# A new antibody/antigen combination rapid test to detect acute HIV infection:

# A synthesis of evidence

**Megan Smallwood**

Department of Epidemiology, Biostatistics, and Occupational Health

McGill University, Montréal

April 2015

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science

## ABSTRACT

**Background**: Rapid and point-of-care diagnostics have expanded access to HIV testing, with fourth generation HIV rapid tests (Ag/Ab combo) now offering the potential of timely detection of acute HIV infection, when HIV is highly infectious. The aim of this thesis is to synthesize evidence on the global diagnostic performance of the only FDA-approved fourth generation rapid test, the Determine HIV 1/2 Ag/Ab Combo RT, to detect acute HIV in adults.

**Methods**: We performed a systematic review and meta-analysis, searching Medline, Embase, PubMed, BIOSIS, The Cochrane Library, LILACS, and African Index Medicus, including conferences, bibliographies, and citations. Studies were included if they evaluated the Determine Combo Rapid Test in adults, against a reference standard. Two reviewers independently extracted data and assessed study quality with QUADAS-2. Our main outcomes of interest were sensitivity and specificity (overall, plus antigen and antibody components). Data from 17 studies (n=21599 patient samples) were pooled using a Bayesian hierarchical random effects meta-analysis model, which assumed a perfect gold standard was available for each study. To explore the extent to which the pooled estimates obtained through this model might differ under different assumptions, data were pooled using a hierarchical random effects model which assumed varying thresholds for positivity, allowing sensitivity and specificity to be correlated within each study. We analyzed subgroups by blood sample and study design for each model.

**Results:**

Using the Bayesian model which assumed a perfect gold standard, the overall pooled sensitivity for the device was 88.5%, 95% credible interval (CrI) [80.1 – 93.4], and overall pooled specificity was 99.1%, 95% CrI [97.3 – 99.8]. Pooled sensitivity of the antigen component was

12.3%, 95% CrI [1.1 – 44.2], with a pooled antigen specificity of 99.7%, 95% CrI [96.8 – 100]. Pooled sensitivity of the antibody component was 97.3%, 95% CrI [60.7 – 99.9], and pooled antibody specificity was 99.6%, 95% CrI [99.0 – 99.8]. Estimates using the continuous threshold model did not differ notably from the perfect gold standard model, giving us confidence that our estimates are robust. Individual study limitations included failure to blind reference standard results, and selecting patients or samples based on HIV status, resulting in potential for bias. None of the included studies were considered to have a low risk of bias. Data limitations prevented sub-group analyses by reference standards, and statistical exploration of the effect of patient case-mix on accuracy.

**Conclusions:**

HIV infection is accurately detected by the Determine HIV Combo in individuals who have seroconverted; however the diagnostic accuracy of the antigen component needs to be improved for detecting acute HIV infections at point-of-care.

**Résumé**

**Contexte:** Le diagnostic rapide au point de service a élargi l'accès au dépistage du VIH. En effet, la quatrième génération de test de dépistage rapide du VIH (Ag/Ab combo) offre maintenant le potentiel de détection rapide de l'infection par le VIH durant sa phase aiguë et contagieuse. L'objectif de cette thèse est de synthétiser les données probantes sur la performance diagnostique globale du Determine HIV 1/2 Ag/Ab Combo RT, qui est le seul test rapide de quatrième génération approuvé par la FDA pour la détection d'une infection aiguë au VIH chez les adultes.

**Méthodes:** Nous avons effectué une revue systématique et méta-analyse des études identifiées par une recherche Medline, Embase, PubMed, BIOSIS, The Cochrane Library, LILACS et l'Index Medicus Africain, tout en incluant des conférences, des références bibliographiques et des citations de ressources d'information. Des études ont été incluses si elles ont évalué la performance du test rapide Determine Combo par rapport à un étalon de référence chez les adultes. Deux examinateurs ont extrait les données de façon indépendante et ont évalué la qualité des études retenues à l'aide de QUADAS-2. Nos principaux résultats d'intérêt étaient la sensibilité et la spécificité (globale, ainsi que les composants antigène et anticorps). Les données de 17 études (n=21599 échantillons de patients) ont été regroupées à l'aide d'une méta-analyse avec un modèle hiérarchique Bayésien à effets aléatoires qui, pour chaque étude, suppose la disposition d'une norme de référence parfaite. Les données ont été rassemblées en utilisant un modèle hiérarchique à effets aléatoires qui a assumé divers seuils de positivité, permettant que la sensibilité et la spécificité soient corrélées au sein de chaque étude. Cela permet d'explorer la mesure dans laquelle les estimations groupées obtenues par ce modèle peuvent varier selon différentes hypothèses. Pour chaque modèle, nous avons aussi effectué des analyses en sous-groupes du type d'échantillon de sang et de la méthodologie de conception de l'étude.

**Résultats:** En utilisant le modèle Bayésien qui suppose une norme de référence parfaite, la sensibilité combinée globale du test était de 88,5%, 95% intervalle de crédibilité (ICR) [80,1 à 93,4], et la spécificité combinée globale était de 99,1%, 95% ICR [97,3 à 99,8]. La sensibilité groupée du composant antigène était de 12,3%, 95% ICR [1,1 à 44,2], et la spécificité antigénique groupée était de 99,7%, 95% ICR [96,8 à 100]. La sensibilité groupée du composant d'anticorps était de 97,3%, 95% ICR [60,7 à 99,9], et la spécificité groupée d'anticorps était de 99,6%, 95% ICR [99,0 à 99,8]. Les estimations obtenues du modèle à divers seuil de positivité ne diffèrent notamment du modèle de norme parfaite, ce qui nous donne confiance que nos estimations sont robustes. Les limitations des études retenues incluent l'échec de faire en sorte que les examinateurs ne connaissent pas les résultats des étalons de référence, et la sélection des patients ou des échantillons fondée sur la base de statut sérologique VIH, d'où les biais résultants. Aucune des études incluses étaient considérés comme ayant un faible risque de biais. Les limitations des données ont empêché les analyses en sous-groupe des étalons de référence, ainsi que l'exploration statistique de l'effet de l'indice de gravité des cas sur la précision.

**Conclusions:** L'infection par le VIH est détectée avec précision par le test Determine HIV Combo chez les sujets qui ont présenté une séroconversion. Mais la précision diagnostique du composant antigène doit être améliorée pour détecter au point de service les infections aiguës au VIH.

## ACKNOWLEDGEMENTS

I would first and foremost like to thank my supervisor, Dr. Nitika Pant Pai for her guidance, support, and encouragement; it has been a pleasure working under her supervision. I would also like to thank Dr. Lawrence Joseph for his invaluable feedback and insight on the statistical analysis, in particular for the Bayesian meta-analysis.

I am very grateful for the countless hours that Rohit Vijh devoted as second reviewer of this systematic review; we encountered many challenges but he remained enthusiastic throughout the process. I am also grateful to Bénédicte Nauche from the McGill University Health Centre Library for her expertise in the literature search, and Dr. Bertrand Lebouché for providing much appreciated insight on the importance acute HIV infection and editorial help with our manuscript.

Thank you to Jana Daher for translating the abstract into French, and for her continuous, never-failing support.

Finally, thank you to my classmates, friends, and teammates, for giving me a rich and fulfilling experience here at McGill.

## PREFACE AND CONTRIBUTION OF AUTHORS

An abstract including work described in this thesis was presented as a poster on July 21st 2015 at the prestigious 8th IAS (International AIDS Society) Conference on HIV Pathogenesis, Treatment and Prevention. This work was also presented at the 11th Annual McGill Department of Epidemiology, Biostatics and Occupational Health (EBOH) Student Research Day, on April 22, 2015.

A manuscript including work presented in this thesis is currently under review for publication at PloS One.

Contributions of co-authors for this thesis are outlined below:

Conception and design of the systematic review: Megan Smallwood, Dr. Nitika Pant Pai, and Dr. Bertrand Lebouché. In addition, Bénédicte Nauche, librarian of the McGill University Health Centre, assisted me in designing the search strategy and adapting it to all databases; Rohit Vijh acted as second reviewer; Dr. Lawrence Joseph conceived the Bayesian meta-analysis, which I replicated under his supervision; I designed and ran the meta-analysis using frequentist methods in Stata for this thesis. The abstract was translated to French by Jana Daher.

I completed all of the writing, literature review, presentation, and figures included in this thesis. Thesis editing was done by Dr. Pant Pai, and Dr. Joseph reviewed the data synthesis methods.

# TABLE OF CONTENTS

## LIST OF ABBREVIATIONS AND ACRONYMS

HIV – Human immunodeficiency virus

AIDS – Acquired immunodeficiency syndrome

WHO – World Health Organization

UNAIDS – Joint United Nations Programme on HIV/AIDS

ART – Antiretroviral therapy

HAART – Highly active antiretroviral therapy

MSM – Men who have sex with men

POC – Point-of-care

POCT – Point-of-care testing

RNA – Ribonucleic acid

CDC – Centers for Disease Control and Prevention

NAAT – Nucleic acid amplification testing

EIA – Enzyme immunoassay

AHI – Acute HIV infection

# CHAPTER 1: INTRODUCTION

In 2013, there were an estimated 35.3 million people [32.2 million - 38.8 million] living with human immunodeficiency virus (HIV) worldwide; with a vast majority residing in low and middle-income countries.[1] Access to antiretroviral therapy (ART) has greatly reduced the morbidity and mortality rates from HIV; in 2012 there were 9.7 million people receiving ART in low and middle income countries, and ART prevented an estimated 5.5 million deaths from acquired immunodeficiency virus (AIDS) in low and middle income countries from 1996 to 2012.[1] However, new infection rates remain a concern, with 2.1 million [1.9 million – 2.4 million] new infections occurring in 2013.[2] Efforts to decrease sexual transmission of HIV have been largely successful, with the number of new infections decreasing by 50% between 2001 and 2012 in 26 countries; however in some sub-Saharan African countries, risky sexual behaviours (e.g. decreasing condom use) have been on the rise.[1] Additionally, 19 million of the 35 million individuals living with HIV globally are unaware of their HIV status,[3] highlighting the fact that although much progress has been made in curbing the global HIV epidemic, many challenges remain.

In 2000, the eight Millennium Development Goals were established by member states of the United Nations, to address extreme global poverty and improve the lives of those affected by it.[4] The sixth goal targets HIV/AIDS, planning to have halted and reversed the spread of HIV by 2015, and to have universal access to HIV treatment by 2010. As we are approaching the end of 2015, much progress has been made in preventing the spread of HIV, with new infection rates decreasing overall;[4] additionally, there has been a rapid increase in access to ART.[4,5] Nevertheless, there is still much work to be done in further expanding ART coverage and

decreasing new cases of HIV infection globally. Global agencies are creating new targets, goals, and programs to push towards the end of this epidemic.

Looking forward, the Joint United Nations Programme on HIV/AIDS (UNAIDS) is now seeking new fast-track targets such as the 90-90-90 treatment initiative: by 2020, 90% of those living with HIV will know their status, 90% of HIV diagnosed individuals will be receiving life-saving ART, and 90% of those on ART will achieve viral suppression.[6] These new targets emphasize equity, quality of outcomes, preventative benefits of HIV treatment, and speed of scale-up. These targets must be achieved in the next five years if the global HIV epidemic is to be ended by 2030, as modelling estimates suggest.[6]

While much of the focus in tackling the HIV epidemic has been centred around countries with the highest burdens of HIV (e.g. sub-Saharan Africa), developed countries such as Canada cannot and should not fall behind in HIV control. An estimated 71 300 people were living with HIV in Canada in 2011, and this number continues to rise.[7] Canada faces problems with individuals seeking testing, particularly in high risk subgroups which are overrepresented in this epidemic. HIV disproportionally affects men who have sex with men (MSM), comprising almost 50% of those living with HIV, and almost 50% of all incident infections.[7] People who inject drugs make up approximately 20% of those living with HIV in Canada.[7] Only 2.2% of the Canadian population was born in an HIV-endemic country; however 15% of prevalent HIV infections are comprised of individuals exposed through heterosexual contact who come from an HIV endemic country.[7] Another group disproportionally affected by HIV is the Aboriginal population in Canada, who represent 9% of all HIV infections, and among whom in 2011 the new infection rate was 3.5 times that of the non-Aboriginal population.[7]

Approximately 25% of Canadians living with HIV are unaware of their serostatus.[7] With increasing access to treatment and counselling, it is essential that these individuals are not left behind, particularly these key high risk populations who may be less likely to access health services. To achieve the 90-90-90 goals, access to HIV diagnostic testing services and quality of service delivery must be improved, so that all people infected with HIV can be empowered with the knowledge of their status, receive life-saving ART, and achieve an undetectable viral load.

As diagnostic technology improves in accuracy, there is a growing need for tests which can detect HIV during early and acute infection. In light of this need to bring people into care early, acute and early identification has become an important research agenda item.

Early identification of HIV infection is crucial for improving treatment initiation, and also for preventing disease transmission. The risk of HIV transmission is highest during acute and early infection, accounting for up to 50% of new infections.[8-12] Timely knowledge of HIV status, staging, and ART initiation, can potentially reduce HIV transmission to sexual partners.[13,14] Other advantages to beginning ART early include improved CD4+ counts, decreased severity of acute retroviral syndrome, suppression of viral load, and control of HIV infection in the community.[15-17] In response to recent evidence, the WHO has issued recommendations for earlier initiation of ART, to increase the quality and duration of the lives of those living with HIV.[18]

Conventional testing for HIV requires laboratories with advanced equipment, and delays ranging from days to weeks for patients to receive results. The introduction of rapid and point-of-care (POC) tests for HIV has allowed expanded access to HIV testing, with their promise of a quick turn-around time and obviated need for complex laboratory resources.[19] Rapid POC tests have been a significant breakthrough for HIV testing, both in resource-limited settings and high

income settings.[20-23] POC testing technology is incredibly innovative in the sense that patients can receive their test results on the same visit they are tested, minimizing the number of HIV positive individuals who do not return to collect their results, expediting clinical decision-making and improving patient outcomes.[24,25]

While these POC tests have high accuracy,[25,26] most HIV POC tests currently in use are second and third generation tests, which are antibody-based and limited to detecting infection after three to four weeks.[12,27] RNA is the first marker of HIV infection, becoming detectable in the blood approximately 10 days following transmission, p24 antigen becomes detectable five to seven days after RNA (15-17 days post transmission), and HIV antibodies are detectable five days after p24 antigen (approximately 20-22 days post transmission).[28] Fourth generation assays have improved upon second and third generation assays by detecting p24 antigen in addition to HIV antibodies, allowing diagnosis to be made five days earlier, during the critical acute phase of infection where the risk of transmission is high.[12,29] However, most of these fourth generation assays require complex laboratory equipment and have been designed for use in resource-rich settings. At present, there is only one commercially available, US Food and Drug Administration (FDA)-approved fourth generation HIV assay designed for use at point of care: The Determine HIV 1/2 Ab/Ag Combo Rapid Test (Alere).[30,31] This rapid test was approved by the FDA on August 9th, 2013,[31] and its diagnostic accuracy has been evaluated in studies from across the world, in both laboratory and field settings. However, accuracy estimates between studies have been inconsistent, particularly in terms of the sensitivity ranges of the test.

This is of importance as the Determine Combo rapid test is currently in use in various countries worldwide including the United States, Australia, South Africa, and many countries in South America and Europe.[32] In 2014, the US Centres for Disease Control and Prevention (CDC)

issued updated guidelines recommending initial screening for HIV infection using fourth generation assays; however, the CDC has yet to recommend the use of the Determine Combo rapid test in place of laboratory-based assays, due to insufficient information on its field performance.[33] To the best of my knowledge, a systematic review of the overall diagnostic performance of the Determine Combo test has not been published to date. This information is needed for HIV testing and screening strategies for populations with limited resources and high incidence of acute infections.

In this context, the objective of this thesis is to synthesize evidence on the overall diagnostic performance of the Determine HIV 1/2 Ag/Ab Combo rapid test. The following chapters will provide a comprehensive review of the literature pertaining to the biological mechanisms of acute HIV infection, HIV diagnostics at POC, the importance of acute diagnosis, and will explore the added value of including a fourth generation POC assay into testing algorithms globally. Independent studies evaluating the performance of the Determine combo test will be reviewed descriptively.

Chapters 3 through 5 will present a systematic review and meta-analysis, synthesizing global evidence for the Determine combo test. The diagnostic performance will be defined through accuracy parameters (sensitivity, specificity), and evidence on the separate components of the test (antibody, p24 antigen) will be synthesized in addition to overall accuracy. Furthermore, the effects of study design and blood sample type will be explored statistically through subgroup analysis, and the effect of patient spectrum on the overall accuracy will be explored descriptively. Chapter 5 will additionally place the results of the meta-analysis in context and discuss future directions for research in this field.

# CHAPTER 2: LITERATURE REVIEW

## 2.1 Molecular biology of HIV transmission

A general understanding of the molecular biology of HIV and progression of the infection is useful for understanding how HIV is diagnosed. HIV is most commonly transmitted through sexual exposure, either in the genital or rectal mucosal tissue.[34] HIV is a lentivirus, in the family of retroviruses, meaning it has an RNA genome contained within an envelope.[35] The envelope contains two structural proteins: gp41 and gp120, and the core of HIV is made up of three structural proteins: p24, p16, and p9.[35] HIV is transmitted through the binding of HIV envelope proteins to host CD4+ T helper cells, or Langerhan's cells, and a co-receptor (either a CCR5 or CXCR4 receptor).[36] The envelope of HIV then fuses to the host cell and the viral capsid is released into the host cell, immediately releasing viral RNA.[35] HIV begins to rapidly replicate, and disseminates to the lymph nodes within two days after infection; the virus enters the bloodstream for widespread dissemination within three days.[37] Clinically, an individual infected with HIV at this time will have no signs or symptoms of infection.

## 2.2 Progression and staging of HIV infection to Acquired Immunodeficiency Syndrome (AIDS)

The stages of HIV infection can be defined in immunological or clinical terms; however there are three general stages of HIV progression (Figure 1). The first stage, which will be the focus of this thesis, is primary HIV infection, and this acute period lasts only a few weeks. It is characterized by high levels of HIV in the blood (known as high viral load), a drop in the number of CD4+ cells, and patients may experience acute retroviral syndrome, which is accompanied by fever and flu-like symptoms.[38,39] As the patient's immune system begins to respond to the virus, antibodies to HIV begin to appear in the blood, through a process known as seroconversion.

The following stage is clinical latency, or chronic infection, which can last up to ten years in patients not receiving treatment.[38] Patients in this stage are asymptomatic, however they are still infectious, and HIV is very active in the lymph nodes. Patients receiving ART will generally remain asymptomatic for decades; however, individuals not receiving treatment will begin to show symptoms, when the immune system begins to fail and CD4+ T helper cell count drops.[38] These individuals will succumb to opportunistic infections, which the immune system would normally prevent. The final stage is when the immune system further deteriorates and HIV infection progresses to Acquired Immunodeficiency Syndrome (AIDS).

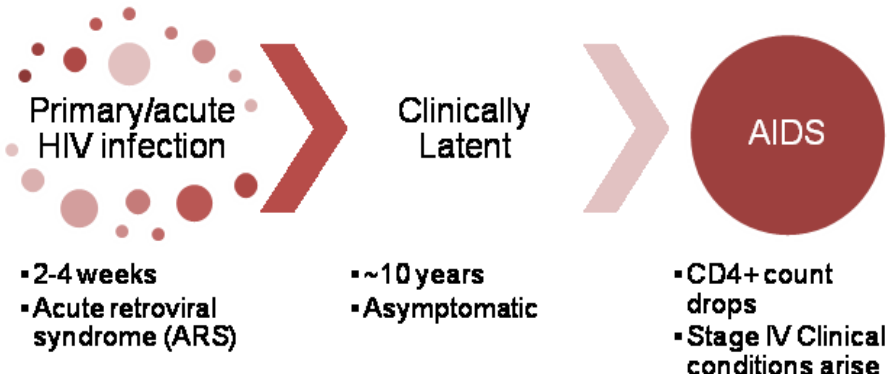


**Figure 1** Progression of HIV

The World Health Organization (WHO) released definitions for HIV clinical staging in adults in 2007, with four clinical stages (following primary HIV infection). Clinical stage 1 is defined as being clinically asymptomatic, or having persistent enlarged or swollen lymph nodes (lymphadenopathy).[39] Clinical stage 2 is characterized by mild symptoms (e.g. moderate weight loss, recurrent respiratory infections); and clinical stage 3 is characterized by more advanced symptoms. Clinical stage 4 is defined by severe clinical conditions such as extrapulmonary tuberculosis, HIV encephalopathy, and Kaposi sarcoma, among many others. A definition of

AIDS in this context is either a diagnosis of a Stage IV clinical condition, or when CD4+ cell count drops below 200 cells/$mm^3$.[40]

If HIV is diagnosed swiftly and treatment is initiated in a timely fashion; individuals receiving ART can delay and potentially prevent progression to AIDS. To optimize diagnosis and screening for HIV, the various stages of early HIV infection must be adequately differentiated.

## 2.3 Diagnostic stages of early HIV infection

Early or recent HIV infection is considered to be the first six months after acquiring the virus,[16] and this is a crucial period for diagnosis and treatment initiation. While long-term infection and progression to AIDS can be staged using clinical symptoms and CD4+ cell count, recent HIV infection is characterized by detecting biomarkers using available technology. The timing of emergence of various markers of HIV progression tends to follow a fairly consistent pattern between patients, and a staging system based on the appearance of these markers in the blood was developed by Fiebig and colleagues in 2003.[36,41]

Following exposure and successful transmission, HIV is undetectable in blood samples, and this stage of infection is known as the eclipse phase which lasts about 10 days.[36] Viral RNA becomes detectable by polymerase chain reaction (PCR) at concentrations of $10^2$ copies of RNA/mL of plasma, lasting 5-7 days, known as Fiebig Stage I.[34,36] Stage II occurs approximately 15-17 days post transmission, and is characterized by the presence of both viral RNA (rapidly replicating), and p24 antigen, lasting 5 days.[41] Stages I and II are commonly considered to comprise acute HIV infection, when RNA or p24 antigen is detectable, but antibodies are not; however, this definition is largely dependent on the accuracy of the tests used to detect these markers. Antibodies to HIV begin to become detectable 20-22 days after transmission (Fiebig Stage III),

as viral RNA is reaching peak levels in the plasma; this stage lasts approximately 3 days. [28,34,41] Once antibodies are present in the blood, p24 antigen becomes more difficult to detect, as the antibodies bind p24 antigen forming immune complexes, reducing its availability in the plasma to be picked up by diagnostic tests.[42,43] Stage IV is defined by an indeterminate Western blot pattern (showing bands but failing to meet FDA criteria), and Stage V is reached when a full Western blot pattern is achieved excluding p31 reactivity (approximately 1 month after infection); this stage is much longer than stages I-IV, lasting approximately 70 days.[41] Stage VI is indefinitely long, characterized by full Western blot positivity including p31.[41]

As mentioned above, the first six months following HIV transmission are defined as early HIV infection, and acute HIV, refers to the period of infection between RNA detection and the appearance of antibodies.[12] For clarity, when acute HIV infection is described in this thesis, it will be referring to this definition. Primary infection is frequently used interchangeably with acute infection, referring to the stage of infection prior to seroconversion (antibodies in the blood).[44]

Because these definitions are reliant on the ability of diagnostic technology to detect these laboratory markers, evolving diagnostic technology can change how we stage early infection. The staging system proposed by Fiebig and colleagues was published over 10 years ago.[41] In 2013, Ananworanich and colleagues proposed a novel system for staging HIV infection, which incorporates the widely used fourth generation immunoassays for detecting acute infection.[45] In contrast to the first four stages of the Fiebig system (before Western blot is positive), the system proposed by Ananworanich and colleagues consists of only three stages until Western blot positivity. Stage 1 is when NAAT is positive but 4$^{th}$ generation and 3$^{rd}$ generation immunoassays are negative, stage 2 is when both NAAT and 4$^{th}$ generation immunoassays are positive but 3$^{rd}$

generation tests are negative, and stage 3 is when NAAT, 4th generation and 3rd generation assays are positive but Western blot is negative or indeterminate.[45] Advantages of this new staging system are that Fiebig Stage I subjects can be split into two groups (4th generation stages 1 and 2) with different levels of DNA and RNA, and duration since exposure to HIV,[45] which may be relevant for designing treatment plans. A limitation of the staging system proposed by Ananworanich is its applicability in resource limited settings, as similarly to the Fiebig system, the fourth generation system still relies on laboratory intensive Western blot technology.[45] A comparison of both the Fiebig staging system and the 4th generation staging system proposed by Ananworanich can be seen in Table 1.

**Table 1:** Comparison of Fiebig and Ananworanich staging systems for HIV infection

| **Fiebig staging** | | **Ananworanich 4th generation staging** | |
|---|---|---|---|
| *Eclipse phase* | HIV undetectable | | |
| *Stage I* | Viral RNA detectable | *Stage 1* | NAAT positive<br>4th generation assay negative<br>3rd generation assay negative |
| *Stage II* | P24 antigen detectable | *Stage 2* | NAAT positive<br>4th generation assay positive<br>3rd generation assay negative |
| *Stage III* | HIV antibodies detectable | *Stage 3* | NAAT positive<br>4th generation assay positive<br>3rd generation assay positive<br>Western blot negative/indeterminate |
| *Stage IV* | Indeterminate Western blot | | |
| *Stage V* | Full Western blot excluding p31 | | |
| *Stage VI* | Full Western blot including p31 | | |

## 2.4 Current laboratory technologies for screening and diagnosing HIV infection

The first immunoassays to screen for HIV were developed in the 1980's; these assays were able to detect antibodies to HIV and were shortly followed by Western blot.[42] Western blot has been used as a confirmatory assay for HIV for the past 25 years; only recently has it begun to be replaced by other confirmatory tests which produce less indeterminate results, and are less time

and labour intensive.[46] The first and second generations of enzyme immunosassays (EIA) were designed to detect IgG antibodies to HIV, with the second generation assays improving on specificity and detecting HIV up to a week earlier compared to first generation assays.[47] Third generation immunoassays evolved to detect infections earlier by detecting IgM in addition to IgG, diagnosing HIV infection 20-25 days after transmission.[47] Furthermore, some newer third generation assays can detect HIV-2 in addition to HIV-1.[42] HIV-1 is the predominant form of HIV worldwide; HIV-2 is only common in West Africa.[46] While both types of HIV share similarities, HIV-2 has lower transmissibility than HIV-1, and HIV-2 leads to slower progression to AIDS.[46] Diagnostic tests which can detect HIV-2 in addition to HIV-1 offer benefit to countries in West Africa where traditional assays for HIV-1 would prove useless.

In the late 1990's, fourth generation immunoassays began to appear on the market. Fourth generation assays offer the advantage of detecting p24 antigen in addition to detecting IgM and IgG antibodies to HIV, and they are often referred to as antigen/antibody combination assays. These assays can detect HIV during the acute period before antibodies are present, shortening detection time by about 5 days.[47] A timeline of HIV detection by various diagnostic technologies can be seen in Figure 2. Fourth generation assays have largely replaced second and third generation immunoassays in Europe and Australia, where fourth generation assays have been in use for well over a decade; however, the United States and Canada have been slow to update their testing algorithms, with the first fourth generation assay licensed in the US in 2010.[48]

Nucleic acid amplification testing (NAAT) remains the most sensitive technology for detecting HIV infection, as NAAT can detect HIV RNA as quickly as 10 days following infection.[44,47] NAAT is expensive and relies on complex equipment and technical skill, making it far from an ideal candidate for a screening test. Without the availability of fourth generation assays in the US

and Canada, NAAT has been used off-label for screening for acute HIV infection by pooling specimens for cost-effectiveness; if the pooled sample reveals a positive result, each blood sample is individually tested. [48,49]

Alternatively, assays which exclusively detect p24 antigen are used to diagnose acute HIV in cases where initial EIA testing is positive but the confirmatory Western blot is negative or indeterminate.[50] In the past, these assays had low sensitivity due to lack of free p24 antigen in the blood once it forms immune complexes with antibodies in the blood; however improvements in disrupting these immune complexes have improved sensitivity.[51] Similar to NAAT, ultrasensitive p24 assays are limited in practicality for HIV screening by their cost, and their need for bulk sampling (cannot test samples individually), making fourth generation assays ideal for initial screening of at-risk individuals unaware of their HIV status.[46,52] Even so, fourth generation assays require technical skill and access to advanced laboratory equipment, which renders these assays more suitable to high-income settings than resource-limited settings.

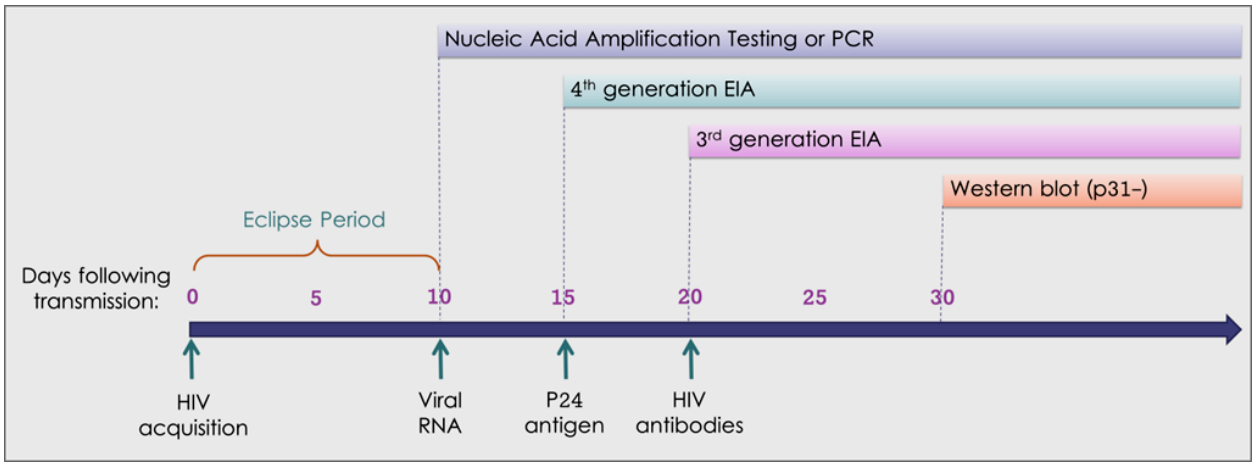


**Figure 2** Timeline of detection of markers for HIV by diagnostic technology. Adapted from[28,41].

## 2.5 Point-of-Care (POC) Diagnostics for HIV

Many low-income countries do not have access to the advanced laboratory equipment and technicians needed for EIAs. Diagnostic POC testing has been a breakthrough in improving access to diagnostic testing in resource-limited settings, particularly for infectious diseases including HIV.

POC testing began in the 1960's, and was introduced into clinics in the 1990's.[24] While there are various definitions for POC tests, and no accepted universal definition, they all focus around the concept of rapid turn-around time of results, allowing a patient to receive results and clinical diagnosis or intervention during the same clinical visit.[53] This has implications for both resource-limited and resource-rich settings; where HIV patients may be deterred from testing and lost to follow up due to stigma, discrimination, and long wait times for results. In Canada, TAT for patients to receive HIV laboratory results can range from a few days to several weeks,[54] in rural areas of developing countries TAT can be several months.[25] It is critical that people living with HIV are diagnosed, as estimates suggest that people who are unaware of their HIV positive status are responsible for over 50% of new HIV infections.[10]

In 2004, the WHO released recommendations for POC diagnostics for sexually transmitted infections (STI's), describing criteria for diagnostic technology to be considered POC in resource limited settings.[54] The WHO concluded that POC diagnostics must be "ASSURED": affordable, sensitive, specific, user-friendly, rapid and robust, equipment-free, and delivered (accessible to end-users).[54] Many HIV POC diagnostics have been successful in meeting the ASSURED criteria as they are rapid, can detect HIV antibodies in whole blood specimens without electricity or lab equipment, and can be carried out by individuals with no formal lab training.[25] While there has been concern about the accuracy of these tests, particularly in terms of the implications of

false positive and false negative results, these tests have been shown to be both accurate and reliable, with sensitivity greater than 95% and specificity greater than 99% in both developed and developing settings.[19,25,26]

POC tests for HIV have been very successful in increasing access to screening in resource limited settings.[55] Many studies conducted in sub-Saharan Africa have found high patient and clinical staff acceptability in using rapid POC testing for HIV;[23,56,57] and POC testing is proven cost-effective for many low-income countries with high HIV prevalence.[58,59] POC tests have also found success in high-income countries, particularly with stigmatized populations such as MSM, and in acute medical admissions or emergency department settings. Patients reported high acceptability in Canadian emergency departments,[60] non-urban Canadian clinical settings,[22] and preference for rapid testing over conventional testing among high-risk sub-groups such as MSM in various high-income countries.[20,61-63]

However, as progress has been made with fourth generation laboratory immunoassays, most POC tests for HIV are second and third generation rapid tests, meaning they can detect HIV antibodies but not p24 antigen, limiting their use in detecting acute HIV infection.[64,65] Across diverse settings, screening for HIV using rapid antibody tests alone is failing to pick up cases of acute HIV infection,[27,66-68] and as our understanding of HIV biology and epidemiology has improved, the importance of detecting HIV as soon as possible has been highlighted.[65]

## 2.6 Acute HIV infection: added value of detection

As defined earlier, acute HIV infection is the period of time between RNA becoming detectable in the blood, and seroconversion. This is a key period for diagnosis for multiple reasons, including individual and public health benefits.

Acute infection is characterized by high viral load of patients, making patients highly contagious and putting them at greater risk of transmitting HIV than those with chronic HIV infection.[28] The stage of acute infection is thought to contribute disproportionally to HIV transmission. In Uganda and Zimbabwe, cervical HIV viral loads in women were found to be highest during acute infection; however, this study was unable to address any long term changes in risk of transmission.[69] A study conducted in 2004 investigated the probability of heterosexual transmission of acute HIV in men, and found that men with average HIV semen concentrations could be expected to infect up to 24% of female sexual partners within the first two months of infection.[70] Another study conducted in Rakai, Uganda, found that heterosexual risk of HIV transmission was highest during early and late-stage HIV infection, and found that rates of transmission were 12 times higher in the first 2.5 months following infection compared with prevalent infections.[11] However, this study was limited by self-reported frequency of sexual intercourse, and was limited in generalizability by only including sero-discordant couples (one HIV positive partner, the other negative), many of whom were monogamous.[11] The data from this study were further analyzed using more robust modelling, and primary HIV infection was found to be 26 times more infectious than asymptomatic infection.[9] Research in other settings continues to provide evidence of high risk of HIV transmission during early and acute infection, with a study in Quebec estimating that early infection accounts for approximately 50% of forward transmission,[8] and a US study estimating that 11% of new HIV infections transmitted sexually are from those acutely infected who are unaware of their HIV status.[10] It should be noted that the results from these studies are largely dependent on the patient populations included, as rates of transmission during acute infection may be different in vaginal compared to

anal intercourse, and sexual behaviours and networks may differ between populations; however, it is evident that acute HIV in general plays an important role in transmission.

Knowing that acute HIV infection represents a potential public health target for slowing the rate of HIV transmission, acute diagnosis is often missed. Firstly, there is a lack of rapid diagnostic tools that can accurately detect acute HIV in resource-limited settings, and in high-income settings where patients prefer rapid testing.[65] Secondly, clinical symptoms of acute HIV infection are non-specific and frequently overlooked. Symptoms occur in an estimated 50-90% of patients, including fever, rash, nausea, or vomiting;[16] symptoms which can be easily mistaken for the flu or mononucleosis.

Recently, a well-designed systematic review and meta-analysis explored the accuracy of signs and symptoms in identifying acute HIV infection, and the authors concluded that there is limited usefulness of physician assessment of symptoms for diagnosing acute infection.[16] Additionally, a qualitative study investigated patient knowledge about acute HIV infection and its symptoms, and found that most patients who experienced acute retroviral syndrome were unaware that their symptoms were indicative of acute infection until after they were diagnosed with HIV.[71] In addition to the difficulty in tracing or picking up acute infection, its clinical presentation is often confused with malaria in malaria-endemic countries, causing patients to seek care for malaria when in fact they are experiencing acute retroviral syndrome.[39] In coastal Kenya, acute HIV was found to be equally common as malaria in young adults seeking care for fever.[39] It is therefore not surprising that patients in these settings do not receive or seek HIV testing, as fever (pyrexia of unknown origin) is often misdiagnosed, as evidence indicates.

Recent research on adults seeking healthcare for fever at outpatient clinics in sub-Saharan Africa was systematically reviewed; the authors found that only 2 out of 43 studies investigating fever tested all patients for HIV, and in 13 out of the 43 studies there were no HIV tests performed at all. However, this systematic review may have missed relevant studies by using too specific of search terms, and the study authors failed to perform quality assessment on individual studies.[72] In sub-Saharan Africa, because there is no POC screening for acute HIV, algorithms based on clinical symptomology have been created to determine risk scores for targeted testing for acute HIV infection.[73,74] One study reported success in finding acute cases using a risk score,[74] however a different study in South Africa noted that almost 30% of patients showed no clinical symptoms of acute infection and tested negative on third generation POC tests, meaning that they would have been missed using this clinical algorithm.[73] The results from these studies point to a gap in diagnostics; a risk score algorithm for targeted testing is only useful in patients showing symptoms for acute HIV. A POC test which could detect these asymptomatic acute cases offers the potential to bridge this gap, removing the need for more expensive tests which must be used selectively.

Among all the evidence supporting the importance of acute HIV detection, it becomes critical to ask whether or not there are any true clinical or public health benefits of early diagnosis; we know that transmission is high during acute HIV infection, but does knowledge of a positive HIV result have any impact on future transmission? In terms of behavioural changes, the evidence is promising. One meta-analysis provided evidence that the prevalence of risky behaviours (unprotected anal or vaginal intercourse with at-risk sexual partners) was lower in HIV positive individuals who were aware of their HIV status compared to HIV positive individuals unaware of their status.[75] However, sexual behaviour was self-reported in all studies included, and those

who are HIV positive (and aware of their status) may underreport risky sexual behaviours, overestimating the results. In addition, unprotected anal or vaginal sex does not necessarily coincide with rates of HIV transmission, due to a range of biological factors such as viral load.[75] Another meta-analysis found that after counselling and testing, HIV positive individuals and HIV sero-discordant couples had higher condom use and less unprotected sex than HIV negative or untested participants.[76]

Evidence on condom use is conflicting, as another study assessed behaviour changes following diagnosis of acute/early HIV infection, and found decreases in number of partners and unprotected sex, but no change in condom use.[77] Whether or not these behavioural changes are maintained in the long-term is a concern, with one study's results supporting long-term stability among sexual behaviour changes in MSM,[13] and another study finding no change in STI incidence (used as a biomarker for risky behaviour) among young MSM before and after acute HIV diagnosis; however this may have been underestimated due to lack of standardization of STI screening rates for the study participants.[78] One study conducted in Malawi assessed whether HIV transmission risk behaviours could be reduced in a motivational-interviewing based counselling intervention compared to brief educational HIV sessions, and no difference existed between the two study arms.[79] However, this study was in fact the first to report findings from a behavioural intervention to prevent forward transmission of HIV, and study participants found the motivational-interview based counselling to be highly acceptable.[79] Overall there appears to be potential for change in behaviour following early HIV diagnosis, but behaviour change is not the main goal of early detection. We hope that early treatment following detection will also have clinical and public health benefits; however, there is a lack of consensus in the scientific and clinical community regarding any long-term health benefits of initiating ART during acute

infection, and when is the optimal time for ART initiation.[15,17] The rationale for treating acute HIV infection centres on treating symptomatic patients, in whom HIV can progress more quickly; maintaining CD4+ count and reducing the viral set point (stabilized viral load after acute infection); reducing latent viral reservoir (which can cause viral rebound if treatment is stopped); and preserving the immune response.[17]

Many of the studies investigating potential health benefits to initiating ART during acute infection have been observational, and results of these studies are conflicting. For example, an observational cohort study showed that ART initiated during acute infection failed to prevent spreading of the virus in the body, but it seemed to prevent shedding of the virus in the genital tract, which has implications for transmission.[80] QUEST is a large prospective cohort study (n=148 acute subjects) evaluating ART during primary infection, including subjects from eight European countries, plus Canada and Australia, and found ART initiation was associated with improved CD4+ counts and lower viral load.[81] Two cohort studies were published in 2006, one in Germany (n=20),[82] and the other in Canada and the US (n=395 acute and early patients),[83]and they investigated beginning HAART (highly active antiretroviral therapy) in acute infection and then stopping, and had contrasting results: the Canada/US study found that there was reduced viral load and improved CD4+ count at six months compared with no treatment;[83] the German study found no differences in viral load or CD4+ count compared with the untreated group.[82] The method of interrupting treatment is thought to limit drug exposure and toxicity, and prevent development of drug-resistant virus.[84] Both studies were limited by self-selection by participants to their treatment group, and small sample sizes, emphasizing the need for larger, randomized interventions.

The first placebo-controlled randomized trials showed benefits such as improved CD4+ cell count six months after treatment during acute infection, but no change in viral load;[85] and improved CD4+ cell counts and reduced HIV-driven opportunistic infections after six months of ART.[86] The SPARTAC trial published in 2013 found that ART initiated for 48 weeks during primary infection slowed disease progression compared to no treatment during primary infection.[87] These results may be limited by the duration of follow up, and longer trials are needed to learn about the long-term benefits of initiating treatment during acute infection.

Rationale against starting ART during acute infection is based on concerns about early exposure to drug toxicity, worries that non-adherent patients may develop early drug resistance, and cost limitations in low-income countries; however as more evidence emerges, the benefits seem to outweigh the risks.[15,17,84,88]

## 2.7 The Determine HIV 1/2 Ag/Ab Combo Rapid Test

On August 9th, 2013, the FDA approved the first fourth generation rapid test, intended to be used as a POC device to aid in the diagnosis of HIV, including acute infection: The Determine HIV 1/2 Ag/Ab Combo rapid test (Alere).[31] The Determine Combo RT is a qualitative, point-of-care, lateral flow strip test which provides results within 20 minutes, and is able to distinguish p24 antigen and HIV antibody results.[89] This test has become widely used in various countries in North and South America, Europe and Africa; however its diagnostic performance has been contested. Individual trials evaluating the performance of this device have been summarized and will be discussed in Chapter 4 through a systematic review. Before any conclusions or recommendations could be made regarding this rapid test, we quantitatively explored the evidence from these studies using statistical models for meta-analysis.

# CHAPTER 3: METHODS

## 3.1 Data sources and searches

We conducted this systematic review and meta-analysis in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.[90] A medical librarian from the McGill University Health Centre (BN) was consulted for the design of the search strategy; to systematically search and retrieve all studies which evaluated the Determine Combo test. The search strategy was developed initially for Medline via OvidSP and was reviewed by two other librarians. The search was then adapted for the following databases: Embase, Biosis Previews, The Cochrane Library (including CENTRAL), PubMed (limited to records "as supplied by publisher" which would not have been captured in the Medline search), LILACS (Literatura Latino Americana em Ciências da Saúde), and African Index Medicus (AIM). The search strategies were designed to retrieve the test Determine and synonyms in all appropriate fields, and a combination of MeSH terms and text words in appropriate fields to capture acute, early and incident HIV infection. No publication year limit was used. The searches were run June 27, 2014, and updated on January 7th 2015 in Medline and Embase, the two largest databases. Relevant conference abstracts were searched through Embase and BIOSIS; bibliographies in relevant primary studies, narrative reviews, editorials, citations from selected studies were hand-searched. In the case of non-availability of full-text articles, abstracts were only included if they provided sufficient information. Authors were contacted for additional information as necessary. Both English and Non-English articles were included. (Please refer to the Appendix for the complete search strategy in Medline).

### 3.2 Study selection and eligibility criteria

Titles and abstracts of all records were screened independently by two reviewers (MS and RV). Studies were included if they met the following eligibility criteria: evaluated the Determine HIV 1/2 Ag/Ab Combo Rapid Test (Alere) against a reference standard, and reported test accuracy measures such as sensitivity and specificity (or raw cell values such as true positives, true negatives, false positives, false negatives).

Studies were excluded if Determine was evaluated in children or infants only, as this review focused on the adult population. Studies were also excluded if no comparison or reference standard was used, if the reference standard was not performed on all samples, or if there was insufficient information for test accuracy calculations. Duplicated data were not included; in cases of duplicate publications, the most complete record was included. Study inclusion was not restricted by country, study design, or patient populations.

Full-text articles of all titles or abstracts which were deemed relevant by either reviewer were retrieved, and independently assessed for eligibility by two reviewers (MS and RV). Disagreements between reviewers were resolved through discussion, and if consensus was not achieved, through consultation of a third reviewer (NPP).

### 3.3 Data extraction and quality assessment

A data extraction form containing sheets for study characteristics, quantitative data, and the QUADAS-2 tool for study quality assessment was created in Microsoft Excel. The data extraction form was pre-piloted by the review team before beginning data extraction. Two reviewers (MS and RV) independently extracted data and performed quality assessment for each study using the QUADAS-2 tool (Quality Assessment of Diagnostic Accuracy Studies) which

consists of four domains: patient selection, index test, reference standard, and flow and timing.[91] Quality is assessed in each domain in terms of risk of bias, and concerns regarding applicability. The signaling questions used in this tool to determine risk of bias were tailored for this review. Summary "quality scores" were not produced as this is not recommended for QUADAS-2; due to the subjective nature of weighting each item of the scale.[91,92] Instead, quality assessment was summarized graphically and descriptively.

Disagreements between reviewers (MS and RV) were resolved through discussion, and when consensus could not be achieved, in consultation with a third reviewer (NPP). Each study was assigned a unique ID number, and the following information was extracted: authors, year of publication, study objectives, study design, country, patient population, high or low risk of HIV in the population, HIV prevalence in sample, reference standards, total sample size, outcomes reported, conflict of interest, blood sample type, proportion of antigen positive subjects in sample, panel or patient population, any exclusions (i.e. due to invalid/indeterminate results), sensitivity and specificity (overall, and antibody and antigen components separately), and raw cell values (true positives, false positives, true negatives, false negatives) if provided.

### 3.4 Data Synthesis: Outcomes

The main outcomes of this meta-analysis were sensitivity and specificity. Data were collected from each study to complete the following 2x2 table:

| | | Reference Standard | |
|---|---|---|---|
| | | **Positive** | **Negative** |
| **Index test (Determine)** | **Positive** | True positive (TP) | False Positive (FP) |
| | **Negative** | False negative (FN) | True Negative (TN) |

Where:

$$\text{Sensitivity} = \frac{TP}{TP + FN}, \qquad \text{Specificity} = \frac{TN}{TN + FP}$$

Random effects models were used to synthesize data. Fixed effects models assume that there is a single common effect for all studies; in contrast, random effects models assume each study produces a different effect, and that these effects can be summarized by a probability density across studies. We opted to use a random effects model since a fixed effects model is not plausible in our context. We first pooled the data using a model which assumed a perfect gold standard was available in each study, allowing sensitivity and specificity to be pooled separately.

There are some caveats with respect to the field of HIV to be kept in mind while deciding on the models, their interpretations and their implications. There is no accepted perfect gold standard for HIV; however HIV tests currently in use generally have high accuracy, even rapid tests.[25,26] The CDC testing guidelines consist of initial screening with fourth generation assays, and confirmation using NAAT and antibody differentiation assays.[33] If these testing guidelines are used as a gold standard definition, then most of reference standards used in the studies we included in our meta-analysis do not meet this requirement. Different reference testing

algorithms and pathways were used in each study, and were largely determined by the context in which the study was conducted (i.e. following standard testing procedures at a clinic, or national guidelines for a specific country, etc.); there was little consistency between studies in terms of testing algorithms. Additionally, the index test (Determine Combo) has two components which provide different results (antigen and antibody component), which further complicates the pathway for reference standard testing; as most reference tests do not test for antigen and antibody simultaneously. The effect of imperfect reference standards can be uniquely explored using Bayesian methods,[93] however we did not have enough data for this type of exploration; thus we assumed each study had a perfect reference standard and we analyzed sensitivity and specificity separately.

Because there were differences in reference standard algorithms between studies, we compared this model to a second model which allowed sensitivity and specificity to be correlated. This second model, using frequentist methods, assumed that the thresholds for test positivity varied between studies. The two models are further described in detail below.

### 3.5 Model 1: Perfect gold standards

In the first analysis each reference standard was treated as a gold standard, and sensitivity and specificity were analyzed separately. Sensitivity was defined as the number of samples which correctly tested positive using the index Determine Combo test (true positives, considered "successes" for positive samples) over the total number of samples identified as positive by the reference standard (true positives and false negatives). Specificity was defined as the number of samples which were correctly identified as negative by the Determine Combo test (true negatives, considered "successes" for negative samples) over the total number of samples which were identified as negative by the reference standard.

A Bayesian hierarchical random effects model was created to synthesize sensitivity and specificity across studies. For the first level of the hierarchical model we assumed that the number of successes for each study ($X_i$) followed a binomial distribution, with sample size $N_i$, and probability of success $P_i$, allowed to vary by study.

**Likelihood function**:

$$X_i \sim \text{binom}(N_i, P_i), \quad i = 1, 2, 3, \ldots I$$

The second level of the hierarchical model allowed the logit transformed probability of success ($P_i$), to follow a normal distribution, with mean sensitivity/specificity across studies ($\mu$), and variance $\sigma^2$ (indicating the between-study variability).

**Hierarchical function**:

$$\text{logit}(P_i) \sim N(\mu, \sigma^2)$$

Since we did not have prior knowledge about the parameters of the second level of the hierarchical model beyond the data included, very wide non-informative prior densities were used for the mean ($\mu$) and the variance ($\sigma^2$) from the second level of the hierarchical model, so that inferences were determined by the data rather than any prior information.

**Prior densities:**

$$\mu \sim N(0, 10^6)$$

$$\sigma \sim \text{Unif}(0, 5)$$

Since the Determine Combo test has two components which provide distinct results, the p24 antigen component and the HIV antibody component, we ran three different analyses. First, we

pooled data from all studies which assessed the "overall" accuracy of the Determine Combo test, meaning any positive result on either the antigen or antibody component was considered as a "success" for sensitivity, and a negative result on both the antigen and antibody components was considered a "success" for specificity. Second, we pooled data from all studies which assessed the antibody component on its own. Third, we pooled data from all studies which assessed the antigen component on its own.

Furthermore, we explored heterogeneity through various subgroup analyses. We ran a subgroup analysis for blood sample type (plasma/serum and whole blood samples), as we hypothesized that plasma or serum samples might have higher accuracy with the Determine Combo test than whole blood samples. Whole blood samples are more practical than serum or plasma specimens for use at point of care as they can be collected via finger stick.[94] We also performed subgroup analyses by study design, as all studies fell into two categories: case-control studies, where samples were selected based on HIV status, and prospective cross-sectional designs, where samples were taken as patients with unknown HIV status enrolled in the study. We expected case control studies to overestimate test accuracy parameters in comparison to cross-sectional studies, as case control studies tend to recruit patients at extremes of the disease spectrum, leading to spectrum bias. In addition to analyzing the effects of blood sample type and study design, we had planned to explore reference standards by subgroup (perfect vs. imperfect reference standards) however the data available limited our ability to explore this effect statistically. Instead, forest plots were created to visually assess the effect of reference standards.

All models were run in WinBUGS (version 1.4.3, MRC Biostatistics Unit, Cambridge UK).

### 3.6 Model 2: Continuous threshold for positivity

In Model 1, we made the assumption that a perfect gold standard was available in each study. This allowed sensitivity and specificity to be modeled separately, as we assumed no correlation between true positive and true negative data. However, because this was an assumption, we wanted to check the robustness of our results across a different set of assumptions; so we explored the extent to which the pooled estimates obtained through the perfect gold standard model might differ if we assumed that each study had a different underlying distribution for positivity. This would mean that different studies had different thresholds for classifying samples as positive or negative. There is no arbitrary cut-off for an HIV diagnosis, either an individual has HIV or they do not; however, even the most sensitive nucleic acid amplification tests can only detect RNA above a certain threshold, and limits of detection vary between different reference standard algorithms. For the purposes of comparing models, in Model 2 we assumed the reference standards included in this meta-analysis followed a continuous rather than dichotomous distribution, and each reference standard has its own cut-off point for a positive vs. negative diagnosis. This means that there is a trade-off between sensitivity and specificity in each study, depending on the threshold for positivity of the reference standard used.

A bivariate hierarchical logistic regression model was fitted to synthesize sensitivity and specificity across studies, allowing sensitivity and specificity to be correlated within each study. The model was generated in Stata statistical software (Version 13, StataCorp, College Station, TX: 2013), using the "metandi" package.[95] In the first level of the hierarchical model, the successes (true positives for sensitivity, true negatives for specificity) are distributed binomially:

Sensitivity (A): $X_{Ai} \sim \text{binom}\ (N_{Ai}, P_{Ai})$

Specificity (B): $X_{Bi} \sim \text{binom}\ (N_{Bi}, P_{Bi})$

In the second level, the logit transformation of the probability of success ($\theta_A$ for sensitivity, $\theta_B$ for specificity), has a bivariate normal distribution across studies, allowing for correlation between sensitivity and specificity by including an additional parameter, $\sigma_{AB}$.[95]

**Second level:**

Where $\theta_A = \text{Logit}\ (P_{Ai})$, and $\theta_B = \text{Logit}(P_{Bi})$

$$\begin{pmatrix}\theta_{Ai}\\ \theta_{Bi}\end{pmatrix} \sim N\left(\begin{matrix}\theta_{\mathbf{A}}\\ \boldsymbol{\theta}_{\mathbf{B}}\end{matrix}, \boldsymbol{\Sigma}\right) \textit{with}\ \Sigma_{AB} = \begin{pmatrix}{\sigma^2}_A & \sigma_{AB}\\ \sigma_{AB} & {\sigma^2}_B\end{pmatrix}$$

Similarly to the analysis in Model 1, we ran the Model 2 analysis for "overall" sensitivity and specificity, as well as for the sensitivity and specificity for the antigen and antibody components separately. Heterogeneity was explored through subgroup analysis by blood type sample, and study design.

Forest plots were created to visually represent individual study results with their confidence intervals using Review Manager (RevMan Version 5.3. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2011).

# CHAPTER 4: RESULTS

## 4.1 Descriptive Results

Through searching the seven databases, 1091 articles and 76 conference abstracts were identified, producing 555 unique records to screen after duplicates were removed. The full-text was retrieved for 47 articles, with 16 articles meeting the eligibility criteria and selected for inclusion. The search was updated in January 2015, producing 55 new results to screen, with 1 additional article selected for inclusion in the systematic review and meta-analysis. A flow diagram illustrating the process of study selection can be seen in Figure 3.

A total of 17 studies were selected for final inclusion, which reported diagnostic accuracy of the Determine Combo rapid test against a reference standard.[66,96-111] Demographic characteristics of all included studies can be found in Appendix Table 1. 10 studies were case-control design, and 7 were cross-sectional. Whole blood samples were collected in 7 studies, with either plasma or serum samples collected in the remaining 10 studies. Three studies (Patel et al, Laperche et al, and Bhowan et al) each contained more than one separate, unique evaluation of Determine, and in these cases each evaluation was treated independently.[66,97,105] Individual study estimates with their confidence intervals are represented through forest plots as seen in Figure 4. Raw data extracted from each individual study can be seen in Appendix Table 2.

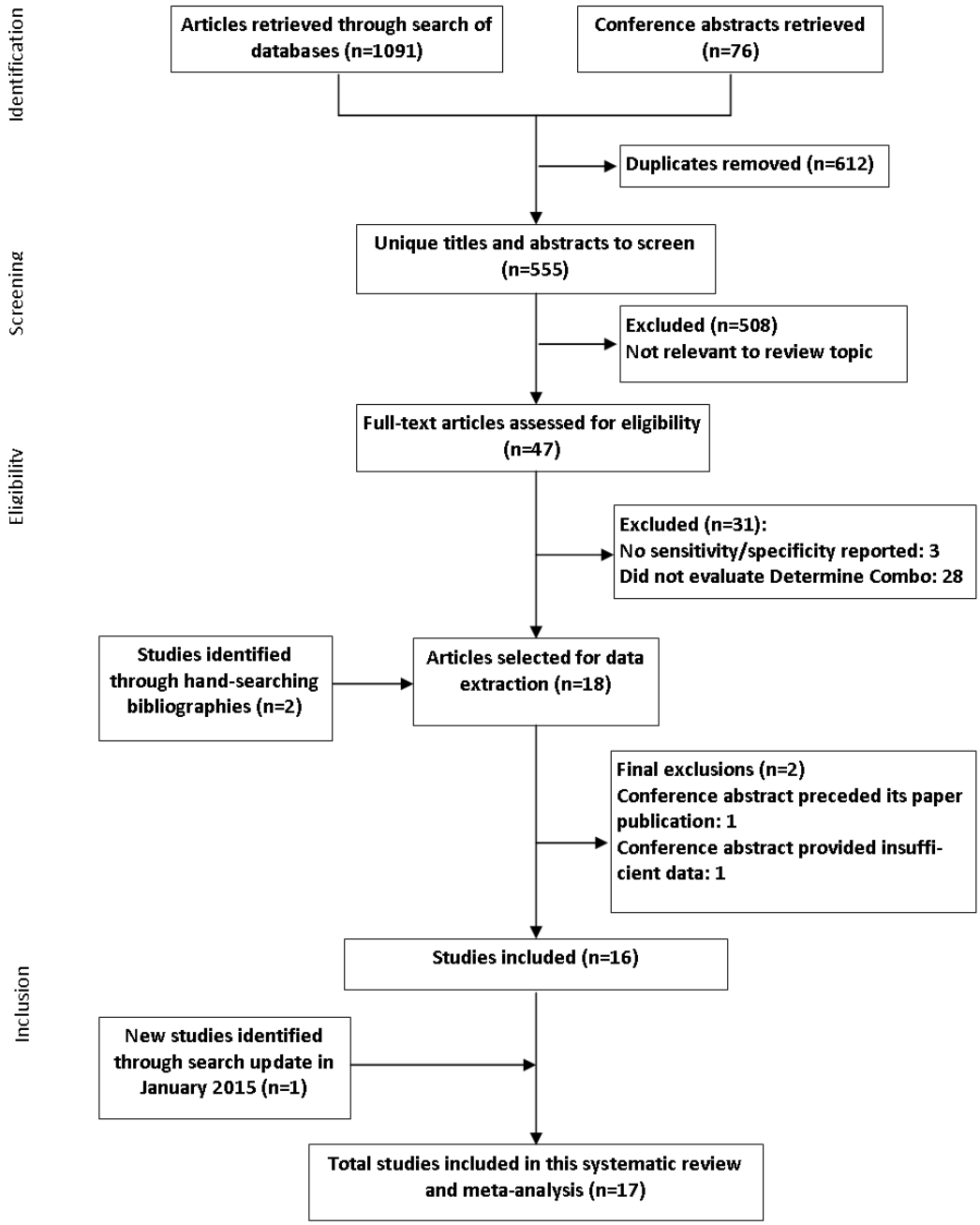


**Figure 3** PRISMA flow diagram of study selection

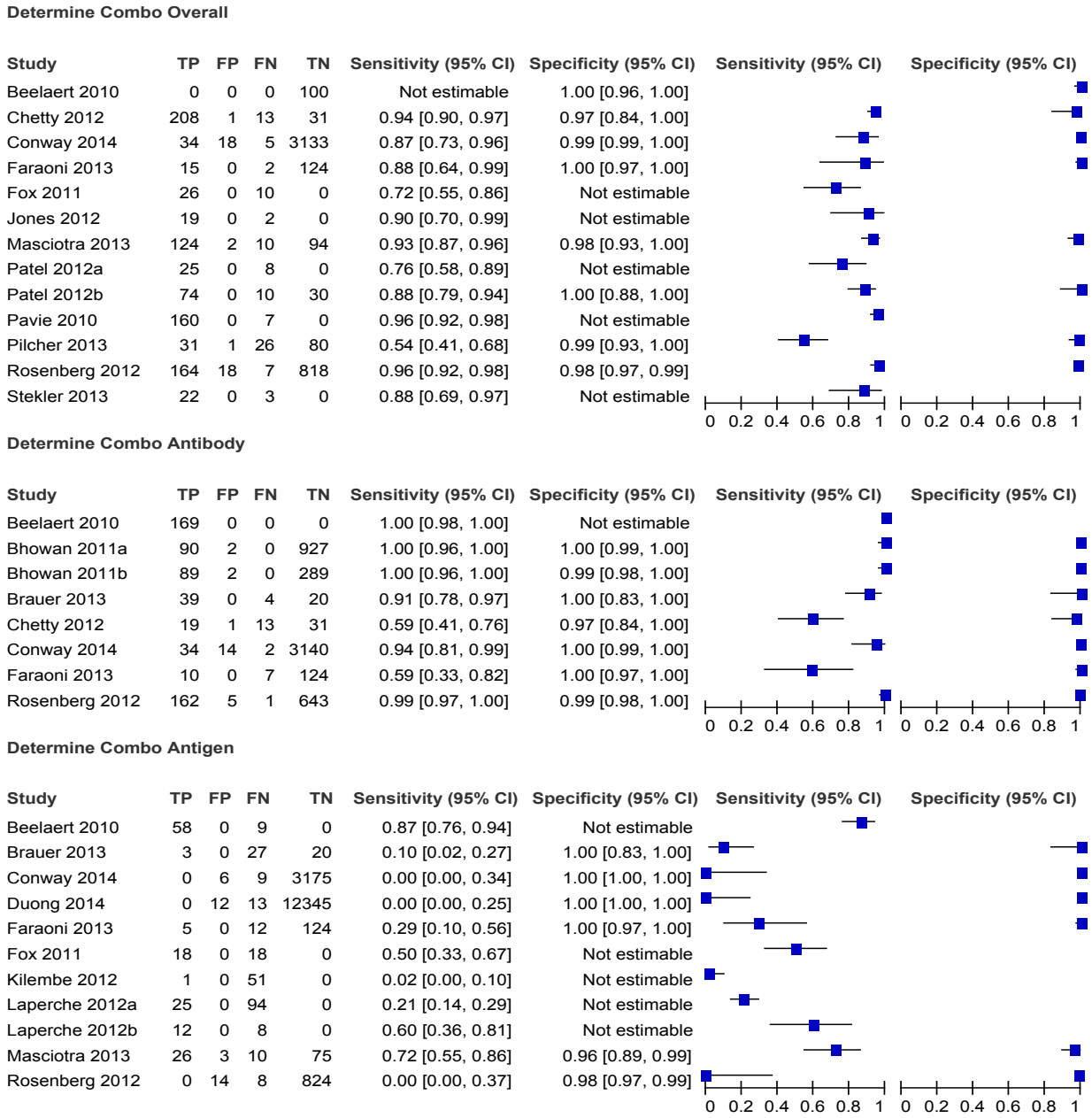


**Figure 4** Forest plots showing performance of the Determine Combo test in individual studies.
It should be noted that missing data (for TP, FP, FN, TN) were treated as having a value of zero by RevMan software; this did not affect the sensitivity and specificity estimates, however please refer to Appendix Table 2 for raw cell values extracted from each study.

## 4.2 Systematic review of studies evaluating Determine HIV 1/2 Ag/Ab Combo Rapid Test

As of January 2015, there have been 17 studies which have evaluated the diagnostic performance of the Determine Combo test against a reference standard, with varying results. The following studies have been summarized and critiqued using the QUADAS-2 tool and STARD checklist for quality assessment and reporting of diagnostic accuracy studies; [91,112] summary results can be found in Figure 5 and Appendix Table 1.

The first studies evaluating the Determine Combo test were published in 2010, in Belgium, France and the United Kingdom. Beelaert and Fransen evaluated the Determine Combo test in Belgium, using a case-control study design and well-characterized HIV specimen panels (n=336).[96] The HIV specimen panels consisted of both archived frozen specimens, and fresh, non-frozen specimens, and specimens were serum, plasma, or whole blood. The Determine Combo test was found to have high accuracy overall and for the antibody component, and good accuracy for the antigen component; however the antigen component was still lower than the sensitivity of laboratory fourth generation assays. There were many limitations of this study. Using a case-control design introduces selection bias by selecting samples based on HIV status, and using a well-characterized specimen panel likely overestimated the test accuracy, as specimens were at extremes of the disease spectrum (spectrum bias). Additionally, not all samples were tested using the same reference standard. The reference standard testing algorithm incorporated two initial EIA's, however different assays were used for different samples; most were third generation EIA's but there was one fourth generation assay used as well. This study was at high risk for verification bias, which occurs when the choice or use of reference standard depends on the results of the index test (or vice versa). In this study, if results of the Determine

Combo test did not correlate with reference standard results, the samples were tested again with the Determine Combo test, likely overestimating test accuracy. This study was funded by Alere (the company which manufactures the Determine Combo test), representing a potential conflict of interest. However, an advantage of using panel specimens in this study was that the investigators were able to incorporate multiple HIV subtypes, finding that the Determine Combo test could detect most but not all HIV subtypes.

That same year, in France, Pavie and colleagues evaluated the Determine Combo test along with four other rapid tests, using whole blood samples from patients with documented HIV infection (n=187).[107] The individual antigen and antibody components of the Determine Combo test were not evaluated separately, but the overall test accuracy was found to be very high. However, the blood samples were taken from patients with established infection, many of whom were receiving treatment to control their viral loads; only two patients had recent HIV infection.[107] This means many of these patients would be unlikely to have p24 antigen detectable in their blood, meaning the results of this study cannot be generalized to populations with acute HIV. As this was a case-control study, this study was at risk of selection bias based on patient selection. Another source of bias is the large number of invalid tests (33 out of 220) which were excluded from the final results. Pavie and colleagues hypothesized that the large number of invalid tests may have been due to the use of whole blood samples hindering migration on the test strip, or binding of antigen or antibody to the strip; however, this also raises concerns that the individuals conducting the test were not adequately trained.

The third evaluation of the Determine Combo test was published at the end of 2010, by Fox and colleagues in the United Kingdom using a case-control design to evaluate the sensitivity overall, and of the antigen component (n=36).[102] Both overall and antigen component sensitivity were

found to be low, which was different than the two studies published previously. One of the main differences, however, was the case mix of the samples tested: this study used stored serum samples which were positive for p24 antigen. Given that the antigen sensitivity was found to be low in this study, and given the very short window for p24 antigen to be detected it was unsurprising that the overall sensitivity was found to also be low, as there were very few additional specimens which were also antibody positive. Similar to the previous two studies, this study was at risk of selection bias due to the case-control design; additionally, this study was at risk of detection (or reviewer) bias because the index test (Determine Combo) was not interpreted independently of reference standard results (i.e. it was known that all samples were antigen positive, and no negative samples were included). While the results of this study were disappointing, it should be noted that the sample size was very small (n=36) leading the study authors to call for more evaluations.

In 2011, Bhowan and colleagues conducted the first cross-sectional field evaluation of the Determine Combo test, which also happened to be the first evaluation in a high-prevalence African population.[97] This study evaluated the accuracy of the antibody component of the Determine Combo test in South African women attending an antenatal clinic (n=1019). Accuracy was evaluated in both whole blood and plasma samples, and in antenatal and post-partum women; antibody accuracy was found to be high in all groups.[97] This study improved over previous studies by using a cross-sectional study design, minimizing risk of spectrum bias. Another strength was evaluating the Determine Combo test on both plasma and whole blood samples. A major limitation of this study was the reference standard algorithm: third generation rapid tests were used, with a fourth generation immunoassay only used when third generation rapid testing results were discordant, introducing verification bias. The comprehensiveness of

this study was limited, as only the antibody component could be tested; it was unknown whether any samples which tested negative via third generation rapid testing were actually p24 antigen positive (acute infection). The applicability of the findings are restricted to pregnant South African women, and may not be applicable to men and non-pregnant or post-partum women.

In 2012, interest and excitement surrounding the Determine Combo test began to expand, with many more studies appearing. Patel and colleagues conducted a case-control study evaluating the Determine Combo test during early infection in the US using stored specimens from the CDC Acute HIV study (n=33), and a blinded specimen panel (n=114).[66] Overall sensitivity was found to be low in the CDC specimens, but higher in the blinded specimen sample, which is not surprising since the blinded specimen panel was likely subject to spectrum bias. This study in particular was interesting because it compared the performance of the Determine Combo test to other rapid tests and a fourth generation lab immunoassay, finding that while the performance of the Determine Combo test was lower than the fourth generation assay, it still performed above all the other rapid tests in detecting early HIV infection. This finding meant that the Determine Combo test could potentially replace second and third generation rapid tests in resource-limited settings to improve early HIV detection. Unfortunately, this study was also subject to bias by using a case-control design; selection bias may have further been introduced since only half of the CDC samples had sufficient blood for index testing, due to the large quantity of rapid tests being conducted on a single blood sample. Strengths of this study were that NAAT was used as a reference standard, which has a higher sensitivity than third generation tests used in previous studies; and index test interpretation was independent of the reference standard results, minimizing risks of incorporation and verification biases.

The next cross-sectional field evaluation was conducted by Rosenberg and colleagues in Malawi (n=844).[109] This study evaluated the antigen component in patients with acute infection, and the antibody component in patients with established infection; the investigators also evaluated an ultrasensitive p24 antigen assay. The Determine Combo test performed well in identifying antibodies, however it failed to detect any of the 8 acute infections. It should be noted, however, that the study authors defined acute infection as having negative or discordant rapid antibody test results but detectable RNA; when the ultrasensitive p24 antigen was tested on the eight acute samples it only detected antigen in five out the eight, suggesting that three of these samples may have not been p24 antigen positive yet. This highlights how the definition of acute HIV infection is reliant upon the diagnostic technology available, which will be discussed later in this thesis. The authors of this study concluded that the Determine Combo test was not adequate for detection of acute HIV infection, but that it performed just as well as its third generation counterpart. This study was at risk of selection bias, because patients were selected based on their risk for acute HIV infection (having negative or discordant antibody rapid test results). This method of patient selection may have unintentionally biased the interpretation of the index test results, leading to detection bias.

Chetty and colleagues performed a cross-sectional field evaluation of the Determine Combo test in South African women, for early detection of HIV during pregnancy (n=253).[99] The study found high accuracy overall, particularly in comparison to third generation rapid tests at baseline, but low accuracy of the antibody component. The antibody component was evaluated in seroconversion specimens, meaning women who tested negative at baseline for antibodies, but tested positive at the next visit. While interesting to examine the seroconverting population, there was a small sample size in this category (n= 32). Similarly to the study by Bhowan and

colleagues, this study may not be generalizable to populations other than pregnant women. Other limitations included that the investigators were unable to assess performance of the antigen component, and that the initial screening assay used was a third generation rapid test (rather than a more sensitive lab assay), introducing imperfect reference standard bias which can underestimate results. Additionally, although this was a cross-sectional study, only a sample of seronegatives and established infections were selected to be tested with the Determine Combo test, which may have introduced selection bias.

Another case-control study was conducted, which assessed the performance of the Determine Combo test on panel specimens from Zambia and Rwanda, by Kilembe and colleagues (n=82).[104] This study found very low antigen sensitivity, however no proper estimates for antibody or overall sensitivity were provided. The investigators used a third generation rapid test and a p24 immunoassay as reference standards, however they found some samples to be false positive by the reference standard, indicating that their reference standard was imperfect, further confusing their results and calling into question the validity of their findings. Imperfect reference standard bias can lead to underestimation of test accuracy. The accuracy results were poorly reported, as they were split up into five groups based on antigen/antibody positivity, failing to provide a clear, concise estimate. These study authors were the second to describe difficulties in interpreting the test results; however faint positive lines were the concern rather than a missing control line. A strength of the study was the patient sample, as it included individuals with non-B subtypes of HIV, which may also have played a role in the accuracy; however this effect was not explored in this study.

Eight studies had been published by the end of 2012, with little consistency in methodology (particularly in reference standard algorithms) and conclusions regarding the diagnostic

performance of the Determine Combo test. There was concern that the antigen component was not strong enough to detect acute infection, and Laperche and colleagues investigated the performance of the antigen component only, against a p24 antigen specific immunoassay reference standard (n=139).[105] The study investigators evaluated both archived specimens from acutely infected individuals, an addition to various dilutions of cell-derived supernatant samples of antigen positive HIV subtypes. Antigen sensitivity was low in both types of samples, and particularly low in non-B HIV subtypes. The authors concluded that the Determine Combo test performs poorly in detecting acute infection; however no information on the overall performance of the Determine combo test was given since the samples were restricted to antigen positives only. This study wasn't able to provide any more insight into how this test might perform in real-life settings, as it was a case control design with panel specimens, once again introducing selection and spectrum biases. Similarly to the study by Fox and colleagues, this study was at risk of detection bias, since it was known that all samples were p24 antigen positive before interpreting index test results.

In contrast to the previous study, Jones and colleagues conducted a cross-sectional field evaluation of the Determine Combo test among high-risk patients attending a clinic in the UK (n=985).[103] They found the Determine Combo test to have good overall sensitivity and specificity, but the antigen and antibody components were not evaluated separately. The sample size was not large enough to capture sufficient antigen positive subjects to determine accuracy during acute infection, which emphasizes the difficulty in evaluating these tests in acute patients, due to the short window period of acute detection. Strengths of this study included use of an appropriate reference standard (fourth generation lab immunoassay) minimizing imperfect reference standard bias; blinding of participants to the index test results minimized verification

and incorporation biases, which happens when the index test is incorporated into the reference standard testing algorithm.

In 2013, interest in the Determine Combo assay continued to grow, and Faraoni and colleagues performed a case-control study on panel specimens in Italy (n=141), in hopes of gaining insight into the performance of the Determine Combo test in acute specimens in a low-prevalence setting.[101] The overall accuracy, as well as the antigen and antibody components were evaluated separately, finding high overall accuracy, but low antibody sensitivity and poor antigen sensitivity. A major limitation of this evaluation was the low sample size, as only 17 acute specimens were evaluated (although 124 negative controls were included, allowing for specificity estimates as well), and the case-control design which introduced selection bias. A strength of this study was the fourth generation immunoassay used as a reference standard, minimizing imperfect reference standard bias. Of interest in this study was the increased overall sensitivity compared to antibody only; the authors suggested that the addition of the antigen component greatly improved the performance of the Determine Combo test, however the number of acute subjects in this sample was not representative of what would be found in real life due to the case-control design, and this is something that would need further studying.

The next two studies to be published on the Determine Combo test were case-control studies. Brauer and colleagues evaluated the antigen and antibody components separately in South African specimens (n=79), but did not report an overall estimate, finding high performance of the antibody component and low performance of the antigen component.[113] A fourth generation immunoassay and NAAT were used as reference standards, but not all samples received both, potentially introducing bias. This study was at risk of selection bias due to its case-control

design, and also at risk of detection bias as the authors failed to report whether or not results of the Determine Combo test were blinded when interpreting reference standards.

The second case-control study evaluated the ability of the Determine Combo test to detect HIV in acute/early infection in US and Ivory Coast samples (n=230).[106] Masciotra and colleagues found that the overall performance of the test was high, and while the antigen component was not as strong, it performed better than the previous case-control study. These results led the study authors to conclude that the Determine Combo test performs somewhere between a third and fourth generation immunoassay, which was only partially supported by previous evidence. In addition to selection bias, this study may have been at risk of detection bias, as it was unclear as to whether or not the results of the Determine Combo test were independent of the reference standard. There was clearly a need for more prospective, cross-sectional studies to see how the Determine Combo test performs in real-life settings.

Stekler and colleagues opted to compare the abilities of various POC tests to detect early HIV infection in a real-time cross-sectional evaluation of MSM (n=1587).[110] The overall sensitivity of the test was similar to findings from Faraoni and colleagues, and Jones and colleagues. While a cross-sectional study at this time was desired to add further evidence to the Determine combo narrative, there were some major flaws in study design that limited the inferences that can be made from the results. The POC tests were not conducted independently of reference standard results, in fact the POC results were part of the reference standard testing algorithm, putting this study at high risk of incorporation bias. This means that estimates of sensitivity for the less sensitive tests were likely overestimated, because these results (faint bands) were interpreted using the other POC test results for reference. This bias could easily have been avoided by

conducting reference standard testing independently of index testing, and excluding the index test (Determine Combo) from the reference standard algorithm.

At the end of 2013, Pilcher and colleagues published an investigation of how the choice of rapid test could impact HIV case finding in San Francisco testing programs.[108] This was another case-control study (n=138), and the overall sensitivity of the Determine Combo test was found to be low. The antibody component could not be evaluated because only antibody negative, and RNA positive samples were included. The reference standards used in this study were inconsistent between samples, since the plasma specimens were archived from various targeted testing programs which all used different reference standards. Some samples were tested for antibody using a first generation immunoassay, which have been long been replaced by second, third, and now fourth generation immunoassays, since first generation assays can take up to 60 days to detect antibody; this introduced imperfect reference standard bias, likely underestimating the results. This study was also at risk of detection bias in addition to selection bias, as the interpretation of the Determine Combo test results was not blinded to reference standard results.[47]

This brings us to 2014, where consensus had still not been reached regarding the usefulness of the Determine Combo test, and studies with flawed designs and poor methodology have produced conflicting results. Appreciating the necessity of prospective field studies for understanding the capacity of the Determine Combo test to detect acute infection, Conway and colleagues implemented a prospective cross-sectional study in public sexual health clinics with high numbers of MSM (n=3190) in Australia.[100] This study had many strengths which had been missing in previous evaluations of the Determine Combo test, including appropriate fourth generation immunoassays as a reference standard which were supplemented by antibody tests,

p24 antigen tests, and Western blot. Whole blood fingerstick samples were tested, and clinicians rather than laboratory technicians collected the samples, mimicking how the test will be used in real-life. The investigators found antibody sensitivity to be high, but antigen sensitivity to be low, as the antigen component did not identify any of the nine antigen positive samples. Similarly to the field study by Rosenberg and colleagues, this study had too few antigen positive samples to properly assess antigen sensitivity. Another limitation is whether or not this study can be generalized outside of MSM testing in clinics.

Later in 2014, Duong and colleagues conducted a large prospective cross-sectional study (n=18172) in a high HIV prevalence setting in Swaziland.[111] Sensitivity of the antigen component was found to be very low, which may have been affected by the low number of acute subjects in the sample, defined as antigen positive but antibody negative. While the large sample size, cross-sectional design, and high HIV prevalence (to increase the number of acute samples) were strengths of the study, a major weakness was the reference standard testing algorithm. All samples were first tested with the index test (Determine Combo), then based on the result of this test, samples were tested with different reference standards. Antibody positive samples were tested on a second rapid test, antigen positive samples received viral load testing, and all antigen and antibody negative samples were tested using pooled NAAT. This method may have introduced incorporation bias, as these reference standards have varying accuracy, and can overestimate test accuracy. Of note in this study was the low number of acute specimens (n=26), which again points to the challenge of detecting acute HIV in a short window period.

To conclude this descriptive review of the literature, the evaluations to date do not offer confidence in the Determine Combo test for detecting acute HIV infection, and many of the evaluations have been limited in methodology. These limitations include timing of detection,

patient populations and case mixes recruited, weak study designs, and inappropriate reference standard algorithms used, which have huge implications on the accuracy estimations of the studies. The results of quality assessment are outlined below.

### 4.3 Quality Assessment

Quality assessment was conducted for each study with the use of the QUADAS-2 tool. Figure 5 presents a graphical display of the proportion of results scoring low, high, or unclear risk of bias in each of the four domains. There were minimal concerns regarding applicability, with the exception of one study which only included patients with established infection.[107] 65% of studies reported high risk of bias in the "patient selection" domain,[66,96,98,101,102,104-109] and 65% of studies reported high risk of bias in the "index test" domain.[98,99,101-106,108-110] 13 out of 17 studies (75%) scored a "high" risk of bias in two QUADAS-2 domains or more,[66,96-98,101,102,104-110] and not a single study scored a "low" risk of bias in all domains.[107] The individual study results for QUADAS-2 assessment can be found in the Appendix Table 3.

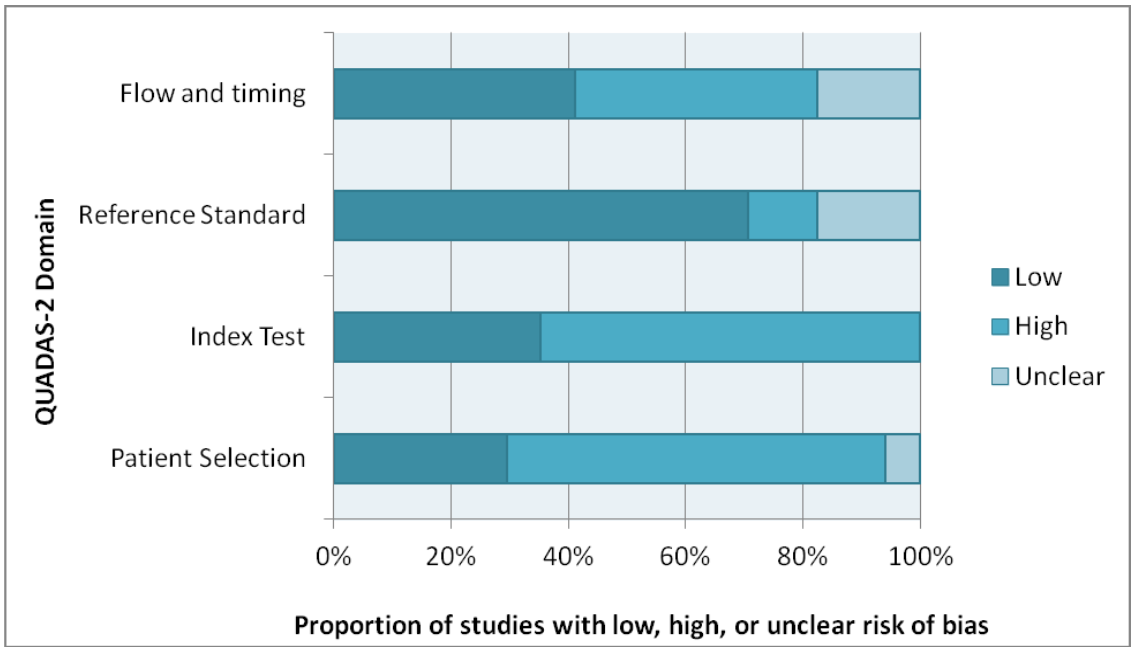


**Figure 5** Graphical representation of quality assessment using QUADAS-2.

## 4.4 Results from Model 1

The data were first synthesized using Model 1, assuming perfect gold standards; results estimating sensitivity and specificity are given in Table 2. The overall pooled sensitivity estimate for the assay was 88.5%, 95% CrI [80.1 – 93.4], and overall pooled specificity was 99.1%, 95% CrI [97.3 – 99.8]. When the individual antigen and antibody components of the assay were evaluated separately, the antigen component produced a pooled sensitivity of 12.3%, 95% CrI [1.1 – 44.2], with a pooled specificity of 99.7%, 95% CrI [96.8 – 100]. The antibody component produced a pooled sensitivity of 97.3%, 95% CrI [60.7 – 99.9], and pooled specificity of 99.6%, 95% CrI [99.0 -99.8].

To compare the diagnostic accuracy of the Determine Combo test in serum/plasma vs. whole blood samples, we performed a subgroup analysis. We found the overall pooled sensitivity in serum/plasma to be 84.3%, 95% CrI [65.0 – 94.2], and specificity to be 99.3%, 95% CrI [94.7 –

100.0]. The overall pooled sensitivity in whole blood was 93.8%, 95% CrI [84.4 – 97.3]. An estimate for pooled specificity in whole blood could not be provided by the data available.

Overall diagnostic accuracy was also assessed in subgroups by study design, with case-control studies reporting a pooled sensitivity of 84.9, 95% CrI [65.0 – 94.5] and pooled specificity of 99.6, 95% CrI [79.1 – 100.0]. Cross-sectional studies had a pooled sensitivity of 93.2%, 95% CrI [83.9 – 96.9], and specificity of 98.8, 95% CrI [37.2 – 100.0]. It should be noted that credible intervals greatly overlap for sensitivity and specificity estimates for case-control and cross-sectional studies.

**Table 2 Results from Model 1 (perfect gold standards).**
Point estimates given are posterior medians, with 95% credible intervals (CrI).

| **Overall** | **Sensitivity** | **95% CrI** | **Specificity** | **95% CrI** |
|---|---|---|---|---|
| **Total** | 88.5 | 80.1 - 93.8 | 99.1 | 97.3 - 99.8 |
| **Serum** | 84.3 | 65.0 - 94.2 | 99.3 | 94.7 - 100.0 |
| **Whole blood** | 93.8 | 84.4 - 97.3 | 98.8 | 0.00 - 100.0 |
| **Case-control** | 84.9 | 65.0 - 94.5 | 99.6 | 79.1 - 100.0 |
| **Cross-sectional** | 93.2 | 83.9 - 96.9 | 98.8 | 37.2 - 100.0 |
| | | | | |
| **Antibody** | **Sensitivity** | **95% CrI** | **Specificity** | **95% CrI** |
| **Total** | 97.3 | 60.7 - 99.9 | 99.6 | 99.0 - 99.8 |
| **Serum** | * | * | 99.8 | 83.9 - 100.0 |
| **Whole blood** | * | * | 99.5 | 93.4 - 99.9 |
| **Case-control** | * | * | * | * |
| **Cross-sectional** | 99.2 | 35.6 - 100.0 | 99.5 | 98.4 - 99.8 |
| | | | | |
| **Antigen** | **Sensitivity** | **95% CrI** | **Specificity** | **95% CrI** |
| **Total** | 12.3 | 1.1 – 44.2 | 99.7 | 96.8 - 100.0 |
| **Serum** | 27.8 | 5.9 - 67.4 | * | * |
| **Whole blood** | * | * | 99.7 | 24.7 - 100.0 |
| **Case-control** | 27.8 | 5.9 - 67.4 | * | * |
| **Cross-sectional** | * | * | 99.7 | 24.7 - 100.0 |

* These parameters could not be estimated due to limited data

Although the diagnostic accuracy across reference standards could not be analyzed statistically, refer to Figure 6 for forest plots grouped by reference standard.

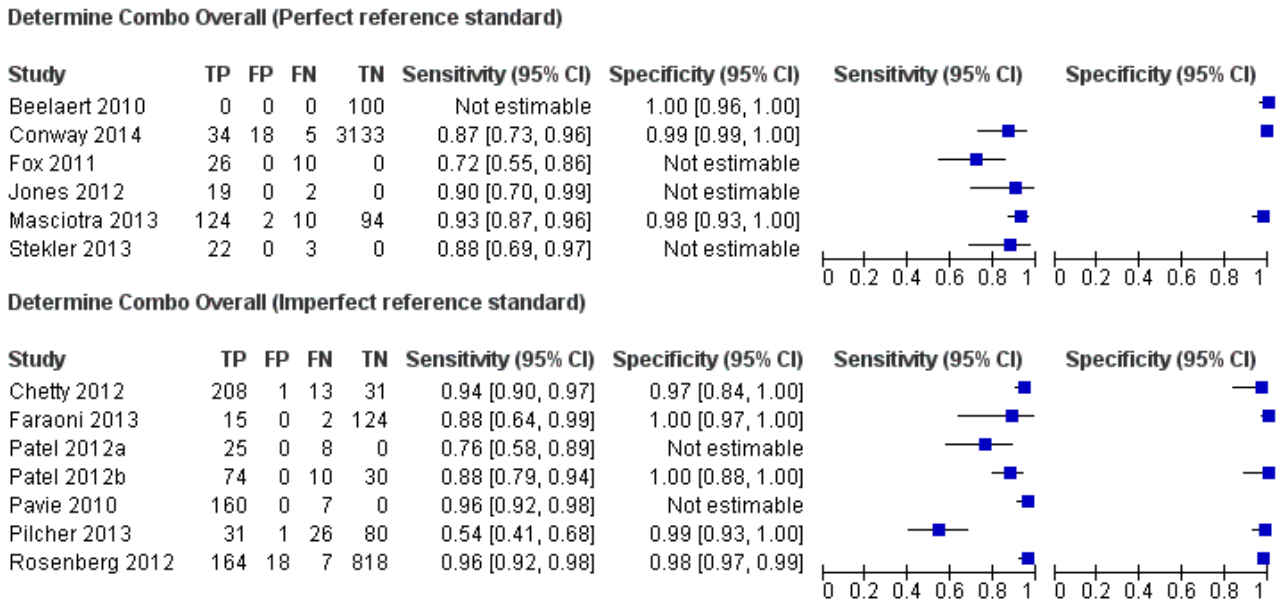


**Figure 6.** Forest plots showing the overall sensitivity and specificity of the Determine Combo test, sub-grouped by reference standard**.** Perfect reference standard was defined as screening using a fourth generation assay and/or NAAT, imperfect reference standard was defined as screening using a second or third generation assay, and/or Western Blot.

## 4.5 Results from Model 2

We then synthesized the data using frequentist methods; assuming varying thresholds between studies and allowing correlation of sensitivity and specificity within studies. Studies which provided data for all measures (i.e. true positives, false negatives, true negatives, and false positives) were included in this meta-analysis. Results for the overall, antibody, and antigen pooled sensitivities and specificities, along with subgroup analysis by blood sample and study design can be found in Table 3. The overall pooled sensitivity estimate was 88.6%, 95% CI [79.1 - 94.2], and the overall pooled specificity estimate was 99.2, 95% CI [98.3 – 99.6]. The pooled antigen sensitivity estimate was 8.9%, 95% CI [1.0 – 48.2], and the pooled antigen specificity

estimate was 99.5, 95% CI [98.3 – 99.9]. The pooled antibody sensitivity estimate was 97.4, 95% CI [69.0 – 99.8], and the pooled antibody specificity was 99.5, 95% CI [99.3 – 99.7].

When the overall data were pooled by blood sample type, the overall sensitivity in serum or plasma was estimated as 87.0, 95% CI [73.3 – 94.2] and overall specificity in serum or plasma was estimated as 98.9, 95% CI [97.0 – 99.6]. Overall sensitivity and specificity for whole blood samples could not be pooled as due to limited data.

**Table 3. Results from Model 2 (varying thresholds).**
Point estimates given are means with 95% confidence interval (CI).

| **Overall** | **Sensitivity** | **95% CI** | **Specificity** | **95% CI** |
|---|---|---|---|---|
| **Total** | 88.6 | 79.1 - 94.2 | 99.2 | 98.3 - 99.6 |
| **Serum** | 87.0 | 73.3 - 94.2 | 98.9 | 97.0 - 99.6 |
| **Whole blood** | * | * | * | * |
| **Case-control** | 84.2 | 66.4 - 93.4 | 99.1 | 97.2 - 99.7 |
| **Cross-sectional** | * | * | * | * |
| | | | | |
| **Antibody** | **Sensitivity** | **95% CI** | **Specificity** | **95% CI** |
| **Total** | 97.4 | 69.0 - 99.8 | 99.5 | 99.3 - 99.7 |
| | | | | |
| **Antigen** | **Sensitivity** | **95% CI** | **Specificity** | **95% CI** |
| **Total** | 8.9 | 1.0 - 48.2 | 99.5 | 98.3 - 99.9 |

* These parameters could not be estimated due to limited data

When the data were pooled by study design, the overall sensitivity for case-control study designs was estimated as 84.2%, 95% CI [66.4 – 93.4], and the overall specificity for case-control designs was estimated as 99.1, 95% CI [97.2 – 99.7]. Overall sensitivity and specificity could not be pooled for cross-sectional study designs due to limited data.

The estimates from Model 2 were comparable to the estimates from Model 1, and credible intervals in Model 2 were comparable to the confidence intervals in Model 1. This demonstrated that our results were robust across the different assumptions that were made between models. Refer to the Appendix for examples of statistical output from Stata and WinBUGS.

# CHAPTER 5: DISCUSSION AND CONCLUSIONS

## 5.1 Summary of findings and comparison of models

This review identified 17 studies evaluating the diagnostic performance and accuracy of the Determine Combo rapid test, across various global settings. We synthesized evidence using two models with different assumptions, to check the robustness of results across different sets of assumptions; however the results were found to be largely consistent between models. Model 1 assumed that a perfect gold standard was available in each study, so that sensitivity and specificity were not correlated within each study. In other words, there was no trade off between sensitivity and specificity. This was an advantageous model to use because many of the studies failed to report both sensitivity and specificity estimates; in this model, studies were included even if the full data were not available. We wanted to know if the estimates obtained through this model were consistent under a different set of assumptions, so the data were analyzed using a second model, assuming there was an underlying threshold for positivity in each study. This means that the reference standards used between studies had different cut-offs for positivity. Different tests for HIV have different limits of detection, either for viral RNA, p24 antigen, or HIV antibody, which can be considered varying thresholds.

The results for overall estimates of the Determine Combo test were very similar between models, with Model 1 estimating sensitivity as 88.5%, 95% CrI [80.1 – 93.4], and specificity as 99.1%, 95% CrI [97.3 – 99.8], and Model 2 sensitivity as 88.6%, 95% CI [79.1 - 94.2], and specificity as 99.2, 95% CI [98.3 – 99.6]. The pooled antigen sensitivity estimates differed slightly between models, however the confidence interval and credible interval were comparable. Model 1 estimated pooled antigen sensitivity as 12.3%, 95% CrI [1.1 – 44.2], with specificity 99.7%, 95% CrI [96.8 – 100]. Model 2 estimated antigen sensitivity as 8.9%, 95% CI [1.0 – 48.2], and

specificity estimate 99.5, 95% CI [98.3 – 99.9]. The antibody estimates were comparable between models, with Model 1 estimating antibody sensitivity as 97.3%, 95% CrI [60.7 – 99.9], and specificity 99.6%, 95% CrI [99.0 -99.8]. Model 2 estimated antibody sensitivity as 97.4%, 95% CI [69.0 – 99.8], and specificity 99.5, 95% CI [99.3 – 99.7].

The results between these two models were comparable in terms of both point estimates and confidence/credible intervals, which gave us confidence in our estimates. The main difference between model estimates was that of the antigen component, which had different sensitivity estimates between models, but the confidence interval and credible interval were quite similar. Fewer studies were included in Model 2 compared to Model 1 since information for both sensitivity and specificity was required in Model 2, so it would be expected that the antigen component would see the largest change, because of the low number of antigen positive samples. For this reason, the discussion and conclusions will focus on the results estimated in Model 1.

The manufacturer of this assay claims an overall sensitivity of 100%, and specificities of 99.23% and 99.66% for the antibody and antigen components, respectively. [30] In this meta-analysis, we found the pooled specificity estimates (antibody specificity 99.6%, antigen specificity 99.5%) to be consistent with the performance claims made by the manufacturer; however the overall pooled sensitivity result (overall sensitivity 88.5%, 95% CrI [80.1 - 93.4]) was not consistent with claims of 100% sensitivity.

In this meta-analysis, the antibody component had high pooled sensitivity (97.3%, 95% CrI 60.7 - 99.9), however the antigen component was found to have very low sensitivity (12.3%, 95% CrI 1.1 – 44.2) for detecting HIV antibodies and p24 antigen, respectively. There are many reasons as to why the antigen component of the Determine Combo test performs poorly. One reason may

be that the Determine Combo test lacks heat dissociation and signal amplification steps, which are not feasible for rapid testing, but have been shown to improve the detection of p24 antigen by breaking apart immune complexes formed through nonspecific binding of p24 antigen to antibodies.[42,51,109] Others have suggested that the limit of detection for p24 antigen in this test is notably lower than laboratory p24 assays.[109,111] The p24 antigen sensitivity may also have been underestimated in this meta-analysis if p24 antigen was evaluated in samples that were defined as "acute" based on being antibody negative and RNA positive. There remains a period of time where RNA is present in the blood but p24 antigen is not yet present; so unless a reference standard was used which was able to detect p24 antigen, the accuracy of this component may be underestimated. The performance of the antigen component would also be underestimated if it was evaluated in samples which had already seroconverted (antibody positive but antigen negative), however we did not include these types of estimates in the meta-analysis. The discrepancies between the two components of the test explain why the Determine Combo test performs similarly to third generation tests (which are only able to detect HIV antibodies), but cannot perform at the same level as fourth generation lab tests which are much more sensitive for detecting acute infection.[114]

With respect to our findings, there were no substantive differences in sensitivity or specificity when the data were pooled by subgroups, for both blood sample type and study design. Studies using whole blood had a higher pooled sensitivity but lower specificity than serum/plasma, and cross-sectional studies had a higher pooled sensitivity and lower specificity than case-control studies, however these estimates were imprecise due to small sample sizes (all credible intervals greatly overlap). While no broad conclusions should be drawn, the subgroup results were different to what we expected, as case control studies tend to overestimate test accuracy,

however the case control studies included in this meta-analysis in general had higher numbers of acute samples, so the patient case mix across subgroups may not have been comparable.

Although many of the included studies concluded that the performance of the Determine rapid test was poor and failed to detect HIV infection, this meta-analysis provides evidence that the Determine Combo test is suitable for use in testing populations that have seroconverted, however in its current form the Determine Combo test cannot reliably detect p24 antigen. It should be noted that the accuracy of this test independent of the antibody component is only relevant for the period preceding seroconversion (the acute phase), when an antibody test would be negative. So even if the antigen component of this test performs poorly, if it is able to detect any antigen at all it might be an improvement on third generation tests. However in settings with access to advanced laboratory equipment, a fourth generation laboratory test may be the most useful for detecting acute infections.

## 5.2 Risk of bias within studies

Quality assessment of studies in this meta-analysis also points to the need to design studies of diagnostic accuracy with careful attention to methodology. Therefore, the overall pooled estimates should be interpreted with caution, as most of the studies included in this meta-analysis were of low methodological quality, particularly in the "index test" and "patient selection" domains, as was seen in Figure 5. Many studies were assigned a high risk of bias due to lack of blinding of the reference test results when interpreting the index test, which is a critical step in preventing detection bias. Studies which selected patients or samples into the study based on their HIV status (i.e. case-control designs) were also assigned a high risk of bias, as this type of patient selection tends to overestimate test accuracy by selecting patients with more obvious or extreme symptoms, rather than a representative sample of the disease spectrum (spectrum bias).

Studies were also at risk of incorporation bias and verification bias by not using the same reference standard for all samples, and in some cases including the index test in the reference standard testing algorithm. Additionally, many studies were at risk of imperfect reference standard bias by using inappropriate reference standards.

Overall, there was a lack of proper reporting of methods used to avoid bias, particularly in terms of the procedure for index and reference standard testing (and whether they were interpreted independently), and also the time interval between index test and reference test. The time interval is not a concern for cross-sectional studies as index tests and reference standard should be performed in parallel; however in case-control studies recruiting patients rather than using stored samples, there is a concern that the condition of the patient may change during this interval.[112]

Although many of the included studies were of low quality, this is not surprising or uncommon for studies of diagnostic accuracy, and has been recognized as a problem. [115] We chose to pool data for exploratory purposes and to analyze various subgroups, and all pooled estimates must be interpreted in the context study quality.

## 5.3 Interpretation of results

The overall results of this meta-analysis should be interpreted with a critical eye, as there are many factors that likely influenced the results. The prevalence of acute HIV infection in the study sample, which affects the case mix and patient spectrum in each study, plays a huge role in predicting the overall accuracy of this rapid test. While not analyzed statistically, descriptively it can be seen that the overall sensitivity reported by each individual study varies based on the patient spectrum used in the study sample. For example, in order to increase the sample size for antigen positive subjects, some studies restricted their analyses to samples of acute or recent

infections only (antigen positive, or RNA positive but Western blot negative). Since antibodies are not yet present in the blood during early stages of infection, the overall performance of the test will be disproportionally weighted to the antigen component of the test, which has a very different performance than the antibody component. The studies conducted by Pilcher et al,[108] Patel et al,[66] Faraoni et al,[101] and Fox et al,[102] each restricted their samples to antigen positive samples only, and reported overall sensitivities of 54.4%, 75.8%, 88.2%, and 72.2%, respectively (Figure 4). The study by Pavie et al restricted their sample to those with established infections only (overall sensitivity 95.8%),[107] and the study by Chetty et al restricted their sample to sero-converters (overall sensitivity 94.1%),[99] both more similar to the pooled antibody sensitivity of 97.3% (95% CrI 60.7 – 99.9). This trend seems to suggest that the overall sensitivity of the Determine Combo test might depend on the prevalence of acute specimens in the study sample. If the study sample consists of both recent and established infections (mirroring the real-life spectrum of patients that show up for testing), the overall sensitivity of the test will likely be higher than a study sample restricted to acute patients.

Furthermore, fourth generation tests are designed to pick up HIV within a very short window of detection. This is very challenging for a real life evaluation, as the likelihood of high antigenemia or high proportion of acutely infected samples is low even in a high incident or prevalent HIV setting. This was noted also by a CDC field trial.[111] Additionally, the role of HIV-1 subtypes, particularly non-B subtypes (which are more common outside of Europe and North America) and HIV-2 in predicting the performance of this test remains unknown. Some studies included a variety of subtypes (Subtype C, D, E) and suggested that this may have impaired performance of the p24 antigen component,[96,109] however the effect of subtype differences was not a main research question of any of the studies, highlighting a need for future research in this

area. Lastly, the optimal concentration of p24 to predict good sensitivity or specificity may only be achievable in the laboratory, which is important as this concentration may vary in field settings where individual patient response to HIV varies. This has also been called optimism bias (performance bias) and is not unique to HIV diagnostics.[5]

The updated CDC recommendations for HIV testing state that the initial test should be a fourth generation laboratory test, and if positive, followed by confirmatory antibody testing, and NAAT if these results are discrepant.[33] Using this algorithm to define the gold standard for HIV diagnosis, then only one of the 16 studies included in this review had a perfect gold standard (Masciotra et al),[106] with many studies initially testing using either 3rd generation assays or rapid tests followed by NAAT for confirmation, with little to no consistency between studies. Using a third generation test or another rapid test as a reference standard is not appropriate for evaluating a fourth generation test, as these tests were not designed to detect p24 antigen. Comparison with an inferior assay may underestimate performance parameters (i.e., sensitivity and specificity). However, as was seen in the forest plots in Figure 6, there did not seem to be an obvious difference between study accuracy estimates when fourth generation reference standards were compared with third generation reference standards; however, a statistical analysis would be needed to further explore this issue.

Since RNA is the earliest detectable marker in the blood, it may be considered the gold standard for HIV testing; however there remains a period of about five days between RNA detection and p24 antigen detection in the blood.[116] If RNA is used as a reference standard for p24 antigen sensitivity, the estimate may be biased due to RNA positive samples that are not yet antigen positive. Future studies evaluating the separate components of the Determine Combo test should use a reference standard that is able to accurately identify p24 antigen, such as a fourth

generation immunoassay or p24 assay. In this systematic review, one study described the antigen sensitivity as 0%, when in fact they did not use a reference standard for p24 antigen, which is quite misleading as there may have been no antigen positives in their study sample, which the authors noted.[97]

## 5.4 Strengths and limitations

Strengths of this systematic review and meta-analysis include the broad and extensive literature search and independent data extraction and study quality assessment. Data were pooled from all the studies using two different hierarchical random effects models to compare robustness across different assumptions. Additionally, sub-group analysis was undertaken by study design and blood samples to investigate possible heterogeneity.

Limitations: We were unable to formally assess publication bias, as using statistical tests and funnel plots to assess publication bias is not recommended for meta-analyses of diagnostic accuracy.[117] This is because many studies of diagnostic accuracy do not generate hypotheses, rendering tests of statistical significance irrelevant; additionally, sample sizes are not necessarily indicative of study quality, particularly when considering study design, as we found in many of the included case-control studies. For these reasons, we must assume that publication bias was present in this systematic review and meta-analysis. Conference abstracts were searched in an effort to minimize publication bias, however after careful review of all abstracts, and recognizing that results from conference abstracts are often preliminary, no abstracts met the inclusion criteria for this systematic review.

Additionally, we were unable to statistically explore the effect of patient case mix on the overall pooled accuracy estimates due to limited data and varying definitions of acute infection between

studies. We had originally planned to explore the effect of imperfect and perfect references standards, however data limitations did not allow for this type of statistical analysis.

## 5.5 Future directions for research

In general, future studies evaluating the diagnostic accuracy of any rapid test, including the Determine Combo test, should be mindful of biases that we observed in the studies included in this meta-analysis. Studies should take note of the spectrum of patients they are including in their sample, to be representative of the spectrum of patients that show up for testing. In light of the results of this systematic review and meta-analysis, more studies on this test are not needed; however, if more studies are to be undertaken which evaluate the Determine HIV combo test, they should be large cross-sectional field studies using appropriate reference standards, and evaluate the overall accuracy as well as each component separately. Case control studies are not needed at this stage. If many more studies are to be published, then perhaps the effect of imperfect and perfect reference standards can be explored using Bayesian methods.[93]

However, Alere (the company who manufactures the Determine Combo test) has recently released an improved version of this test, called the Alere HIV Combo, which has yet to have its diagnostic performance independently evaluated in field settings.[118] If the antigen performance of this test has been improved (without compromising specificity of the test) this may become the first rapid test to accurately detect acute HIV infection, however methodologically rigorous evaluations must be conducted to produce evidence on the performance of this test in real life settings. At this point in time, studies evaluating this new test are needed more than new studies evaluating the Determine Combo test.

## 5.6 Conclusion

This is the first meta-analysis to synthesize evidence on the overall global performance of the Determine HIV 1/2 Ag/Ab Combo rapid test. We found that the Determine Combo test is accurate in identifying HIV antibodies, but based on the studies which evaluated p24 antigen independently of antibodies, the antigen component should be improved to rival fourth generation laboratory assays. Based on the evidence, this assay in its current form may be suitable for identifying HIV infections in those who have already seroconverted, but falls short in picking up acute HIV infections; its performance is better compared to many third generation antibody based assays. To pick up acute HIV infection in resource-rich settings, fourth generation laboratory assays and NAAT may be a better option, but in resource-limited settings the Determine Combo assay may prove to be a useful alternative, or a step up to third generation rapid testing. The assay may be used as another POC test for settings which have a proportional mix of acute, early and seroconverting patients. It may also offer strong competition to the available repertoire of antibody based tests; however it needs to improve performance of its antigen component to detect acute infection with sufficient accuracy.

## REFERENCES


**1.** UNAIDS. *Global Report: UNAIDS report on the global AIDS epidemic 2013.* Geneva 2013.

**2.** WHO. Global summary of the AIDS epidemic 2013. [Table]. 2014; http://www.who.int/hiv/data/epi_core_dec2014.png?ua=1. Accessed August 9, 2014.

**3.** UNAIDS. *The Gap Report.* Geneva, Switzerland2014.

**4.** United Nations. *The Millennium Development Goals Report*: United Nations;2014.

**5.** United Nations. *We can end poverty: Millenium Development Goals and Beyond 2015 Fact Sheet*: United Nations;2013.

**6.** UNAIDS. *90-90-90: An ambitious treatment target to help end the AIDS epidemic*2014.

**7.** Public Health Agency of Canada. Summary: Estimates of HIV Prevalence and Incidence in Canada, 2011. 2012; http://www.phac-aspc.gc.ca/aids-sida/publication/survreport/estimat2011-eng.php. Accessed May 30, 2014.

**8.** Brenner Bluma G, Roger M, Routy JP, et al. High Rates of Forward Transmission Events after Acute/Early HIV-1 Infection. *J Infect Dis.* 2007;195(7):951-959.

**9.** Hollingsworth TD, Anderson Roy M, Fraser C. HIV-1 Transmission, by Stage of Infection. *J Infect Dis.* 2008;198(5):687-693.

**10.** Prabhu VS, Hutchinson AB, Farnham PG, Sansom SL. Sexually acquired HIV infections in the United States due to acute-phase HIV transmission: an update. *AIDS.* Aug 24 2009;23(13):1792-1794.

**11.** Wawer MJ, Gray RH, Sewankambo NK, et al. Rates of HIV-1 transmission per coital act, by stage of HIV-1 infection, in Rakai, Uganda. *J Infect Dis.* May 1 2005;191(9):1403-1409.

**12.** Branson BM. The Future of HIV Testing. *J Acquir Immune Defic Syndr.* 2010;55:S102-S105.

**13.** Dombrowski JC, Harrington RD, Golden MR. Evidence for the Long-Term Stability of HIV Transmission–Associated Sexual Behavior After HIV Diagnosis. *Sex Transm Dis.* 2013;40(1):41-45.

**14.** Marks G, Crepaz N, Janssen RS. Estimating sexual transmission of HIV from persons aware and unaware that they are infected with the virus in the USA. *AIDS.* Jun 26 2006;20(10):1447-1450.

**15.** Bell Sigall K, Little Susan J, Rosenberg Eric S. Clinical Management of Acute HIV Infection: Best Practice Remains Unknown. *J Infect Dis.* 2010;202(S2):S278-S288.

**16.** Wood E, Kerr T, Rowell G, et al. Does This Adult Patient Have Early HIV Infection? *JAMA.* 2014;312(3):278.

**17.** O'Brien M, Markowitz M. Should We Treat Acute HIV Infection? *Curr HIV/AIDS Rep.* 2012;9(2):101-110.

**18.** Thomas G. WHO issues new HIV recommendations calling for earlier treatment. 2013; http://www.who.int/mediacentre/news/releases/2013/new_hiv_recommendations_20130630/en/. Accessed May 30 2014.

**19.** Piwowar-Manning EM, Tustin NB, Sikateyo P, et al. Validation of Rapid HIV Antibody Tests in 5 African Countries. *J Int Assoc Provid AIDS Care.* 2010;9(3):170-172.

**20.** Yang M, Prestage G, Maycock B, et al. The acceptability of different HIV testing approaches: cross-sectional study among GMSM in Australia. *Sex Transm Infect.* 2014;0:1-4.

**21.** Burns F, Edwards SG, Woods J, et al. Acceptability, feasibility and costs of universal offer of rapid point of care testing for HIV in an acute admissions unit: results of the RAPID project. *HIV Med.* Oct 2013;14 Suppl 3:10-14.

**22.** Lewis NM, Gahagan JC, Stein C. Preferences for rapid point-of-care HIV testing in Nova Scotia, Canada. *Sex Health.* May 2013;10(2):124-132.

**23.** Appiah LT, Havers F, Gibson J, Kay M, Sarfo F, Chadwick D. Efficacy and acceptability of rapid, point-of-care HIV testing in two clinical settings in Ghana. *AIDS Patient Care STDS.* May 2009;23(5):365-369.

**24.** Drain PK, Hyle EP, Noubary F, et al. Diagnostic point-of-care tests in resource-limited settings. *Lancet Infect Dis.* 2014;14(3):239-249.

**25.** Branson BM. Point-of-Care Rapid Tests for HIV Antibodies. *Laboratoriums Medizin.* 2003;27(7-8):288-295.

**26.** Delaney KP, Branson BM, Uniyal A, et al. Evaluation of the performance characteristics of 6 rapid HIV antibody tests. *Clin Infect Dis.* Jan 15 2011;52(2):257-263.

**27.** Stekler Joanne D, Swenson Paul D, Coombs Robert W, et al. HIV Testing in a High-Incidence Population: Is Antibody Testing Alone Good Enough? *Clin Infect Dis.* 2009;49(3):444-453.

**28.** Yerly S, Hirschel B. Diagnosing acute HIV infection. *Expert Rev Anti Infect Ther.* 2012;10(1):31-41.

**29.** Branson BM, Stekler JD. Detection of Acute HIV Infection: We Can't Close the Window. *J Infect Dis.* 2011;205(4):521-524.

**30.** Alere International Limited. The Alere Determine™ HIV-1/ 2 Ag /Ab Combo Brochure. 2012; http://alerehiv.s3.amazonaws.com/wp-content/uploads/12000140-v03-Alere-Determine-HIV-12-AgAb-Combo-Broc.pdf. Accessed September 2 2014.

**31.** Epstein JS. August 9, 2013 Approval Letter - Alere Determine HIV-1/2 Ag/Ab Combo. 2013; http://www.fda.gov/BiologicsBloodVaccines/BloodBloodProducts/ApprovedProducts/PremarketApprovalsPMAs/ucm364653.htm. Accessed September 2, 2014.

**32.** Alere International Limited. Alere HIV Studies and Implementation World Map2015.

**33.** Branson BM, Owen SM, Wesolowski LG, et al. Laboratory Testing for the Diagnosis of HIV Infection: Updated Recommendations. 2014; http://www.cdc.gov/hiv/pdf/HIVtestingAlgorithmRecommendation-Final.pdf. Accessed September 2 2014.

**34.** McMichael AJ, Borrow P, Tomaras GD, Goonetilleke N, Haynes BF. The immune response during acute HIV-1 infection: clues for vaccine development. *Nat Rev Immunol.* Jan 2010;10(1):11-23.

**35.** Chinen J, Shearer WT. Molecular virology and immunology of HIV infection. *J Allergy Clin Immunol.* 8// 2002;110(2):189-198.

**36.** Cohen MS, Shaw GM, McMichael AJ, Haynes BF. Acute HIV-1 Infection. *N Engl J Med.* May 19 2011;364(20):1943-1954.

**37.** Kahn JO, Walker BD. Acute human immunodeficiency virus type 1 infection. *N Engl J Med.* Jul 2 1998;339(1):33-39.

**38.** Centers for Disease Control and Prevention. About HIV/AIDS. 2014; http://www.cdc.gov/actagainstaids/basics/whatishiv.html. Accessed March 12, 2015.

**39.** Sanders EJ, Mugo P, Prins HA, et al. Acute HIV-1 infection is as common as malaria in young febrile adults seeking care in coastal Kenya. *AIDS.* Jun 1 2014;28(9):1357-1363.

**40.** World Health Organization. *Interim WHO clinical staging of HIV/AIDS and HIV/AIDS case definitions for surveillance.* Geneva2005.

**41.** Fiebig EW, Wright DJ, Rawald BD, et al. Dynamics of HIV viremia and antibody seroconversion in plasma donors: implications for diagnosis and staging of primary HIV infection. *AIDS.* Sep 5 2003;17(13):1871-1879.

**42.** Nishanian P, Huskins KR, Stehn S, Detels R, Fahey JL. A simple method for improved assay demonstrates that HIV p24 antigen is present as immune complexes in most sera from HIV-infected individuals. *J Infect Dis.* Jul 1990;162(1):21-28.

**43.** Schupbach J, Flepp M, Pontelli D, Tomasik Z, Luthy R, Boni J. Heat-mediated immune complex dissociation and enzyme-linked immunosorbent assay signal amplification render p24 antigen detection in plasma as sensitive as HIV-1 RNA detection by polymerase chain reaction. *AIDS.* Sep 1996;10(10):1085-1090.

**44.** Daar ES, Pilcher CD, Hecht FM. Clinical presentation and diagnosis of primary HIV-1 infection. *Curr Opin HIV AIDS.* Jan 2008;3(1):10-15.

**45.** Ananworanich J, Fletcher JL, Pinyakorn S, et al. A novel acute HIV infection staging system based on 4th generation immunoassay. *Retrovirology.* 2013;10:56.

**46.** Gurtler L, Muhlbacher A, Michl U, et al. Reduction of the diagnostic window with a new combined p24 antigen and human immunodeficiency virus antibody screening assay. *J Virol Methods.* Nov 1998;75(1):27-38.

**47.** Owen SM. Testing for acute HIV infection. *Curr Opin HIV AIDS.* 2012;7(2):125-130.

**48.** Murphy G, Aitken C. HIV testing—The perspective from across the pond. *J. Clin. Virol.* 2011;52:S71-S76.

**49.** Lee BE, Plitt SS, Jayaraman GC, Chui L, Singh AE, Preiksaitis JK. Use of Quantitative HIV RNA Detection for Early Diagnosis of HIV Infection in Infants and Acute HIV Infections in Alberta, Canada. *J Clin Microbiol.* 2011;50(2):502-505.

**50.** Steinberg M, Cook DA, Gilbert M, et al. Towards targeted screening for acute HIV infections in British Columbia. *J Int AIDS Soc.* 2011;14(1):39.

**51.** Schüpbach J, Tomasik Z, Knuchel M, et al. Optimized virus disruption improves detection of HIV-1 p24 in particles and uncovers a p24 reactivity in patients with undetectable HIV-1 RNA under long-term HAART. *J Med Vir.* 2006;78(8):1003-1010.

**52.** Busch MP, Hecht F. Nucleic acid amplification testing for diagnosis of acute HIV infection: has the time come? *AIDS.* Aug 12 2005;19(12):1317-1319.

**53.** Pai NP, Vadnais C, Denkinger C, Engel N, Pai M. Point-of-care testing for infectious diseases: diversity, complexity, and barriers in low- and middle-income countries. *PLoS Med.* 2012;9(9):e1001306.

**54.** Kettler H, White K, Hawkes S. *Mapping the landscape of diagnostics for sexually transmitted infections.* Geneva: World Health Organization;2004.

**55.** Peeling RW, Mabey D. Point-of-care tests for diagnosing infections in the developing world. *Clin Microbiol Infect.* Aug 2010;16(8):1062-1069.

**56.** Morin SF, Khumalo-Sakutukwa G, Charlebois ED, et al. Removing barriers to knowing HIV status: same-day mobile HIV testing in Zimbabwe. *J Acquir Immune Defic Syndr.* Feb 1 2006;41(2):218-224.

**57.** Kassler WJ, Alwano-Edyegu MG, Marum E, Biryahwaho B, Kataaha P, Dillon B. Rapid HIV testing with same-day results: a field trial in Uganda. *Int J STD AIDS.* Mar 1998;9(3):134-138.

**58.** Vickerman P, Watts C, Peeling RW, Mabey D, Alary M. Modelling the cost effectiveness of rapid point of care diagnostic tests for the control of HIV and other sexually transmitted infections among female sex workers. *Sex Transm Infect.* Oct 2006;82(5):403-412.

**59.** Walensky RP, Wood R, Fofana MO, et al. The clinical impact and cost-effectiveness of routine, voluntary HIV screening in South Africa. *J Acquir Immune Defic Syndr.* Jan 1 2011;56(1):26-35.

**60.** Becker ML, Thompson LH, Pindera C, et al. Feasibility and success of HIV point-of-care testing in an emergency department in an urban Canadian setting. *Can J Infect Dis Med Microbiol.* Spring 2013;24(1):27-31.

**61.** Koester KA, Collins SP, Fuller SM, Galindo GR, Gibson S, Steward WT. Sexual Healthcare Preferences among Gay and Bisexual Men: A Qualitative Study in San Francisco, California. *PLoS ONE.* 2013;8(8):e71546.

**62.** Spielberg F, Branson BM, Goldbaum GM, et al. Choosing HIV Counseling and Testing Strategies for Outreach Settings: A Randomized Trial. *J Acquir Immune Defic Syndr.* Mar 1 2005;38(3):348-355.

**63.** Kendrick SR, Kroc KA, Couture E, Weinstein RA. Comparison of point-of-care rapid HIV testing in three clinical venues. *AIDS.* Nov 5 2004;18(16):2208-2210.

**64.** Louie B, Wong E, Klausner JD, et al. Assessment of Rapid Tests for Detection of Human Immunodeficiency Virus-Specific Antibodies in Recently Infected Individuals. *J Clin Microbiol.* 2008;46(4):1494-1497.

**65.** Pilcher Christopher D, Christopoulos Katerina A, Golden M. Public Health Rationale for Rapid Nucleic Acid or p24 Antigen Tests for HIV. *J Infect Dis.* 2010;201(s1):S7-S15.

**66.** Patel P, Bennett B, Sullivan T, Parker MM, Heffelfinger JD, Sullivan PS. Rapid HIV screening: Missed opportunities for HIV diagnosis and prevention. *J. Clin. Virol.* 2012;54(1):42-47.

**67.** Patel P, Klausner JD, Bacon OM, et al. Detection of acute HIV infections in high-risk patients in California. *J Acquir Immune Defic Syndr.* May 2006;42(1):75-79.

**68.** Bassett IV, Chetty S, Giddy J, et al. Screening for acute HIV infection in South Africa: finding acute and chronic disease*. *HIV Med.* 2011;12(1):46-53.

**69.** Morrison CS, Demers K, Kwok C, et al. Plasma and cervical viral loads among Ugandan and Zimbabwean women during acute and early HIV-1 infection. *AIDS.* 2010;24(4):573-582.

**70.** Pilcher CD, Tien HC, Eron JJ, Jr., et al. Brief but efficient: acute HIV infection and the sexual transmission of HIV. *J Infect Dis.* May 15 2004;189(10):1785-1792.

**71.** Remien RH, Higgins JA, Correale J, et al. Lack of Understanding of Acute HIV Infection among Newly-Infected Persons—Implications for Prevention and Public Health: The NIMH Multisite Acute HIV Infection Study: II. *AIDS Behav.* 2009;13(6):1046-1053.

**72.** Prins HAB, Mugo P, Wahome E, et al. Diagnosing acute and prevalent HIV-1 infection in young African adults seeking care for fever: a systematic review and audit of current practice. *Int Health.* 2014;6(2):82-92.

**73.** Mlisana K, Sobieszczyk M, Werner L, et al. Challenges of Diagnosing Acute HIV-1 Subtype C Infection in African Women: Performance of a Clinical Algorithm and the Need for Point-of-Care Nucleic-Acid Based Testing. *PLoS ONE.* 2013;8(4):e62928.

**74.** Powers KA, Miller WC, Pilcher CD, et al. Improved detection of acute HIV-1 infection in sub-Saharan Africa: development of a risk score algorithm. *AIDS.* Oct 18 2007;21(16):2237-2242.

**75.** Marks G, Crepaz N, Senterfitt JW, Janssen RS. Meta-Analysis of High-Risk Sexual Behavior in Persons Aware and Unaware They are Infected With HIV in the United States: Implications for HIV Prevention Programs. *J Acquir Immune Defic Syndr.* 2005;39 (4):446-453.

**76.** Weinhardt LS, Carey MP, Johnson BT, Bickham NL. Effects of HIV counseling and testing on sexual risk behavior: a meta-analytic review of published research, 1985-1997. *Am J Public Health.* Sep 1999;89(9):1397-1405.

**77.** Steward WT, Remien RH, Higgins JA, et al. Behavior change following diagnosis with acute/early HIV infection-a move to serosorting with other HIV-infected individuals. The NIMH Multisite Acute HIV Infection Study: III. *AIDS Behav.* Dec 2009;13(6):1054-1060.

**78.** Cope AB, Crooks AM, Chin T, et al. Incident Sexually Transmitted Infection as a Biomarker for High-Risk Sexual Behavior After Diagnosis of Acute HIV. *Sex Transm Dis.* 2014;41(7):447-452.

**79.** Corneli A, Pettifor A, Kamanga G, et al. HPTN 062: A Feasibility and Acceptability Pilot Intervention to Reduce HIV Transmission Risk Behaviors Among Individuals with Acute and Early HIV Infection in Lilongwe, Malawi. *AIDS Behav.* 2014.

**80.** Pilcher CD, Shugars DC, Fiscus SA, et al. HIV in body fluids during primary HIV infection: implications for pathogenesis, treatment and public health. *AIDS.* May 4 2001;15(7):837-845.

**81.** Hoen B, Cooper DA, Lampe FC, et al. Predictors of Virological Outcome and Safety in Primary HIV Type 1–Infected Patients Initiating Quadruple Antiretroviral Therapy: QUEST GW PROB3005. *Clin Infect Dis.* August 1, 2007 2007;45(3):381-390.

**82.** Streeck H, Jessen H, Alter G, et al. Immunological and Virological Impact of Highly Active Antiretroviral Therapy Initiated during Acute HIV-1 Infection. *J Infect Dis.* September 15, 2006 2006;194(6):734-739.

**83.** Hecht FM, Wang L, Collier A, et al. A multicenter observational study of the potential benefits of initiating combination antiretroviral therapy during acute HIV infection. *J Infect Dis.* Sep 15 2006;194(6):725-733.

**84.** Hamlyn E, Jones V, Porter K, Fidler S. Antiretroviral treatment of primary HIV infection to reduce onward transmission. *Curr Opin HIV AIDS.* 2010;5(4):283-290.

**85.** Niu MT, Bethel J, Holodniy M, Standiford HC, Schnittman SM. Zidovudine treatment in patients with primary (acute) human immunodeficiency virus type 1 infection: a randomized, double-blind, placebo-controlled trial. DATRI 002 Study Group. Division of AIDS Treatment Research Initiative. *J Infect Dis.* Jul 1998;178(1):80-91.

**86.** Kinloch-De Loes S, Hirschel BJ, Hoen B, et al. A controlled trial of zidovudine in primary human immunodeficiency virus infection. *N Engl J Med.* Aug 17 1995;333(7):408-413.

**87.** Fidler S, Porter K, Ewings F, et al. Short-course antiretroviral therapy in primary HIV infection. *N Engl J Med.* Jan 17 2013;368(3):207-217.

**88.** Sigaloff KCE, Lange JMA, Montaner J. Global Response to HIV: Treatment as Prevention, or Treatment for Treatment? *Clin Infect Dis.* 2014;59(suppl 1):S7-S11.

**89.** Alere International Limited. Alere Determine™ HIV-1/2 Ag/Ab Combo. 2014; http://alerehiv.com/hiv-screening/features. Accessed August 10 2014.

**90.** Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* Jul 21 2009;6(7):e1000097.

**91.** Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med.* Oct 18 2011;155(8):529-536.

**92.** Whiting P, Harbord R, Kleijnen J. No role for quality scores in systematic reviews of diagnostic accuracy studies. *BMC Med Res Methodol.* 2005;5.

**93.** Dendukuri N, Schiller I, Joseph L, Pai M. Bayesian meta-analysis of the accuracy of a test for tuberculous pleuritis in the absence of a gold standard reference. *Biometrics.* Dec 2012;68(4):1285-1293.

**94.** UNAIDS/WHO. *Guidelines for using HIV testing technologies in surveillance: selection, evaluation and implementation – 2009 update.* Geneva2009.

**95.** Harbord RM, Whiting P. Metandi: meta-analysis of diagnostic accuracy using hierarchical logistic regression. *The Stata Journal.* 2009;9(2):18.

**96.** Beelaert G, Fransen K. Evaluation of a rapid and simple fourth-generation HIV screening assay for qualitative detection of HIV p24 antigen and/or antibodies to HIV-1 and HIV-2. *J Virol Methods.* 2010;168(1-2):218-222.

**97.** Bhowan K, Kalk E, Khan S, Sherman G. Identifying HIV infection in South African women: How does a fourth generation HIV rapid test perform? *Afr J Lab Med.* 2011;1(1).

**98.** Brauer M, De Villiers JC, Mayaphi SH. Evaluation of the Determine™ fourth generation HIV rapid assay. *J Virol Methods.* 2013;189(1):180-183.

**99.** Chetty V, Moodley D, Chuturgoon A. Evaluation of a 4th generation rapid HIV test for earlier and reliable detection of HIV infection in pregnancy. *J. Clin. Virol.* 2012;54(2):180-184.

**100.** Conway DP, Holt M, McNulty A, et al. Multi-centre evaluation of the Determine HIV Combo assay when used for point of care testing in a high risk clinic-based population. *PLoS ONE* 2014;9(4):e94062.

**101.** Faraoni S, Rocchetti A, Gotta F, et al. Evaluation of a rapid antigen and antibody combination test in acute HIV infection. *J. Clin. Virol.* 2013;57(1):84-87.

**102.** Fox J, Dunn H, O'Shea S. Low rates of p24 antigen detection using a fourth-generation point of care HIV test. *Sex Trans Infect.* 2010;87(2):178-179.

**103.** Jones CB, Kuldanek K, Muir D, et al. Clinical evaluation of the Determine HIV-1/2 Ag/Ab Combo test. *J Infect Dis.* Dec 15 2012;206(12):1947-1949; author reply 1949-1950.

**104.** Kilembe W, Keeling M, Karita E, et al. Failure of a novel, rapid antigen and antibody combination test to detect antigen-positive HIV infection in African adults with early HIV infection. *PLoS ONE.* 2012;7(6):e37154.

**105.** Laperche S, Leballais L, Ly TD, Plantier JC. Failures in the detection of HIV p24 antigen with the Determine HIV-1/2 Ag/Ab Combo rapid test. *J Infect Dis.* Dec 15 2012;206(12):1946-1947; author reply 1949-1950.

**106.** Masciotra S, Luo W, Youngpairoj AS, et al. Performance of the Alere Determine HIV-1/2 Ag/Ab Combo Rapid Test with specimens from HIV-1 seroconverters from the US

and HIV-2 infected individuals from Ivory Coast. *J. Clin. Virol.* Dec 2013;58 Suppl 1:e54-58.

**107.** Pavie J, Rachline A, Loze B, et al. Sensitivity of five rapid HIV tests on oral fluid or finger-stick whole blood: a real-time comparison in a healthcare setting. *PLoS ONE* 2010;5(7):e11581.

**108.** Pilcher CD, Louie B, Facente S, et al. Performance of Rapid Point-of-Care and Laboratory Tests for Acute and Established HIV Infection in San Francisco. *PLoS ONE.* 2013;8(12):e80629.

**109.** Rosenberg NE, Kamanga G, Phiri S, et al. Detection of acute HIV infection: a field evaluation of the Determine HIV-1/2 Ag/Ab combo test. *J Infect Dis.* Feb 15 2012;205(4):528-534.

**110.** Stekler JD, O'Neal JD, Lane A, et al. Relative accuracy of serum, whole blood, and oral fluid HIV tests among Seattle men who have sex with men. *J. Clin. Virol.* 2013;58:e119-e122.

**111.** Duong YT, Mavengere Y, Patel H, et al. Poor performance of the determine HIV-1/2 Ag/Ab combo fourth-generation rapid test for detection of acute infections in a National Household Survey in Swaziland. *J Clin Microbiol.* Oct 2014;52(10):3743-3748.

**112.** Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. *Ann Intern Med.* Jan 7 2003;138(1):40-44.

**113.** Brauer M, De Villiers JC, Mayaphi SH. Evaluation of the Determine fourth generation HIV rapid assay. *J Virol Methods.* Apr 2013;189(1):180-183.

**114.** Brennan CA, Yamaguchi J, Vallari A, Swanson P, Hackett Jr JR. ARCHITECT® HIV Ag/Ab Combo assay: Correlation of HIV-1 p24 antigen sensitivity and RNA viral load using genetically diverse virus isolates. *J. Clin. Virol.* 2013;57(2):169-172.

**115.** Fontela PS, Pant Pai N, Schiller I, Dendukuri N, Ramsay A, Pai M. Quality and reporting of diagnostic accuracy studies in TB, HIV and malaria: evaluation using QUADAS and STARD standards. *PLoS ONE.* 2009;4(11):e7753.

**116.** Cohen Myron S, Gay Cynthia L, Busch Michael P, Hecht Frederick M. The Detection of Acute HIV Infection. *J Infect Dis.* 2010;202(S2):S270-S277.

**117.** Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Y T. Chapter 10: Analysing and Presenting Results. In: Deeks JJ, Bossuyt PM, C G, eds. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* The Cochrane Collaboration; 2010:46-47.

**118.** Alere International Limited. Alere HIV Combo (Package insert): Alere; 2015.

# APPENDICES

## Search Strategy

### Detailed count of records per database

| | Results | Duplicates deleted | Unique results to screen |
|---|---|---|---|
| **Medline articles** | 374 | 1 | 373 |
| **Embase articles** | 404 | 347 | 57 |
| **Embase conferences** | 65 | 1 | 64 |
| **Biosis articles** | 255 | 242 | 13 |
| **Biosis conferences** | 11 | 0 | 11 |
| **Cochrane** | 33 | 21 | 12 |
| **Pubmed** | 3 | 0 | 3 |
| **AIM** | 0 | 0 | 0 |
| **LILACS** | 22 | 0 | 22 |
| **TOTAL** | 1167 | 612 | 555 |

Total records identified through Medline and Embase search update on January 7th, 2015: **55**

### Detailed search strategy (Medline via Ovid)

Database: Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily, Ovid MEDLINE(R) and Ovid OLDMEDLINE(R) <1946 to Present>

Search Strategy:

---

1 ((Alere or inverness) and determine*).tw,kf,ot,oa. (34)
2 (((ag adj2 ab) or (antigen* adj2 antibod*)) adj4 determine*).tw,kf,ot,oa. (167)
3 (Determine* adj4 (HIV* or kit or combo)).tw,kf,ot,oa. (4224)
4 (Determine* adj4 ((4th or "4" or fourth) adj2 generation)).tw,kf,ot,oa. (12)
5 or/1-4 (4413)
6 exp HIV/ or exp HIV infections/ or HIV Antibodies/ or exp HIV Antigens/ (266037)
7 hiv*.mp,jw,in. (279924)
8 (acquired adj2 immun* adj2 syndrome*).mp,jw,in. (93499)
9 (acquired immun* adj3 deficien*).mp,jw,in. (13318)
10 (human immunedeficien* or human immunodeficien* or human immunideficien*).mp,jw,in. (79379)
11 or/6-10 (330779)
12 Acute disease/ (182858)
13 Early diagnosis/ (13258)
14 (acute* adj5 (infect* or disease)).tw,kf. (69445)
15 ((recent* or earl* or primary) adj2 (infect* or disease)).tw,kf. (62739)
16 (immediately adj2 (post or after or following) adj2 (infect* or acquisit* or transmit*)).tw,kf. (616)

17 (AHI or AHIs or PHI or PHIs).tw,kf. (21194)
18 (preseroconver* or pre-seroconver* or ((recent* or primary or new* or initial* or preced* or before) adj2 (seroconver* or sero-conver*)) or ((acquisition* or infection*) adj6 (seroconver* or sero-conver*))).tw,kf. (1896)
19 ((acute or initial or ramp-up) adj2 (phase* or period* or stage*)).tw,kf. (76918)
20 (newly adj2 infected).tw,kf. (764)
21 incident infection*.tw,kf. (257)
22 or/12-21 (395591)
23 11 and 22 (12936)
24 acute retroviral syndrome.tw,kf. (79)
25 incident hiv*.tw,kf. (245)
26 (Acute adj4 (hiv* or ((acquired or human) adj2 immun* adj3 (syndrome* or deficien*)) or human immunedeficien* or human immunodeficien* or human immunideficien*)).tw,kf. (2071)
27 ((recent* or earl* or primary) adj2 (hiv* or ((acquired or human) adj2 immun* adj3 (syndrome* or deficien*)) or human immunedeficien* or human immunodeficien* or human immunideficien*)).tw,kf. (5767)
28 24 or 25 or 26 or 27 (7709)
29 23 or 28 (16904)
30 5 and 29 (406)
31 limit 30 to humans (375)
32 limit 30 to animals (21)
33 30 not (31 or 32) (25)
34 31 or 33 (400)
35 remove duplicates from 34 (374)

## Appendix Table 1: Demographic characteristics of included studies

| Authors | Year | Country | Study Objective | Study Design | High or Low risk population | Reference Standard(s) used | Patient Population (general characteristics) | Specimen | Total Sample Size (n=) |
|---|---|---|---|---|---|---|---|---|---|
| Marieke Brauer, et al | 2013 | South Africa | To compare the performance of the Determine Combo assay with automated fourth generation HIV ELISAs already in use | Case-control | High risk | 4th generation ELISA followed by confirmatory testing with another fourth generation assay. Seroconversion panel: antibody assay followed by NAAT and p24 assay. | Serum specimens that had been submitted for routine HIV serology testing, plus seroconversion specimens from the South African National Blood service | serum | 79 |
| Damian P. Conway, et al | 2014 | Australia | To gain a better understanding of the potential of Determine for use as a point of care screening assay | Cross-sectional | High risk | 4th generation ELISA, followed by comfirmatory antibody testing, HIV p24 assay, and Western Blot | Free access publicly funded sexual health clinics with high caseloads of MSM | whole blood | 3190 |
| Christopher D. Pilcher, et al | 2013 | United States | To determine how use of newer tests might perform given their variable ability to detect acute infections, and how new tests might influence the performance of HIV testing programs in San Francisco. | Case-control | High risk | Initial testing with 1st generation, 3rd generation, or rapid antibody tests, followed by confirmatory antibody testing and RNA. | Stored remnants of blood plasma from STD clinic populations, MSM, Non-occupational post-exposure prophylaxis, partner services testing population, sex worker population, UCSF options study (suspicion of having acute infection) | plasma | 138 |
| Pragna Patel, et al | 2012 | United States | To evaluate the sensitivity for early HIV infection of several rapid tests and third and fourth generation assays compared with NAAT. | Case-control | Unclear | 1st, 2nd, or 3rd generation antibody test or rapid test, followed by NAAT. | Patients from STD clinics, public health clinics, and MSM. | plasma | 33 |
| | | | | | | | Blinded specimen panel including AHI, Ab-positive and HIV-negative specimens | plasma | 114 |
| Nora E. Rosenberg, et al | 2012 | Malawi | To assess the antigen portion of Determine Combo RT to detect persons with acute HIV infection. Also to assess the antibody portion of Combo RT against a standard rapid test antibody algorithm. | Cross-sectional | High risk | Rapid antibody test followed by NAAT. | Patients presenting for HIV testing in HIV testing and counselling centre, patients with negative or discordant HIV rapid test results (at high risk for acute) in STI clinic | whole | 1009 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Juliette Pavie, et al | 2010 | France | French agency for health product safety mandated that a real-time comparison of five approved rapid tests be carried out on samples from patients with documented HIV infection. Compared fingerstick with whole blood. | Case-control | High (HIV positive) | Western blot (previously confirmed) | Adults with documented HIV infection and HIV seronegative volunteers | whole | 220 |
| William Kilembe, et al | 2012 | Zambia and Rwanda | "We present results generated by testing specimen panels from adults with or at risk for acquiring HIV infection, and a panel of serially diluted p24 antigen positive controls" | Case-control | High | Rapid antibody test, followed by p24 antigen assay, unclear as to whether or not this was followed up with NAAT. | Specimen panels from adults with or at risk for acquiring HIV infection, and a panel of serially diluted p24 antigen postitive controls | plasma | 52 (acute and recent infections) |
| Silvia Faraoni, et al | 2013 | Italy | To establish Determine performances in acute HIV infection (AHI) in a setting at low HIV-1 prevalence | Case control | Low | Antibody test and Western blot, followed by a PCR-based assay. | Serum samples from patients with AHI, controls from those (HIV negative) seeking HIV testing | serum | 141 |
| Vani Chetty, et al | 2012 | South Africa | To evaluate Determine against two third generation tests for early detection of HIV infection in pregnancy | Cross-sectional | High | Rapid antibody test | Pregnant women participating in a HIV incidence cohort study who tested negative at baseline | plasma | 253 |
| Julie Fox, et al | 2011 | United Kingdom | To investigate the ability of the Determine assay to detect p24 antigen in samples identified as p24-positive using standard of care fourth generation assays. | Case-control | Not reported | 4th generation assay | Stored serum samples following standard laboratory testing | serum | 36 |
| Silvina Masciotra, et al | 2013 | United States | To evaluate the ability of Determine Combo RT to detect acute/early HIV-1 infections and HIV-2 antibody in well-characterized plasma specimens | Case-control | unclear | 4th generation assay | HIV-1 seroconverters previously used to evaluate FDA-approved assays | plasma | 230 |
| Joanne D. Stekler, et al | 2013 | United States | To compare the abilities of POC tests to detect early infection in real time | Cross-sectional | High | 3rd and 4th generation assays, followed by RNA testing | Men and transgender persons reporting sex with men, considered to be at high risk, plus those with suspected/confirmed acute diagnosis | whole | Unclear - 25 HIV positive specimens were tested |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| G. Beelaert, and K. Fransen | 2010 | Belgium | To evaluate the Determine fourth generation Combo assay using well-characterized HIV panels and a dilution series of HIV culture supernatants. | Case-control | Not reported | Negative specimens: 4th generation ELISA, confirmatory antibody assay, then p24 antigen assay. Positive specimens: 3rd or 4th generation ELISA, confirmation LIA. | Well-characterized archived HIV-postive/negative serum/plasma samples; fresh sepcimens, serum, plasma, and whole blood; supernantants from cultures of different HIV groups and subtypes; p24 antigen control | serum, plasma, and whole blood | 436 |
| Clifford B. Jones, et al | 2012 | United Kingdom | To compare the Determine Combo POCT with the Abbott Architect fourth generation laboratory assay and the Determine antibody only POCT, in high-risk patients attending a UK genitourinary clinic | Cross-sectional | High | 4th generation ELISA | MSM, persons from HIV endemic areas, injection drug users, or partners of persons in these groups | whole | 985 |
| Syria Laperche, et al | 2012 | France | To evaluate the Determine Combo in cell-derived supernatant samples of HIV-1 and HIV-2 subtypes, and in archived plasma specimens | Case-control | Not reported | p24 antigen assay | cell-derived supernatant samples of HIV-1 and HIV-2 subtypes, and in archived plasma specimens from individuals with acute-phase HIV-1 infection | serum and plasma (supernatant samples) | 119 |
| | | | | | | | | plasma (archived samples) | 20 |
| Kapila Bhowan, et al | 2011 | South Africa | To assess the performance of the Determine Combo RT against third generation rapid tests in antenatal and post-partum women | Cross-sectional | High | 3rd generation rapid test, discordant results followed by 4th generation testing | Antenatal and early post-partum women without documented HIV status | plasma | 1019 |
| | | | | | | | | whole blood | 380 |
| Yen T. Duong, et al | 2014 | Swaziland | To evaluate the performance of the Determine Combo test in detecting acute infections in the Swaziland Incidence Measurement Survey (SHIMS), compared to NAAT results. | Cross-sectional | High | NAAT for Ag-/ Ab+ and Ag- /Ab- specimens, rapid testing for Ab+ (Unigold) confirmed by EIA/NAAT if negative | Adults 18-49 years old, identified through a a nationally representative, household-based cross-sectional survey | Whole blood | 18 172 |

**Appendix Table 2: Raw data included in meta-analysis**

| Author | Year | Study Design | Sample | Test component | Sensitivity | Specificity | TP | FN | TN | FP |
|---|---|---|---|---|---|---|---|---|---|---|
| **Brauer** | 2013 | case-control | serum/plasma | antibody | 90.7 | 100 | 39 | 4 | 20 | 0 |
| **Brauer** | 2013 | case-control | serum/plasma | antigen | 10 | 100 | 3 | 27 | 20 | 0 |
| **Conway** | 2014 | cross-sectional | whole blood | overall | 87.2 | 99.4 | 34 | 5 | 3133 | 18 |
| **Conway** | 2014 | cross-sectional | whole blood | antibody | 94.4 | 99.6 | 34 | 2 | 3140 | 14 |
| **Conway** | 2014 | cross-sectional | whole blood | antigen | 0 | 99.8 | 0 | 9 | 3175 | 6 |
| **Pilcher** | 2013 | case-control | serum/plasma | overall | 54.4 | 98.8 | 31 | 26 | 80 | 1 |
| **Patel** | 2012 | case-control | serum/plasma | overall | 75.8 | NR | 25 | 8 | NR | NR |
| **Patel** | 2012 | case-control | serum/plasma | overall | 88.1 | 100 | 74 | 10 | 30 | 0 |
| **Rosenberg** | 2012 | cross-sectional | whole blood | antigen | 0 | 98.3 | 0 | 8 | 824 | 14 |
| **Rosenberg** | 2012 | cross-sectional | whole blood | antibody | 99.4 | 99.2 | 162 | 1 | 643 | 5 |
| **Rosenberg** | 2012 | cross-sectional | whole blood | overall | 95.9 | 97.8 | 164 | 7 | 818 | 18 |
| **Pavie** | 2010 | case-control | whole blood | overall | 95.8 | NR | 160 | 7 | NR | NR |
| **Kilembe** | 2012 | case-control | serum/plasma | antigen | 1.9 | NR | 1 | 51 | NR | NR |
| **Faraoni** | 2013 | case-control | serum/plasma | overall | 88.2 | 100 | 15 | 2 | 124 | 0 |
| **Faraoni** | 2013 | case-control | serum/plasma | antigen | 29.4 | 100 | 5 | 12 | 124 | 0 |
| **Faraoni** | 2013 | case-control | serum/plasma | antibody | 58.8 | 100 | 10 | 7 | 124 | 0 |
| **Chetty** | 2012 | cross-sectional | serum/plasma | antibody | 59.4 | 96.9 | 19 | 13 | 31 | 1 |
| **Chetty** | 2012 | cross-sectional | serum/plasma | overall | 94.1 | 96.9 | 208 | 13 | 31 | 1 |
| **Fox** | 2011 | case-control | serum/plasma | antigen | 50 | NR | 18 | 18 | NR | NR |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Fox** | 2011 | case-control | serum/plasma | overall | 72.2 | NR | 26 | 10 | NR | NR |
| **Masciotra** | 2013 | case-control | serum/plasma | overall | 92.5 | 97.9 | 124 | 10 | 94 | 2 |
| **Masciotra** | 2013 | case-control | serum/plasma | antigen | 72.2 | 96.2 | 26 | 10 | 75 | 3 |
| **Stekler** | 2013 | cross-sectional | whole blood | overall | 88.0 | 99.7 | 22 | 3 | NR | NR |
| **Beelaert** | 2010 | case-control | serum, plasma, and whole blood | antibody | 100 | NR | 169 | 0 | NR | NR |
| **Beelaert** | 2010 | case-control | serum, plasma, and whole blood | antigen | 86.6 | NR | 58 | 9 | NR | NR |
| **Beelaert** | 2010 | case-control | serum, plasma, and whole blood | overall | NR | 100 | NR | NR | 100 | 0 |
| **Jones** | 2012 | cross-sectional | whole blood | overall | 90.5 | 99.8 | 19 | 2 | NR | NR |
| **Laperche** | 2012 | case-control | serum/plasma | antigen | 21.0 | NR | 25 | 94 | NR | NR |
| **Laperche** | 2012 | case-control | serum/plasma | antigen | 60 | NR | 12 | 8 | NR | NR |
| **Bhowan** | 2011 | cross-sectional | serum/plasma | antibody | 100 | 99.8 | 90 | 0 | 927 | 2 |
| **Bhowan** | 2011 | cross-sectional | whole blood | antibody | 100 | 99.3 | 89 | 0 | 289 | 2 |
| **Duong** | 2014 | cross-sectional | Whole blood | antigen | 0 | 99.9 | 0 | 13 | 12345 | 12 |

*NR = data not reported

**Appendix Table 3: QUADAS-2 assessment**

| 1st author | Could the selection of patients have introduced bias? (Risk: low/high/unclear) | Concern that included patients do not match the review question? (Concern: low/high/unclear) | Could the conduct or interpretation of the index test (Determine) have introduced bias? (Risk: high/low/unclear) | Is there concern that the index test, its conduct, or interpretation differ from the review question? (low/high/unclear) | Could the reference standard, in its conduct, or its interpretation, have introduced bias? (Risk: low/high/unclear) | Any concern that the target condition as defined by the reference standard does not match the review question? (low/high/unclear) | Could the patient flow have introduced bias? (low/high/unclear) |
|---|---|---|---|---|---|---|---|
| Brauer | high | low | high | low | low | low | High |
| Conway | low | low | low | low | unclear | low | unclear - because of invalids |
| Pilcher | high - only selected suspected acute cases | low | high | low | low | low | unclear |
| Patel | high | low | low | low | low | low | high |
| Rosenberg | high - only suspected acute cases in one clinic | low | high | low | low | low | low |
| Pavie | high - only established infections included | high | low | low | low | high - established infection | high - many exclusions due to invalid results |
| Kilembe | high | low | high | low | low | low | low |
| Faraoni | high | low | high | low | low | low | low |
| Chetty | low | low | high | low | low | low | unclear |
| Fox | high | low | High | low | low | low | low |
| Masciotra | high | low | High | low | low | low | low |
| Stekler | unclear | low | High | low | high | low | high |
| Beelaert | high | low | low | low | unclear | low | high |
| Jones | low | low | high | low | low | low | low |
| Laperche | high | low | high | low | unclear | low | low |
| Bhowan | low | low | low | low | low | low | high - |
| Duong | low | low | low | low | high | low | high |

**WinBUGS output for Model 1 (overall, antigen and antibody components, sub-group analyses)**

| theta.mu | | | | | | |
|---|---|---|---|---|---|---|
| File | *mean* | *sd* | *MC error* | *median* | *2.5%* | *97.5%* |
| Antibody+CaseControl-sens-WinBUGSlog.txt | 0.6416 | 0.3627 | 0.003814 | 0.7918 | 4.561E-7 | 1.0 |
| Antibody+CaseControl-spec-WinBUGSlog.txt | 0.8894 | 0.2888 | 0.004555 | 0.9998 | 9.828E-7 | 1.0 |
| Antibody+CrossSectional-sens-WinBUGSlog.txt | 0.9428 | 0.1539 | 0.001468 | 0.9917 | 0.3556 | 1.0 |
| Antibody+CrossSectional-spec-WinBUGSlog.txt | 0.9938 | 0.01768 | 1.766E-4 | 0.9951 | 0.9844 | 0.9982 |
| Antibody+Serum-sens-WinBUGSlog.txt | 0.798 | 0.2692 | 0.002391 | 0.9167 | 0.01707 | 0.9999 |
| Antibody+Serum-spec-WinBUGSlog.txt | 0.9804 | 0.09917 | 0.001094 | 0.9979 | 0.8388 | 1.0 |
| Antibody+Whole-sens-WinBUGSlog.txt | 0.9175 | 0.2229 | 0.002934 | 0.9939 | 0.01654 | 1.0 |
| Antibody+Whole-spec-WinBUGSlog.txt | 0.983 | 0.08498 | 0.00137 | 0.9947 | 0.9339 | 0.9994 |
| Antibody-sens-WinBUGSlog.txt | 0.9373 | 0.1127 | 0.001044 | 0.973 | 0.6068 | 0.9994 |
| Antibody-spec-WinBUGSlog.txt | 0.9951 | 0.006049 | 5.712E-5 | 0.9955 | 0.9901 | 0.9984 |
| Antigen+CaseControl-sens-WinBUGSlog.txt | 0.2994 | 0.1565 | 0.001288 | 0.2775 | 0.05873 | 0.674 |
| Antigen+CaseControl-spec-WinBUGSlog.txt | 0.9203 | 0.2253 | 0.002761 | 0.9974 | 0.009292 | 1.0 |
| Antigen+CrossSectional-sens-WinBUGSlog.txt | 0.03697 | 0.1601 | 0.002251 | 2.202E-4 | 2.538E-7 | 0.7394 |
| Antigen+CrossSectional-spec-WinBUGSlog.txt | 0.9547 | 0.1642 | 0.002145 | 0.9968 | 0.247 | 1.0 |
| Antigen+Serum-sens-WinBUGSlog.txt | 0.2994 | 0.1565 | 0.001288 | 0.2775 | 0.05873 | 0.674 |
| Antigen+Serum-spec-WinBUGSlog.txt | 0.9203 | 0.2253 | 0.002761 | 0.9974 | 0.009292 | 1.0 |
| Antigen+Whole-sens-WinBUGSlog.txt | 0.03697 | 0.1601 | 0.002251 | 2.202E-4 | 2.538E-7 | 0.7394 |
| Antigen+Whole-spec-WinBUGSlog.txt | 0.9547 | 0.1642 | 0.002145 | 0.9968 | 0.247 | 1.0 |
| Antigen-sens-WinBUGSlog.txt | 0.1483 | 0.1143 | 0.001064 | 0.1234 | 0.01079 | 0.4418 |
| Antigen-spec-WinBUGSlog.txt | 0.9926 | 0.03118 | 2.708E-4 | 0.9967 | 0.9677 | 0.9998 |
| Overall+CaseControl-sens-WinBUGSlog.txt | 0.8367 | 0.07616 | 6.276E-4 | 0.8491 | 0.65 | 0.9451 |
| Overall+CaseControl-spec-WinBUGSlog.txt | 0.9753 | 0.1072 | 0.001379 | 0.9959 | 0.7914 | 1.0 |
| Overall+CrossSectional-sens-WinBUGSlog.txt | 0.925 | 0.04097 | 6.231E-4 | 0.9324 | 0.8394 | 0.969 |
| Overall+CrossSectional-spec-WinBUGSlog.txt | 0.9476 | 0.1523 | 0.002185 | 0.9876 | 0.3717 | 0.9998 |
| Overall+Serum-sens-WinBUGSlog.txt | 0.8317 | 0.07609 | 6.502E-4 | 0.8431 | 0.6504 | 0.9418 |
| Overall+Serum-spec-WinBUGSlog.txt | 0.9848 | 0.0565 | 5.926E-4 | 0.9926 | 0.9469 | 0.9998 |
| Overall+Whole-sens-WinBUGSlog.txt | 0.9301 | 0.04229 | 6.494E-4 | 0.9378 | 0.8435 | 0.9733 |
| Overall+Whole-spec-WinBUGSlog.txt | 0.8361 | 0.3193 | 0.004574 | 0.9879 | 1.71E-6 | 1.0 |
| Overall-sens-WinBUGSlog.txt | 0.8814 | 0.03493 | 3.234E-4 | 0.8854 | 0.8013 | 0.9375 |
| Overall-spec-WinBUGSlog.txt | 0.989 | 0.0184 | 1.719E-4 | 0.9906 | 0.9729 | 0.9984 |

### Stata output for Model 2 (overall, antigen, and antibody components)

```
//overall pooled//
. metandi tp_num1 fp_num1 fn_num1 tn_num1 if component_num==3, gllamm

True  positives: tp_num1              False positives: fp_num1
False negatives: fn_num1              True  negatives: tn_num1

Meta-analysis of diagnostic accuracy

Log likelihood   = -35.289564                     Number of studies =        7
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Bivariate    |
  E(logitSe) |   2.053607   .3701995                      1.32803    2.779185
  E(logitSp) |   4.760624   .3528367                     4.069076    5.452171
Var(logitSe) |   .8030789    .479358                      .2492733    2.587263
Var(logitSp) |   .4955538   .4302006                      .0903951    2.716668
Corr(logits) |         -1          .                             .           .
-------------+----------------------------------------------------------------
Summary pt.  |
          Se |   .8863116   .0373025                      .7905145    .9415406
          Sp |   .9915124   .0029693                      .9831941    .9957313
         DOR |   910.7157    262.149                       518.042    1601.034
         LR+ |   104.4241   33.67882                     55.49727    196.4853
         LR- |   .1146616   .0373878                      .0605154    .2172552
       1/LR- |   8.721316   2.843769                     4.602882    16.52472
------------------------------------------------------------------------------
Covariance between estimates of E(logitSe) & E(logitSp)  -.0893421


. //antigen pooled//
. metandi tp_num1 fp_num1 fn_num1 tn_num1 if component_num==2, gllamm

Log likelihood   = -28.41614                     Number of studies =        6
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Bivariate    |
  E(logitSe) |  -2.322671   1.147822                     -4.572361   -.0729816
  E(logitSp) |   5.282402    .629596                      4.048416    6.516387
Var(logitSe) |   4.134625   3.847469                      .6673471    25.61654
Var(logitSp) |   1.646312   1.085625                      .4520699     5.99541
Corr(logits) |  -.7369785   .3611966                     -.9864372    .5411062
-------------+----------------------------------------------------------------
Summary pt.  |
          Se |   .0892626    .093312                      .0102278    .4817627
          Sp |   .9949455   .0031662                      .9828493    .9985232
         DOR |   19.29277   20.07938                     2.508925    148.3547
         LR+ |   17.65991    16.8807                     2.712344    114.9826
         LR- |   .9153641   .0925521                      .7508086    1.115985
       1/LR- |   1.092461   .1104583                      .8960691    1.331897
------------------------------------------------------------------------------
Covariance between estimates of E(logitSe) & E(logitSp)    -.31534
```

```
. //antibody pooled//
. metandi tp_num1 fp_num1 fn_num1 tn_num1 if component_num==1, gllamm


Log likelihood   = -29.93053                     Number of studies =        7
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Bivariate    |
  E(logitSe) |   3.642835   1.450737                      .7994432    6.486227
  E(logitSp) |   5.364565   .2113343                      4.950358    5.778773
Var(logitSe) |   6.247469   5.131216                      1.249077    31.24777
Var(logitSp) |   .0038278   .0397485                      5.55e-12     2642135
Corr(logits) |          1          .                             .           .
-------------+----------------------------------------------------------------
Summary pt.  |
          Se |   .9744898   .0360645                      .6898554     .998478
          Sp |   .9953423   .0009797                      .9929689     .996917
         DOR |   8163.272   12184.41                      437.8947    152180.4
         LR+ |   209.2213   45.65178                      136.4191    320.8757
         LR- |   .0256296   .0362365                      .0016042    .4094651
       1/LR- |    39.0174   55.16493                      2.442211    623.3522
------------------------------------------------------------------------------
Covariance between estimates of E(logitSe) & E(logitSp)   .0392622
```