. This indicates a low incidence of misdiagnosis by the STD clinic staff and reduces the potential for bias in the study variable.

There is also a limitation in the 3% population prevalence calculated from the number of syphilis diagnoses reported in the study patients and the total number of syphilis tests performed by the western region public health laboratory. A sampling bias may exist due to the exclusion of 984 samples that either had no diagnosis or were collected from children. In addition, there was the potential for false negative Syphilis IgG results among the 34,340 Syphilis IgG EIA tests performed. However, this potential bias was reduced by the enhanced surveillance in use at the

STD clinic with all three syphilis tests ordered for patients who presented with a risk of syphilis infection or who were members of a targeted surveillance group. Additionally, the public health laboratory director approved reflex testing for nonreactive Syphilis IgG samples with an S/CO value greater or equal to 0.450. There were 115 false negative Syphilis IgG test results with an S/CO between 0.450 and 0.899 which represents 0.3% (115/34430) of the total Syphilis IgG tests performed, indicating the possibility of a false negative Syphilis IgG within the clinic population was low.

Human error may also occur with incorrect data input of the variables in the Excel spreadsheet. The Syphilis IgG S/CO value provided an additional check on the Syphilis IgG test result. To reduce this limitation, the data sheet was reviewed to ensure that the numerical S/CO value matched the Syphilis test result with values below 0.9 resulted as nonreactive, values between 0.9-1.1 resulted as equivocal and values greater than or equal to 1.1 resulted as reactive. Additionally, the student researcher compared the spreadsheet variables to all patient charts prior to database deidentification and corrected spreadsheet errors in 191 samples. The proportion of agreement for the chart review was 0.95 (3886/4077), with 1.0 being perfect agreement, thus reducing the potential for bias in the study variables.

The study assumption that the samples were collected and transported properly provides a potential limitation in the Donabedian theoretical model structure domain which reflects the preanalytical (sample collection, processing and transport) phase of laboratory testing. All clinic and laboratory staff involved in sample collection, processing and transport were trained by laboratory staff and followed established SOPs. Additionally, the laboratory staff utilized established sample rejection criteria including rejection (no testing) of grossly hemolyzed specimens, unlabeled or mislabeled specimens, insufficient specimen quantity for testing,

leaking or broken specimen, and delayed specimen transport. Samples were centrifuged and stored refrigerated off site until delivered to the laboratory by courier using refrigerated transport practices. Refrigerated samples must be received at the laboratory within five days or placed in frozen storage. Samples over five days old and not frozen were rejected for testing due to potential loss of syphilis antibody during storage which may result in a false negative serologic test. All laboratory and clinic staff were trained on the sample rejection criteria and clinic staff were notified that a rejected sample required recollection of the sample. A total of 13 out of 34,430 samples (0.04%) were rejected during the two year study time period. In 2011, rejected samples included one unlabeled tube and 11 that exceeded the allowable five day non-frozen storage time. In 2012, one sample was rejected due to improper storage or transport. The extremely low rejection rate indicates that clinic and laboratory staff followed the established procedures for collection, processing and transport of all samples.

Conclusions and Future Studies

The purpose of this study was to determine the usefulness of the Captia™ Syphilis IgG EIA test and the reverse algorithm for detecting syphilis infection in a public health setting. Useful laboratory testing is an important diagnostic tool for determining individual syphilis infection and preventing community-wide disease spread.

This study provided supporting evidence for the higher sensitivity of the reverse algorithm versus the traditional algorithm for detection of syphilis infection. However, the Captia™ Syphilis IgG EIA test method showed a lower sensitivity than previous studies performed by the test manufacturer and lower sensitivity than the TP-PA test method. This may be due to the inability of the test method to detect IgM antibodies present in early syphilis stages or the high number of latent syphilis cases in the study population. The Captia™ Syphilis IgG was

implemented at the western region laboratory in 2008. The decision to implement automated syphilis testing was based on the high volume of manual RPR tests performed per day. The western region laboratory was performing over 100 manual RPR and HIV tests per day and was also experiencing a significant labor shortage in 2007. In order to continue to support the volume of testing submitted from the public health STD clinic, the laboratory had to implement automated immunoassay test methodology which would require discontinuing use of the traditional algorithm and implementing the syphilis reverse algorithm. In addition to the labor shortage, the laboratory was also limited on space and only had room for one automated analyzer. In 2008, the only automated immunoassay analyzer that could perform both a Syphilis and HIV test was the BioRad EVOLIS analyzer and the only syphilis test method FDA approved for that analyzer was the Captia™ Syphilis IgG EIA. Other automated analyzers could perform a total Syphilis EIA (both IgG and IgM), but could not perform the HIV test. The STD clinic requested that the public health laboratory maintain the ability to test both HIV and syphilis on the same sample; therefore, the EVOLIS analyzer was installed and both the Captia™ Syphilis IgG and BioRad HIV EIA test methods were verified. The study confirmed that utilizing the less sensitive Captia™ Syphilis IgG EIA test method for initial screening followed the CDC reverse algorithm recommendations of using the more sensitive confirmatory TP-PA test for discordant test results. Additional studies comparing the Captia™ Syphilis IgG EIA test method to treponemal EIA test methods that detect both syphilis IgG and IgM antibodies could provide further insight as to the usefulness of the IgG only EIA test method for detection of syphilis infection.

Both the reverse and traditional algorithms had high positive likelihood ratios above 10; however, the negative likelihood ratios for each algorithm only provided moderate (reverse

algorithm) and weak (traditional algorithm) evidence for ruling out a diagnosis of syphilis infection. Likelihood ratios are not often reported by clinical laboratories, but could be calculated during initial method verification and included as a comment on the final laboratory test report. This would allow healthcare providers to utilize Bayes' nomogram to determine the likelihood of syphilis infection in their individual patients. A method of estimating pretest probability for syphilis infection, such as the use of clinical decision rules would be needed to utilize Bayes' nomogram. Deumaresq, et al. (2013) developed clinical prediction rules for diagnosis of neurosyphilis in HIV-infected patients. However, there are no known clinical decision rules for detecting syphilis infection in the general population. These rules could be developed based on the following syphilis risk factors provided by the CDC treatment guidelines (2015), with points given for each type of risk factor and a pretest probability percentage calculated based on the total patient score:

- Age 15 – 24 years
- MSM
- Pregnant female
- Contact with a syphilis infected person
- HIV infection
- Symptoms of syphilis infection
- Visual or hearing complaints
- Stroke like symptoms or meningitis
- Dementia, muscle weakness, or paralysis

Based on the clinical decision rule score, the healthcare provider could then determine which syphilis tests should be ordered initially. If the patient had a high pretest probability score, then

the physician might consider ordering both a treponemal and nontreponemal test at the same time. Performing two tests on the same sample will increase the sensitivity of the screening testing. Additionally, performing a treponemal test at the same time as the nontreponemal test would reduce the possibility of missing a syphilis infection due a false negative test from the prozone effect which often occurs with nontreponemal tests. The FDA recently approved a treponemal and nontreponemal antibody multiplex test method performed on the BioRad Bioplex 2220 automated test platform (BioRad, 2016) which would provide the healthcare practitioner with this option. The data from this study was analyzed comparing the reverse and traditional algorithms separately. Because samples in this study had all three syphilis tests performed, future studies could re-analyze the data to determine the usefulness of performing the treponemal and nontreponemal tests at the same time using syphilis diagnosis as the gold standard.

The analysis highlighted the potential for use of the S/CO value to resolve discordant test results. Cutoff values equal to and above 2.5 showed 100% specificity and PPV for syphilis infection. According to Table 59, there were 15 discordant (Syphilis IgG reactive with nonreactive RPR) test results within those cutoff values and 93% (14/15) were confirmed by TPPA analysis. If the laboratory director determined that S/CO values equal to and above 2.5 did not require TPPA confirmation, there would be a laboratory cost savings of $231.45 (15 tests x $15.43 fully loaded TPPA cost). The use of a Syphilis EIA test that measured both IgG and IgM antibodies could provide a higher TPPA confirmation rate for discordant results and might allow the laboratory director, after completing appropriate studies, to lower the TPPA confirmation S/CO cutoff value resulting in greater laboratory cost savings. With the FDA approval of the BioRad treponemal/nontreponemal multiplex method which will perform both

tests simultaneously on the same sample, clinical laboratories will see discordant results similar to those identified in this study. As shown in Table 59, discordant results, either reactive Syphilis IgG with nonreactive RPR (246 samples) or reactive RPR with nonreactive Syphilis IgG (90 samples) were detected in a total of 336 samples for a discordant rate of 8% (336/4077) within the study sample population. The usefulness of the S/CO values for providing a semi-quantitative estimate of the likelihood of syphilis infection should be further explored, especially for discordant results.

The true cost of laboratory testing provided in this study included both direct and indirect costs. Prior researchers have not utilized the fully loaded cost of laboratory testing to perform cost effectiveness analysis. In this study, the fully loaded cost of Syphilis IgG and RPR testing were within $0.49 of each other when the volume was high; however, there was a large difference in the labor involved in performing manual RPR (371 minutes) versus automated Syphilis IgG (92 minutes) tests when used in a high volume laboratory setting (Table 13). With new testing methods, such as the BioRad Bioplex 2200 total Syphilis and RPR or the Gold Standard RPR, becoming available, the labor involved in RPR testing will be reduced. While smaller volume laboratories may not see the reduction in cost per test, they may be able to justify adding automated testing if the testing platform can also perform other immunoassay tests (Dunseth, Ford, & Krasowski, 2017). There also is a reduced labor benefit in the ability to perform multiple STD tests, such as syphilis and HIV on the same sample at the same time.

Laboratory automation may be one administrative response to the current shortage of medical laboratory professionals. The Bureau of Labor Statistics projects that there will be over 11,000 new laboratory technologist/technician job openings annually up into the year 2020; however, there are fewer than 5,000 students graduating each year from accredited training programs

(Rothenberg, 2016). The Bureau does not report the average age of workers; however, both the American Society for Clinical Pathology (ASCP) and Medical Laboratory Observer (MLO) send out a laboratory salary survey which provides a snapshot of the age of respondents. In the 2017 MLO survey, 42.3% of the respondents were in the 56-65 age group and 28.3% of respondents were in the 46-55 age group (MLO, 2017). With more baby boomers moving into retirement age, there will be fewer new graduates to take their place in the laboratory and labor costs may increase due to the shortage. As laboratory managers make decisions regarding automation, they need to determine the true cost of laboratory testing, including labor, as a component of cost effectiveness studies.

The reliability or precision of lot to lot variation within syphilis immunoassay reagents has been determined using only qualitative test results. Automated equipment that report numerical data such as the optical density (OD) may provide an additional mechanism to measure lot to lot variation among immunoassay reagents. The lot to lot assessment tool developed in this study used three statistical measurements of the OD values of an external control and identified a lot number with potential performance issues (Table 19). This tool may be useful for clinical laboratories as immunoassay testing becomes more automated.

Syphilis disease reports continue to rise at an alarming rate with serious health consequences for the men, women and babies infected with *T. pallidum*. A coordinated effort by epidemiologists, healthcare providers, STD clinic and laboratory staff is necessary to attempt to reduce the level of syphilis infection. A laboratory test only provides a snapshot in time of the patient's immune response to syphilis infection. The test results must be correlated to the patient's risk factors and medical history and, according to current CDC recommendations, require performance of both a treponemal and nontreponemal test. This study concluded that the

Captia™ Syphilis IgG test method and the reverse algorithm provided useful tools for detecting syphilis infection in a public health population; however, it is important for healthcare practitioners to understand the limitations of any laboratory test method within the population being tested. Laboratory professionals can provide valuable insight as to the true cost, reliability, validity, and result interpretation of test methods, as well as guidance on sample collection, processing, and transport. In April 2017, the CDC issued a Call to Action: Let's Work Together to Stem the Tide of Rising Syphilis in the U.S. (CDC, 2017) and called for the development of new tools for the detection and diagnosis of syphilis. The data from this study can be utilized by future researchers and scientists who are developing or improving syphilis detection tools.

List of References

Aamir, A. & Hamilton, R.G. (2014). Predictive Value Model for Laboratory tests: Diagnostic Specificity, Positive and Negative Predictive Value, Efficiency, Likelihood Ratio ([positive and negative]), Incidence and Prevalence. In I.R. Mackay, N.R. Rose, & D.K. Ledford (Eds.), *Encyclopedia of Medical Immunology* (pp.581-586). New York, NY: Springer.

Addison, J. (2017). Verifying New Reagent Lot Performance. Retrieved from http://moodle.kpsahs.edu/pluginfile.php/67779/mod_resource/content/2/Addison%20Handout.pdf

Algeciras-Schimnich, A., Bruns, D.E., Boyd, K.C., Bryant, S.C., La Fortune, K.A. & Grebe, S.K.G. (2013). Failure of Current Laboratory Protocols to Detect Lot-to-Lot Reagent Differences: Findings and Possible Solutions. *Clinical Chemistry*,*59*(8):1187-1194.

Algeciras-Schimnich, A. (2014). Tackling Reagent Lot-to-Lot verification in the Clinical Laboratory. *Clinical Laboratory News*. Retrieved from https://www.aacc.org/publications/cln/articles/2014/july/bench-matters

American Association of Clinical Chemists (AACC) (2015). Direct-to-Consumer Laboratory Testing Position Paper. Retrieved from https://www.aacc.org/health-and-science-policy/advocacy/position-statements

American Society for Clinical Laboratory Science (ASCLS) (2005). Value of Clinical Laboratory Services in Health Care. Retrieved from http://www.ascls.org/position-papers/177-value-of-clinical-laboratory-services/153-value-of-clinical-laboratory-services

Association of Public Health Laboratories (APHL) (2009). Laboratory Diagnostic Testing for *Treponema pallidum* Expert Consultation Meeting Summary Report. Retrieved from http://www.aphl.org/aphlprograms/infectious/std/Documents/ID_2009Jan_Laboratory-Guidelines-Treponema-pallidum-Meeting-Report.pdf

Association of Public Health Laboratories (2011). The Role of Public Health Laboratories in STD Testing. Retrieved from http://www.aphl.org/AboutAPHL/publications/Documents/ID_2011May_STDIssueBrief.pdf

Association of Public Health Laboratories (2017). The Role of Public Health Laboratories in STD Testing (Updated). Retrieved from https://www.aphl.org/aboutAPHL/publications/Documents/ID_2017Feb-PHL-STD-Testing--update.pdf#search=the%20role%20of%20the%20public%20health%20laboratory%20in%20STD%20testing

Armelagos, G.J., Zuckerman, M.K., & Harper, K.N. (2012). The Science behind Pre-Columbian Evidence of Syphilis in Europe: Research by Documentary. *Evolutionary Anthropology 21*(2);50-57.

Attia, J. (2003). Moving beyond sensitivity and specificity: using likelihood ratios to help interpret diagnostic tests. *Australian Prescriber 26*;111-113.

Ball, M.D. (2014). *Clinical Laboratory Mathematics*. Upper Saddle River, New Jersey, Pearson Education, Inc.

Beasley, D.M. (2016). The laboratory's role in the transformation to patient-centered care. *Medical Laboratory Observer, 48*(4): 46-47.

Becton Dickinson and Company, (2010). *Macro-Vue RPR Card Tests*. Franklin Lakes, NJ: Becton Dickinson and Company.

Begg, C. (1987). Biases in the assessment of diagnostic tests. *Statistics in Medicine,4*:411-423.

Bergman, J., Gratrix, J., Plitt, S., Fenton, J., Archibald, C., Wong,T., & Singh, A.E. (2013). Feasibility and Field Performance of a Simultaneous Syphilis and HIV Point-of-Care Based Screening Strategy in at Risk Populations in Edmonton, Canada. *AIDS Research and Treatment, 2013*:1-8.

Berry, G.J. & Loeffelholz, M.J. (2016). Use of Treponemal Screening Assay Strength of Signal to Avoid Unnecessary Confirmatory Testing. *Sexually Transmitted Diseases, 43*(12):737-740.

Binnicker, M.J. (2012). Which algorithm should be used to screen for syphilis? *Current Opinion for Infectious Diseases, 25*:79-85.

Binnicker, M.J., Jespersen, D. J., & Rollins, L.O. (2012). Direct Comparison of the Traditional and Reverse Syphilis Screening Algorithms in a Population with a Low Prevalence of Syphilis. *Journal of Clinical Microbiology, 50*(1):148-150.

Blackhawk BioSystems, Inc. (2003). *VIROTROL RPR Panel*. San Ramon, CA: Blackhawk BioSystems, Inc.

Bio-Rad. (2010). *Virotrol Syphilis total*. Bio-Rad. Hercules, CA.

Bossuyt, P.M.M, Reitsman, J.B., Linnet, K., & Moons, K.G.M. (2012). Beyond Diagnostic Accuracy: The Clinical Utility of Diagnostic Tests. *Clinical Chemistry,58*(12):1636-1643.

Bowen, V., Su, J., Torrone, E., Kidd, S., & Weinstock, H. (2015). Increase in Incidence of Congenital Syphilis-United States, 2012-2014. *Morbidity and Mortality Weekly Report, 64*(44):1241-1245.

Branger, J., Van Der Meer, J.T.M., Van Ketel, R.J., Jurriaans, S., & Prins, J.M. (2009). High Incidence of Asymptomatic Syphilis in HIV-Infected MSM Justifies Routine Screening. *Sexually Transmitted Diseases,36*(2):84-85.

Bujang, M.A. & Adnan, T.H. (2016). Requirements for Minimum Sample Size for Sensitivity and Specificity Analysis. *Journal of Clinical and Diagnostic Research, 10*(10): YE01-YE06.

Cantor, A.G., Pappas, M., Daeges, M., & Nelson, H.D. (2016). Screening for Syphilis Updated Evidence Report and Systematic Review for the US Preventive Services Task Force. *Journal of the American Medical Association, 315*(21):2328-2337.

Castro, A., Jost, H., Cox, D., Fakile, Y., Kikkert, S., Tun, Y. Zaidi, A., & Park, M. (2013). A comparison of the analytical level of agreement of nine treponemal assays for syphilis and possible implications for screening algorithms. *British Medical Journal*, *3*:1-4.

Caswell, R.J., Hathorn, E., & Manavi, K. (2016). The Significance of Isolated Reactive Treponemal Enzyme Immunoassay in the Diagnosis of Early Syphilis. *Sexually Transmitted Diseases, 43*(6):365-368.

Cellum, C.L., Bolan, G., Krone, M.,Code, K., Leone,P., Spaulding, C., Henry, K., Clarke, P., Smith, M., & Hook, E.W. (1997). Patients Attending STD Clinics in an Evolving Health Care Environment. *Sexually Transmitted Diseases, 24*(10):599-605.

Centers for Disease Control and Prevention (1999). The National Plan to Eliminate Syphilis from the United States. Retrieved from http://www.cdc.gov/stopsyphilis/plan-1999.htm

Centers for Disease Control and Prevention (2003). Internet Use and Early Syphilis Infection Among Men Who Have Sex with Men-San Francisco, California, 1999-2003. *Morbidity and Mortality Weekly Report, 52*(50): 1229-1232.

Centers for Disease Control and Prevention (2003). Guidelines for Laboratory Testing and Result Reporting of Antibody to Hepatitis C Virus. *Morbidity and Mortality Weekly Report*, *52*(RR03): 1-16.

Centers for Disease Control and Prevention (2006). Syphilis Elimination Effort. Retrieved from http://www.cdc.gov/stopsyphilis/seeplan2006.pdf

Centers for Disease Control and Prevention (2011). Discordant Results from Reverse Sequence Syphilis Screening-Five Laboratories, United States, 2006-2010. *Morbidity and Mortality Weekly Report, 60*(5):133-137.

Centers for Disease Control and Prevention (2012). STD Surveillance case definitions. Retrieved from http://www.cdc.gov/std/stats12/2012-casedefinitions.pdf

Centers for Disease Control and Prevention (2013). 2013 STD Surveillance Syphilis. Retrieved from http://www.cdc.gov/std/stats13/syphilis.htm

Centers for Disease Control and Prevention (2015). Clinical Advisory: Ocular Syphilis in the United States. Retrieved from http://www.cdc.gov/std/syphilis/clinicaladvisoryos2015.htm

Centers for Disease Control and Prevention (2015). 2015 Sexually Transmitted Disease Treatment Guidelines Syphilis. Retrieved from http://www.cdc.gov/std/tg2015/syphilis.htm

Centers for Disease Control and Prevention. (2017). Sexually Transmitted Diseases-Syphilis, 2016. Retrieved from https://www.cdc.gov/std/stats16/syphilis.htm

Centers for Disease Control and Prevention (2017). About HIV and AIDS. Retrieved from https://www.cdc.gov/hiv/basics/whatishiv.html

Centers for Disease Control and Prevention (2017). CDC Call to Action: Let's Work Together to Stem the Tide of Rising Syphilis in the United States. Retrieved from https://www.cdc.gov/std/syphilis/SyphilisCalltoActionApril2017.pdf

Centers for Medicare and Medicaid Services (CMS). (2017). Clinical Laboratory Improvement Amendments. Retrieved from https://www.cms.gov/Regulations-and-Guidance/Legislation/CLIA/Interpretive_Guidelines_for_Laboratories.html

Centers for Medicare and Medicaid Services (CMS). (2017). Direct Access Testing. Retrieved from https://www.cms.gov/Regulations-and-Guidance/Legislation/CLIA/index.html?redirect=/CLIA

Chesson, H.W., Collins, D., & Koski, K. (2008). Formulas for estimating the costs averted by sexually transmitted infection (STI) programs in the United States. *Cost Effectiveness and Resource Allocation, C/E*, 6(10). http://doi.org/10.1186/1478-7547-6-10.

Chesson, H.W., Pinkerton, S.D., Irwin, K.L., Rein, D., & Kassler, W.J. (1999). New HIV cases attributable to syphilis in the USA: estimates from a simplified transmission model. *Journal of Acquired Immune Deficiency Syndromes,13*:1387-1396.

Chuck, A., Ohimas, A., Tilley, P., Singh, A., & Jacobs, P. (2008). Cost effectiveness of enzyme immunoassay and immunoblot testing for the diagnosis of syphilis. *International Journal of STD and AIDS,19*:393-399.

Clinical Laboratory Standards Institute (2004). *Application of a Quality Management System Model for Laboratory Services: Approved Guideline-Third Edition.* (NCCLS document GP26-A3). Wayne, Pennsylvania.

Cohen, J. (1990). Things I Have Learned (So Far). *American Psychologist, 45*(12):1304-1312.

Cohen, J. (1992). A Power Primer. *Psychological Bulletin, 112*(1):155-159.

Cohen, S.E., Klausner, J.D., Engelman, J. & Philip, S. (2013). Syphilis in the Modern Era An Update for Physicians. *Infectious Disease Clinics of North America,27*:705-722.

College of American Pathologists. (2016). *Syphilis Serology Participant Summary G-B.* Northfield, IL: College of American Pathologists.

Cortez, K.J. & Greenwald, M.A. (2014). Current Trends in Donor Testing to Detect Syphilis Infection. *Current Infectious Disease Reports,16*(423):1-8.

Council of State and Territorial Epidemiologists (CSTE). (2004). Public Health Practice versus Research. Retrieved from https://public.health.oregon.gov/ProviderPartnerResources/EvaluationResearch/InstitutionalReviewBoard/Documents/phpractice.pdf

Cronbach, J.J. & Meehl, P.E. (1955). Construct Validity in Psychological Tests. *Psychological Bulletin,52*(4):281-302.

Crosby, A.W. (1969). The Early History of Syphilis: A Reappraisal. *American Anthropologist, 71:*218-227.

Cummings, M.C., Lukehart, S.A., Marra, C., Smith, B.L., Shaffer, J., Demeo, L.R., Castro, C., & McCormack, W.M. (1995). Comparison of Methods for the Detection of *Treponema pallidum* in Lesions of Early Syphilis. *Sexually Transmitted Diseases,23*(5):366-369.

Dai, S., Chi, P., Lin, Y., Zheng, X., Liu,W.,Zhang, J., Zeng, Q., Qu,W., Liu, W. &Wang, J. (2014). Improved Reverse Screening Algorithm for *Treponema pallidum* Antibody using Signal-to-Cutoff Ratios from Chemiluminescence Microparticle Immunoassay. *Sexually Transmitted Diseases,41*(1):29-34.

DeLong, E.R., DeLong, D.M., & Clarke-Pearson, D.L. (1988). Comparing the Areas under Two or more Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. *Biometrics,44*(3):837-845.

Dodig, S. (2009). Interferences in quantitative immunochemical reactions. *Biochemica Medica,19*(1):50-62.

Donabedian, A. (1980). *Explorations in Quality Assessment and Monitoring: The Definition of Quality and Approaches to its Assessment-Volume 1*. Anne Arbor, MI, Health Administration Press.

Domantay-Apostol, G.P., Handog, E.B., & Gabriel, T.G. (2008). Syphilis: The International Challenge of the Great Imitator. *Dermatologic Clinics,26:*191-202.

Dufour, D.R. (2004). Lot-to-Lot Variation in Anti-Hepatitis C Signal-to-Cutoff Ratio. *Clinical Chemistry, 50*(5): 958-960.

Dumaresq, J., Langevin, S., Gagnon, S., Serhir, B., Deligne, B., Tremblay, C., Tsang, R.S.W., Fortin, C., Coutlee, F., & Roger, M. (2013). Clinical Prediction and Diagnosis of Neurosyphilis in HIV-Infected Patients with Early Syphilis. *Journal of Clinical Microbiology, 51*(12): 4060-4066.

Dunseth, C.D., Ford, B.A., & Krasowsi, M.D. (2017). Traditional versus reverse syphilis algorithms: A comparison at a large academic medical center. *Practical Laboratory Medicine,8*: 52-59.

Ezzelle, J., Rodriguez-Chavez, I.R., Darden, J.M., Stirewalt, M., Kunwar, N., Hitchcock, R., Walter, T., & D'Souza, M.P. (2008). Guidelines on Good Clinical Laboratory Practice: Bridging Operations between Research and Clinical Research Laboratories. *Journal of Pharmaceutical and Biomedical Analysis,46*(1):18-29.

Fenton, K.A., Breban, R., Vardavas, R., Okano, J.T., Martin, T., Ami, S., & Blower, S. (2008). Infectious syphilis in high-income settings in the 21st century. *Lancet Infectious Disease. 8*:244-253.

Ferguson, C.J. (2009). An Effect Size Primer: A Guide for Clinicians and Researchers. *Professional Psychology Research and Practice, 40*(5): 532-538.

Food and Drug Administration (FDA). (1980). Reliability of Manufactured Products. Retrieved From https://www.fda.gov/ICECI/Inspections/InspectionGuides/InspectionTechnicalGuides/ucm072912.htm

Food and Drug Administration (FDA). (1994). Points to Consider for Collection of Data in Support of In Vitro Device Submissions for 510(k) Clearance. Retrieved from https://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM094246.pdf

Food and Drug Administration (FDA). (2007). Statistical Guidance on Reporting Results from Studies Evaluating Diagnostic Tests. Retrieved from https://www.fda.gov/RegulatoryInformation/Guidances/ucm071148.htm

Food and Drug Administration (FDA). (2011). Guidance for Industry and FDA Staff-Establishing the Performance Characteristics for In Vitro Diagnostic Devices for the Detection or Detection and Differentiation of Influenza Viruses. Retrieved from https://www.fda.gov/RegulatoryInformation/Guidances/ucm079171.htm#8

Food and Drug Administration (FDA). (2015). FDA 510(k) Premarket notification for Gold Standard Diagnostics AIX1000 Rapid Plasma Reagin (RPR) Automated Test System. Retrieved from http://www.accessdata.fda.gov/cdrh_docs/pdf15/k150358.pdf

Food and Drug Administration (FDA). (2017). BioRad Labs Receives FDA Clearance for BioPlex 2200 Syphilis Total & RPR Assay. Retrieved from http://www.fdanews.com/articles/print/182141-bio-rad-labs-receives-fda-clearance-for-bioplex-2200-syphilis-total-rpr-assay

Frith, J. (2012). Syphilis-Its early history and Treatment until Penicillin and the Debate on its Origins. *Journal of Military and Veterans' Health, 20*(4):49-58.

Fujirebio Diagnostics, Inc. (2006). *Serodia^{R} TP-PA Reagents for the Detection of Antibodies to Treponema pallidum.* Malvern, PA: Fujirebio Diagnostics, Inc.

Galen, R.S. & Gambino, S.R. (1975). *Beyond Normality The Predictive Value and Efficiency of Medical Diagnoses.* New York, NY, John Wiley & Sons.

Genc, M. & Ledger, W.J. (2000). Syphilis in Pregnancy. *Sexually Transmitted Infections,76*:73-79.

Gold Standard Diagnostics (2016). AIX1000^{TM} Fully Automated RPR Syphilis Testing for the Traditional Algorithm. Retrieved from http://www.gsdx.us/instrument-catalog/

Golden, M.B., Marra, C.M., & Holmes, K.K. (2003). Update on Syphilis Resurgence of an Old Problem. *Journal of the American Medical Association,290*(11):1510-1513.

Gratix, J., Plitt, S., Lee, B.E., Ferron, L., Anderson, B., Verity, B…Singh, A. (2012). Impact of Reverse Sequence Syphilis Screening on New Diagnosis of Late Latent Syphilis in Edmonton, Canada. *Sexually Transmitted Diseases,39* (7):528-530.

Gray, C.D. & Kinnear, P.R. (2012). *IBM SPSS statistics 19 made simple*: New York, NY; Psychology Press.

Grimes, D.A. & Schulz, K.F. (2005). Refining clinical diagnosis with likelihood ratios. *Lancet Infectious Disease,365*:1500-1505.

Guyatt, G. H., Tugwell, P.X., Feeny, D.H., Haynes, R.B., & Drummond, M. (1986). A framework for clinical evaluation of diagnostic technologies. *Canadian Medical Association Journal,134*:587-594

Hajian-Tilaki, K. (2013). Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. *Caspian Journal of Internal Medicine, 4*(2): 627-635.

Hajian-Tilaki, K. (2014). Sample size estimation in diagnostic test studies of biomedical informatics. *Journal of Biomedical Informatics, 48*: 193-204.

Halkitis, P.N., Wilton, L., & Drescher, J. (2005). *Barebacking: Psychosocial and Public Health Approaches*: Binghamton, NY: The Haworth Medical Press.

Halling, V.W., Jones, M.F., Bestrom, J.E., Wold, A.D., Rosenblatt, J.E., Smith, T.F., & Cocerrill, F.R. (1999). Clinical Comparison of the *Treponema pallidum* CAPTIA Syphilis-G Enzyme Immunoassay with the Fluorescent Treponemal Antibody Absorption Immunoglobulin G Assay for Syphilis Testing. *Journal of Clinical Microbiology, 37*(10):3233-3234.

Harper, K.N., Zuckerman, M.K., Harper, M.L., Kingston, J.D., & Armelagos, G.J. (2011). The Origin and Antiquity of Syphilis Revisited: An Appraisal of Old World Pre-Columbian Evidence for Treponemal Infection. *Yearbook of Physical Anthropology, 54*:99-133.

Hernandez, J.S. (2003). Cost-Effectiveness of Laboratory Testing. *Archives of Pathology and Laboratory Medicine,127*:440-445.

Ho, E.L. & Lukehart, S. A. (2011). Syphilis: using modern approaches to understand an old disease. *The Journal of Clinical Investigation,121*(12):4584-4592.

Holzel, W. (1991). Analytical Variation in Immunoassays and its Importance for Medical Decision Making. *Scandinavian Journal of Clinical and Laboratory Investigation*, 50:113-119.

Hoover, K. & Park, I. (2015). Reverse Sequence Syphilis Screening: Frequently Asked Questions. Retrieved from https://www.cdph.ca.gov/programs/std/Documents/STD-301-Reverse-sequence-FAQs.pdf

Hulley, S.B., Cummings, S.R., Browner, W.S., Grady, D.G. & Newman, T.B. (2013). *Designing Clinical Research*: Philadelphia, PA: Lippincott Williams & Wilkins.

Ilahi, M. (2016). What is Direct Access Testing and How Can We Afford it? *Health Law Perspectives*, January 2016:1-7.

Jaeschke, R., Guyatt, G.H., & Sackett, D. (1994). Users' Guides to Medical Literature: How to Use an Article About a Diagnostic Test. *Journal of American Medical Association,271* (9):703-707.

Japan International Cooperation Agency (JICA). (2017). Part One: Failure of Medical Equipment. Retrieved from http://gwweb.jica.go.jp/km/FSubject0201.nsf/3b8a2d403517ae4549256f2d002e1dcc/0cddc7a82dc274d94925728b0003f65d/$FILE/Part%20I.pdf

Kaji, A.H., Schriger, D., & Green, S. (2014). Looking through the Retrospectoscope: Reducing bias in Emergency Medicine Chart Review Studies. *Annals of Emergency Medicine, 64*(3):292-298.

Kaplan, C. (1990). Use of the Laboratory. In H.K. Walker, W.D. Hall, & J.W. Hurst. (Eds,), *Clinical Methods: The History, the Physical, and Laboratory Examinations.* Retrieved from http://www.ncbi.nlm.nih.gov/books/NBK1755/

Karras, D. J. (1997). Statistical Methodology: II. Reliability and Validity Assessment in Study Design. Part A. *Academic Emergency Medicine, 4(*1):64-71.

Karras, D. J. (1997). Statistical Methodology: II. Reliability and Validity Assessment in Study Design. Part B. *Academic Emergency Medicine, 4(*2):144-147.

Kent, P. & Hancock, M.J. (2016). Interpretation of dichotomous outcomes: sensitivity, specificity, likelihood ratios, and pretest and posttest probability. *Journal of Physiotherapy*, *62*:231-233.

Kim, H.S., Kang, H.J., Whang, D.H., Lee, S.G., Park, M.J., Park, J., & Lee, K.M. (2012). Analysis of Reagent Lot-to-Lot Comparability Tests in Five Immunoassay Items. *Annals of Clinical and Laboratory Science, 42*(2):165-173.

Kitchen, A.D. & Newham, J.A. (2010). Lot release testing of serological infectious disease assays used for donor and donation screening. *Vox Sanguinis*, 98:508-516.

Klausner, J.D., Wolf, W., Fisher-Ponce, L., Zolt, I., Katz, M.H. (2000). Tracing a Syphilis Outbreak Through Cyberspace. *Journal of the American Medical Association,284*(4):447-449.

Klick, R. (1997). Decision Making in the Clinical Laboratory: A Quantitative and Statistical Approach for Methods Evaluation. Retrieved from http://www.medtechnet.com/public_pdf/mtc19.pdf

LaFond, R.E. & Lukehart, S.A. (2006). Biological Basis for Syphilis. *Clinical Microbiology Reviews,19*(3):29-49.

Larsen, S.A., Pope, V., Johnson, R.E., Kennedy, E. J. (1998). *A Manual of Tests for Syphilis*. Washington, D. C.: American Public Health Association.

Lefevre, J., Bertrand, M., & Baurlaud, R. (1990). Evaluation of the Captia Enzyme Immunoassays for Detection of Immunoglobulins G and M to *Treponema pallidum* in Syphilis. *Journal of Clinical Microbiology*,*28*(8):1704-1707.

Levett, P.N., Fonseca, K., Tsang, R.SW., Kadkhoda, K., Serhir, B., Radons, S. M., & Morshed, M. (2015). Canadian Public Health Laboratory Network laboratory guidelines for serological tests (excluding point-of-care tests) for the diagnosis of syphilis in Canada. *Canadian Journal of Infectious Disease Medical Microbiology, 26* (Suppl A):6A-12A.

Linnet, K., Bossuyt, P.M.M., Moons, K.G.M., & Reitsma, J.B. (2012). Quantifying the Accuracy of a Diagnostic Test or Marker. *Clinical Chemistry*,*58*(9):1292-1301.

Lipsey, M.W. & Hurley, S.M. (2009). Design Sensitivity: Statistical Power for Applied Experimental Research. In L. Bickman & D.J. Rog (Eds.), *The SAGE Handbook of Applied Social Research Methods* (pp.44-76). doi: http://dx.doi.org/10.4135/9781483348858.n2

Liu, H., Rodes, B., Chen, C.-Y., & Steiner, B. (2001). New Tests for Syphilis: Rational Design of a PCR Method for Detection of *Treponema pallidum* in Clinical Specimens Using Unique Regions of the DNA Polymerase I Gene. *Journal of Clinical Microbiology*,*39*(5):1941-1946.

Loeffelholz, M.J. & Binnicker, M.J. (2012). It Is Time to Use Treponema Specific Antibody Screening Tests for Diagnosis of Syphilis. *Journal of Clinical Microbiology*,*50*(1):1-6.

Lord, S.J., Staub, L.P., Bossuyt, P.M.M., & Irwig, L.M. (2011). Target practice: choosing target conditions for test accuracy studies that are relevant to clinical practice. *British Medical Journal*,*343*:1-5. doi:10.1136/bmj.d4684.

Lynn, W. A. & Lightman, S. (2004). Syphilis and HIV: a dangerous combination. *The Lancet Infectious Diseases*,*4*:456-466.

Malm, K., Andersson, S., Fredlund, H., Norrgren, H., Biague, A., Mansson, F., & Ballard, R., (2015). Analytical evaluation of nine serological assays for diagnosis of syphilis. *Journal of European Academy of Dermatology and Venerology*,*29*:2369-2376.

Manavi, K., Young, H., McMillan, A. (2006). The sensitivity of syphilis assays in detecting different stages of early syphilis. *International Journal of STD & AIDS*,*17*:768-771.

Martindale, R.A., Cembrowski, G.S., Journault, L.J., Crawford, J.L., Tran, C., Hofer, T.L.,…LeGatt, D.E. (2006). Validating New Reagents: Roadmaps Through the Wilderness. *LABMEDICINE*,*37*(6):347-351.

Matthias J., Dwiggins P., Totten Y., Blackmore C., Wilson C., & Peterman, T.A. (2016). Notes from the Field. Evaluation of the Sensitivity and Specificity of a Commercially Available Rapid Syphilis Test — Escambia County, Florida, 2016. *Morbidity and Mortality Weekly Report 65*:1174–1175. DOI: http://dx.doi.org/10.15585/mmwr.mm6542a5

McKenzie, S.B. & Williams, J. L. (2015) *Clinical Laboratory Hematology*, Upper Saddle River, NJ: Elsevier Saunders.

McPherson, R.A. & Pincus, M.R. (2011). *Henry's Clinical Diagnosis and Management by Laboratory Methods*. Pittsburgh, PA: Elsevier Saunders.

Messick, S. (1995). Validity of Psychological Assessment Validation of Inferences From Persons' Responses and Performances as Scientific Inquiry Into Score Meaning. *American Psychologist* (50):9:741-749.

Miller, J. (2004). Interference in immunoassays: avoiding erroneous results. *Clinical Laboratory International*,*28*:14-17.

Mishra, S., Boily, M., Jg, V., Gold, W.L., Okura, T., Shaw, M., Mazzulli, T., & Fisman, D.N. (2011). The Laboratory Impact of Changing Syphilis Screening from the Rapid-Plasma Reagin to a Treponemal Enzyme Immunoassay: A Case-Study from the Greater Toronto Area. *Sexually Transmitted Diseases*,*38*(3):190-196.

MLO staff (2017, February). MLO's 2017 laboratory professional annual salary survey: a snapshot of our time. *Medical Laboratory Observer*. Retrieved from https://www.mlo-online.com/mlo%E2%80%99s-2017-laboratory-professional-annual-salary-survey-snapshot-time

Moore, R. D. & Chaisson, R.E. (1999). Natural history of HIV infection in the era of combination antiretroviral therapy. *AIDS*,*13*:1933-1942.

Morshed, M.G. & Singh, A.E. (2015). Recent Trends in Serologic Diagnosis of Syphilis. *Clinical and Vaccine Immunology*,*22*(2):137-147.

Motulsky, H. (2014). *Intuitive Biostatistics*: New York, NY: Oxford University Press.

National Academies of Science, Engineering and Medicine (NASEM) (2013). Crossing the Quality Chasm: The IOM Health Care Quality Initiative. Retrieved from http://www.nationalacademies.org/hmd/Global/News%20Announcements/Crossing-the-Quality-Chasm-The-IOM-Health-Care-Quality-Initiative.aspx

National Quality Forum (NQF) (2017). Improving Diagnostic Quality and Safety: Draft Measurement Framework. Retrieved from http://www.qualityforum.org/Improving_Diagnostic_Quality_and_Safety.aspx

Number Cruncher Statistical Software (NCSS). (2016) [Computer software]. Kaysville, UT: NCSS, LLC.

Obuchowski, N.A. (1998). Sample size calculations in studies of test accuracy. *Statistical Methods in Medical Research,7*:371-392.

Ostrow, D.G., Silverberg, M.J., Cook, R.L., Clamiel, J.S., Johnson, L., Li, X., & Jacobson, L.P. (2008). Prospective study of attitudinal and relationship predictors of sexual risk in the multicenter AIDS cohort study. *AIDS and Behavior, 12,* 127-138.

Owusu-Edusei, K., Peterman, T.A., & Ballard, R.C. (2011). Serologic Testing for Syphilis in the United States: A Cost-effectiveness Analysis of Two Screening Algorithms. *Sexually Transmitted Diseases,38*(1):1-7.

Owusu-Edusei, K., Koski, K.A., & Ballard, R.C. (2011). The Tale of Two Serologic Tests to Screen for Syphilis-Treponemal and Nontreponemal: Does the Order Matter? *Sexually Transmitted Diseases,38*(5):448-456.

Pandori, M.W. (2017). Establishing quality in molecular testing. *Medical Laboratory Observer,49*(8):74-75.

Park, I.U., Chow, J.M, Bolan, G., Stanley, M., Shieh, J., & Schapiro, J.M. (2011) Screening for Syphilis with the Treponemal Immunoassay:Analysis of Discordant Serology Results and Implications for Clinical Management. *Journal of Infectious Diseases,204*:1297-1304.

Pathela, P., Klingler, E.J., Guerry, S., Bernstein, K.T., Kerani, R.P., Llata, L., Mark, H.D., Tabidze, I., & Rietmeijer, C.A. (2015). Sexually Transmitted Infection Clinics as Safety Net Providers: Exploring the Role of Categorical Sexually Transmitted Infection Clinics in an Era of Health Care Reform. *Sexually Transmitted Diseases,42*(5):286-293.

Patton, M.E., Xu, J.R., Nelson, R., & Weinstock, H. (2014). Primary and Secondary Syphilis – United States, 2005-2013. *Morbidity and Mortality Weekly Report,63*(18):402-406.

Peeling, R.W. & Ye, H. (2004). Diagnostic tools for preventing and managing maternal and congenital syphilis: an overview. *Bulletin of the World Health Organization,82*:439-446

Perkins, N.J. & Schisterman, E.F. (2006). The Inconsistency of "Optimal" Cut-points Using Two ROC Based Criteria. *American Journal of Epidemiology,163*(7):670-675.

Pfister, J. (1997, Spring). Public health applications of new laboratory technologies for communicable diseases. *Wisconsin State Laboratory of Hygiene Results*. pp. 7-8.

Plebani, M. (1999). The clinical importance of laboratory reasoning. *Clinica Chemica Acta 280*: 35-45.

Plebani, M. (2002). Charting the course of medical laboratories in a changing environment. *Clinica Chemica Acta, 319*: 87-100.

Plebani, M. & Lippi, G. (2010). Is laboratory medicine a dying profession? Blessed are those who have not seen and yet have believed. *Clinical Biochemistry, 43*: 939-941.

Plebani, M., Laposata, M. & Lundberg, G.D. (2011). The Brain-to-Brain Concept for Laboratory Testing 40 Years After Its Introduction. *American Journal of Clinical Pathology, 136*: 829-833.

Polit, D.F. & Beck, C.T. (2012). *Nursing Research: Generating and Assessing Evidence for Nursing Practice*: Philadelphia, PA: Lippincott Williams & Wilkins.

Power Analysis and Sample Size (PASS) 15. (2017). [Computer Software]. Kaysville, UT: NCSS, LLC.

Ratnam, S. (2005). The laboratory diagnosis of syphilis. *Canadian Journal of Infectious Diseases and Medical Microbiology,16*(1):45-51.

Reed, G.F., Lynn, F. & Meade, B.D. (2002). Use of Coefficient of Variation in Assessing Variability of Quantitative Assays. *Clinical and Diagnostic Laboratory Immunology,9*(6): 1235-1239.

Reisner, B.S., Mann, L.M., Tholeken, C.A., Waite, R.T., & Woods, G.L. (1997). Use of the *Treponema pallidum*-specific Captia syphilis IgG assay in conjunction with the rapid plasma reagin to test for syphilis. *Journal of Clinical Microbiology,35*(5):1141-1143.

Reitmeijer, C.A., Donnelly, J., Bernstein, K.M., Bissette, J.M., Martins, S., Pathela, P., Schillinger, J.A., Stenger, M.R., Seinstock, H., & Newman, L.M. (2009). Here Comes the SSuN: Early Experiences with the STD Surveillance Network. *Public Health Reports, 124*(2):72-77.

Ridge, S.E. & Vizard, A.L. (1993). Determination of the Optimal Cutoff Value for a Serological Assay: an Example Using the Johne's Absorbed EIA. *Journal of Clinical Microbiology,31*(5): 1256-1261.

Rothenberg, I. (2016, December). Laboratory Personnel Shortages. *ADVANCE*. Retrieved from http://laboratory-manager.advanceweb.com/laboratory-personnel-shortages/

Rothschild, B.M. (2005). History of Syphilis. *Clinical Infectious Diseases,40*:1454-1463.

Rowniak, S. (2009). Safe Sex Fatigue, Treatment Optimism, and Serosorting: New Challenges to HIV Prevention Among Men Who Have Sex With Men. *Journal of the Association of Nurses in AIDS Care*, 20(1): 31-38.

Selby, C. (1999). Interference in immunoassay. *Annals of Clinical Biochemistry, 36*: 704-721.

Sena, A.C., White, B.L., & Sparling, P.F. (2010). Novel *Treponema pallidum* Serologic Tests: A Paradigm Shift in Syphilis Screening for the 21st Century. *Clinical Infectious Diseases, 51*(6):700-708.

Shi, L. & Singh, D.A. (2012). *Delivering Health Care in America A Systems Approach*: Burlington, MA: Jones & Bartlett Learning.

Siletti, R.P. (1995). Comparison of CAPTIA syphilis G enzyme immunoassay with rapid plasma reagin test for detection of syphilis. *Journal of Clinical Microbiology,33*(7):1829-1831.

Singh, A.E. & Romanowski, B. (1999). Syphilis: Review with Emphasis on Clinical, Epidemiologic, and Some Biologic Features. *Clinical Microbiology,21*(2):187-209.

Smith, B.C., Simpson, Y., Morshed, M.G., Cowen, L.L.E., Hof, R., Wetherell, C., & Cameron, C.E. (2013). New Proteins for a New Perspective on Syphilis Diagnosis. *Journal of Clinical Microbiology,51*(1):105-111).

Smith, M. (2014, May). Syphilis Rising in a Connected Age. *Medpage Today*. Retrieved from https://www.medpagetoday.com/infectiousdisease/stds/46018

Snowden, J. M., Konda, K.A., Leon, S.R., Giron, J.M., Escobar, G., Coates, T.J., Caceres, C.F., Klausner, J.D. (2010). Recent Syphilis Infection Prevalence and Risk Factors Among Male Low-Income Populations in Coastal Peruvian Cities. *Sexually Transmitted Diseases,37*(2):75-80.

Soper, D.S. (2017). Critical F-value Calculator (Software). Retrieved from http://www.danielsoper.com/statcalc

Soreng, K., Levy, R. & Faile, Y. (2014). Serologic Testing for Syphilis: Benefits and Challenges of a Reverse Algorithm. *Clinical Microbiology Newsletter 36*(24): 195-202.

Sullivan, G.M. & Feinn, R. (2012). Using Effect Size – or Why the *P* Value is Not Enough. *Journal of Graduate Medical Education*. DOI: http://dx.doi.org/10-4300/JGME-D-12-00156.1.

Tabachnick, B.G. & Fidell, L.S. (2013). *Using Multivariate Statistics*: Upper Saddle River, NJ: Pearson Education, Inc.

Tampa, M., Serbu, I., Matei, C., Venea, V. & Georgescu, SR. (2014). Brief History of Syphilis. *Journal of Medicine and Life,7*(1):4-10.

Texas Department of State Health Services (DSHS) Laboratory. (2014). Standardized Protocol for Method Validation/Verification. Retrieved from https://www.aphl.org/mrc/documents/lss_2014_texas-clia-standardized-method-validation.pdf

Tipple, C. & Taylor, G.P. (2015) Syphilis testing, typing and treatment follow-up: a new era for an old disease. *Current Opinion in Infectious Disease, 28*(1): 53-60.

Tong, M., Lin. L., Zhang, H., Huang, S., Chen, Y., Guo, X., Xi, Y., Liu,L., Chen, F., Zhang, Y., & Yang, T. (2014) Analysis of 3 Algorithms for Syphilis Serodiagnosis and Implications for Clinical Management. *Clinical Infectious Diseases,58*(8):1116-1124.

Trinity Biotech. (2003). *Captia™ Syphilis (T.pallidum)-G*. Carlsbad, CA: Trinity Biotech

Unemo, M., Ballard, R., Ison, C., Lewis, D., Ndowa, F., & Peeling, R. (2013). Laboratory diagnosis of sexually transmitted infections, including human immunodeficiency virus. World Health Organization. Retrieved from http://apps.who.int/iris/bitstream/10665/85343/1/9789241505840_eng.pdf

University of Ottawa. (2017). Validity and Reliability of Measurements. Retrieved from http://www.med.uottawa.ca/sim/data/measurement_validity.htm

Vihinen, M. (2012). How to evaluate performance of prediction methods? Measures and their interpretation in variation effect analysis. *BioMed Central Genomics*,3(Suppl 4):52:1-10.

Wawrzynski, J.J. & Hall, J.S. (2013). Cost/Benefit Analysis. In D.M. Harmening (Ed.). *Laboratory Management Principles and Processes* (pp.272-277). St. Petersburg, FL: D.H. Publishing & Consulting, Inc.

Wicher, K. Horowitz, H.W., & Wicher, V. (1999). Laboratory methods of diagnosis of syphilis for the beginning of the third millennium. *Microbes and Infection,1*:1035-1049.

Wilkinson, D.S., Dilts, T.J., Woolf, W., & Lifshitz, M.S. (2011). Financial Management. In R. McPherson & M. Pincus (Eds.), *Henry's Clinical Diagnosis and Management by Laboratory Methods* (pp. 142-154). Philadelphia, PA:Elsevier Saunders.

Willard, R., Shah, G.H., Leep, C., & Ku. L. (2012) Impact of the 2008-2010 Economic Recession on Local Health Departments. *Journal of Public Health Management Practice, 18*(2):106-114.

Woolston, S., Cohen, S.E., Fanfair, R.N, Lewis, S.C., Marra, C.M., & Golden, M.R. (2015). Notes from the field: A cluster of Ocular Syphilis cases – Seattle, Washington and San Francisco, California, 2014-2015. *Morbidity and Mortality Weekly Report,64*(40):1150-1151.

Wong, E.H., Klausner, J.D., Caguin-Grygiel, G., Madayag, C., Barber, K.O., Qui, J.S., Liska, S., & Pandori, M.W. (2011). Evaluation of an IgG/IgM Sensitive Enzyme Immunoassay and the Utility of Index Values for the Screening of Syphilis Infection in a High-Risk Population. *Sexually Transmitted Diseases,38*(6):528-532.

Xie, Q. (2017). Agree or Disagree? A Demonstration of An Alternative Statistic to Cohen's Kappa for Measuring the Extent and Reliability of Agreement between Observers. Retrieved from https://s3.amazonaws.com/sitesusa/wp-content/uploads/sites/242/2014/05/J4_Xie_2013FCSM.pdf

Yen-Liberman, B., Daniel, J., Means, C., Waletzky, J. & Daly, T.M. (2011). Identification of False-Positive Syphilis Antibody Results Using a Semiquantitative Algorithm. *Clinical and Vaccine Immunology*,*18*(6):1038-1040.

Yolken, R. H. (1979). Enzyme-Linked Immunosorbent Assay (ELISA): A Practical Tool for Rapid Diagnosis of Viruses and Other Infectious Agents. *The Yale Journal of Biology and Medicine*,*53*:85-92.

Young, H., Moyes, A., McMillan, A., & Patterson, J. (1992). Enzyme immunoassay for anti-treponemal IgG: Screening or confirmatory test? *Journal of Clinical Pathology*,*45*:37-41.

Young, H., Moyes, A., & Ross, J.D. (1995). Markers of past syphilis in HIV infection comparing Captia Syphilis G anti-treponemal IgG enzyme immunoassay with other treponemal antigen tests. *International Journal of STD & AIDS*,*6*:101-104.

Zady, M. (1999). Z-4: Mean, Standard Deviation, and Coefficient of Variation. Retrieved from https://www.westgard.com/lesson34.htm#6

Zanto, S.N. (2010). Changing Algorithms in Syphilis Laboratory Diagnosis. *Clinical Microbiology Newsletter*,*32*(8):59-63.

Zetola, N.M. & Klausner, J.D. (2007). Syphilis and HIV Infection: An Update. *Clinical Infectious Diseases*,*44*:1222-1228.

Zhang, W., Yen-Liberman, B., Means, C., Kreller, R., Waletxky, J., & Daly, T.M. (2012). The Impact of Analytical Sensitivity on Screening Algorithms for Syphilis. *Clinical Chemistry*,*58*(6):1065-1066.

Zwieg, M.H. & Campbell, G. (1993). Receiver-Operating Characteristic (ROC) Plots: A Fundamental Tool in Clinical Medicine. *Clinical Chemistry*,*39*(4):561-577.

Vita

Patricia A. Armour was born on December 4, 1950 in Superior, Wisconsin and is an American citizen. She graduated from Superior Cathedral High School, Superior, Wisconsin in 1969. She received a Bachelor of Science in Medical Technology from University of Wisconsin, Superior and certification from American Society for Clinical Pathologists as a Medical Technologist in 1973. She subsequently worked as a Medical Technologist in multiple laboratory settings including physician office, small and large hospital, public health, and environmental laboratories. She received a Master of Public Administration degree from University of Las Vegas, Nevada in 2011. She was accepted as a doctoral candidate at Virginia Commonwealth University in 2015. She is currently employed as a Clinical Laboratory Science Faculty Instructor at the College of Southern Nevada, Las Vegas, Nevada.