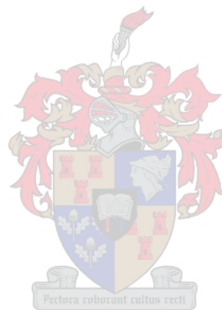
Evaluation of the Mean Duration of Recent Infection (MDRI) and the False Recent Rate (FRR) for the Limiting Antigen Avidity Enzyme Immune Assay (LAg) and Bio-Rad HIV ½ Plus O Avidity Incidence Assay (BRAI)

Elizabeth Gonese

Dissertation presented for the Degree of Doctor in Epidemiology in the Faculty of Medicine and Health Sciences at Stellenbosch University



Supervised by:

Professor John Hargrove	South Africa Centre for Epidemiological Modelling and Analysis, Stellenbosch University
Professor Jean Nachega Dr Gert van Zyl	Faculty of Medicine and Health Sciences, Stellenbosch University, Western Cape
Dr Peter H. Kilmarx	Faculty of Medicine and Health Sciences, Stellenbosch University, Western Cape Division of HIV/AIDS and TB Prevention; Centers for Disease Control and Prevention Zimbabwe.

December, 2017

Stellenbosch University https://scholar.sun.ac.za

Stellenbosch University https://scholar.sun.ac.za

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Copyright © 2017 Stellenbosch University All Rights Reserved

Stellenbosch University https://scholar.sun.ac.za

ABSTRACT

Background: The evaluation of laboratory assays in estimating HIV incidence has become a priority because of the complexity of HIV epidemics and the need to measure the impact of public health interventions targeting reduction of HIV incidence. Biomarkers should have test properties that allow the lowest possible False Recent Rate (FRR, or probability of diagnosing a long-term infection as recently infected) over the longest possible period (Mean Duration of Recent Infection or MDRI) during which the case is considered as a recent infection.

Methods: We compared the BED Capture Enzyme Immunoassay (BED), Sedia Limiting Antigen (LAg) and Bio-Rad HIV ½ Plus O Avidity Incidence Assay (BRAI) using samples from a prospective cohort trial, the Zimbabwe Vitamin A for Mothers and Babies Project (ZVITAMBO) 1997–2000. We determined MDRI using 591 samples from 184 seroconverting women, and determined FRR by testing 2825 cases known to be HIV-positive for >12 months. We used these results to estimate HIV incidence over the first 12 months postpartum, and during the period prior to childbirth.

Results: At recommended cut-offs MDRI values were: BRAI, 135 days (120 – 151) at Avidity Index (AI) 30%; LAg, 104 days (98 - 110) at ODn cut-off 1.5; BED,188 days (180 - 196) at ODn cut-off 0.8. All error bounds in this thesis signify 95% confidence intervals. The coefficients of variation (CV) of the MDRI estimates for BRAI, LAg and BED were 5.9%, 2.9% and 2.1%, respectively. Corresponding FRRs were 1.1% (0.7-1.5) for BRAI, 0.6% (0.3-0.9) for LAg and 4.8% (4.1-5.7) for BED. MDRI and FRR estimates, all derived using postpartum women, were lower than in other published studies. Using original ZVITAMBO HIV diagnoses, adjusted HIV incidence over the first 12 months postpartum was estimated as; BRAI, 2.7% (1.8-3.7); LAg, 3.7% (2.7-4.8); BED, 3.6% (2.4 -4.9). Follow-up incidence was 3.4% (3.0-3.8).

When cases with viral load <1000 copies/ml were defined as long-term infections, regardless of serological biomarker level, FRRs were; BRAI, 1.0% (0.7-1.5); LAg, 0.2% (0.2 -0.7); BED 3.8% (3.1-4.6). MDRIs were; BRAI, 133 days (113-154); LAg, 101 days (87-115); BED, 177 days (155 - 199). Corresponding incidences, unadjusted for FRR, were: BRAI,

Stellenbosch University https://scholar.sun.ac.za

3.9% (2.9-4.9); LAg, 3.1% (2.1-4.0); BED, 6.2 % (5.0-7.3). Adjusted estimates were 2.7% (1.5-4.0), 2.5% (1.6-3.5) and 2.6% (1.6-3.7) respectively.

At baseline, with no follow-up estimate for comparison, adjusted incidence for serological biomarkers used alone were; BRAI, 8.1% (6.6-9.7); LAg, 6.9% (5.7-8.1); BED 6.7% (5.5-7.9). When viral load was also used, the adjusted and unadjusted incidence estimates were; BRAI, 7.3% (5.7-8.8) and 8.4% (6.8-10.0); LAg, 5.1% (3.9-6.3) and 5.7% (4.5-6.9); BED, 5.4% (4.1-6.7) and 8.6% (7.3-10.0).

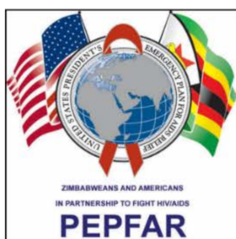
Conclusion: At recommended cut-offs; BRAI FRR was 1.9 times higher than that of LAg. BRAI MDRIs were also 1.3 times higher, but with a relative standard error 2.4 times as high. Postpartum BRAI incidence estimates were consistently lower than follow-up estimates. Adjusted biomarker estimates under-estimated follow-up incidence when we used viral load in combination with either serological test.

Stellenbosch University https://scholar.sun.ac.za

Funding Acknowledgements

This project was supported by funding from the President's Emergency Plan for AIDS Relief (PEPFAR) through the US Centers for Disease Control and Prevention under terms of grant 5U2GGH00315-04 to Department of Community Medicine, University of Zimbabwe and a sub-grant to ZVITAMBO Project. The original ZVITAMBO study was conducted under the auspices of the Ministry of Health and Child Care Zimbabwe.

Ms Elizabeth Gonese received a bursary from South Africa DST/NRF Centre of Excellence for Epidemiological Modelling and Analysis (SACEMA), Stellenbosch University for her PhD studies.



Stellenbosch University https://scholar.sun.ac.za

Acknowledgements

I would like to express my heartfelt gratitude to all my supervisors for walking this journey with me and for their unwavering support. John Hargrove, Gavin Hitchcock, Alex Welte, Lynnemore Scheepers and Amanda October: thank you for helping me navigate the complex waters of the joint registration at SACEMA and Department of Medicine and Health Studies office and logistical support for all the sessions I was able to attend in Cape Town.

I will forever be indebted to John Hargrove as my academic supervisor and mentor and Peter Kilmarx my professional and academic mentor. You both went the extra mile, helped me see things scientifically and held my hand to the end. Peter your passion for science is a bright torch that blazes bright, you saw things that were not obvious to a novice like me.

I would like acknowledge the ZVITAMBO laboratory and management team. Special thanks to Kuda Mutasa, Robert Ntozini and Jean Humphrey for allowing me to use your samples, tap into your laboratory skills that are second to none in Zimbabwe and generally keeping things on track. I thank my colleagues in CDC Atlanta, Bharat Parekh, Trudy Dobbs and Silvina Masciotra for the technical support for this project.

A big thank you for statistical support received from Cari van Van Schalkwyk, Eduard Grebe and John Hargrove. I would never have been able to make sense of that data.

Finally, a big thank you to my family for asking almost every day when I was going to finish working on this. Thanks for affording me time and space complete this work. I wish all my siblings, children, nephews and nieces' success in their own studies. Lastly to God be the Glory Great things He has done.

Stellenbosch University https://scholar.sun.ac.za

Collaborators

CDC
Dr Bharat Parekh
Dr Yen Duong (Pottinger)
Dr Trudy Dobbs
Dr Anindya De
Dr Michele Owen
Ms Silvina Masciotra
ZVITAMBO
Dr J. Humphrey
Mrs K. Mutasa
Mr Robert Ntozini
SACEMA
Prof Alex Welte
Mrs Cari Van Schalkwyk

Author's contribution to this work

Stellenbosch University https://scholar.sun.ac.za

This evaluation was of great interest to a number of entities working on HIV incidence specifically;

1. The International Laboratory Branch, CDC Atlanta have previously supported the development and evaluation of BED using the ZVITAMBO samples. They are currently the lead developers of LAg avidity assay
2. The Diagnostics and Incidence Team, HIV Laboratory Branch, CDC Atlanta which is the CDC Atlanta Domestic Laboratory Program were the lead developers of the Bio-Rad assay which up to 2016 was in use in their domestic incidence work
3. CDC Zimbabwe office, because they are routinely involved in large scale cross-sectional surveys were keenly interested in participating in the local validation of laboratory assays that could potentially be used in impact evaluation
4. South Africa Centre for Epidemiological Modelling and Analysis, Stellenbosch University because of their involvement in previous evaluations of laboratory incidence assay through the Gates Foundation funding of CEPHIA Trials
5. Zvitambo Institute for Maternal and Child Health Research, Harare, Zimbabwe, the owners of the ZVITAMBO cohort samples that were used in the BED evaluation in 2002 and
6. Faculty of Medicine and Health Sciences, Stellenbosch University because of the multi-disciplinary nature of the study

The author of this Thesis developed the protocol for the evaluation with support from the major entities and individual collaborators listed above. She was also central to setting up the collaboration between CDC Zimbabwe, CDC Atlanta, Zvitambo and SACEMA. She also oversaw the development of funding modalities, working with the Department of Community Medicine of the University of Zimbabwe and worked on the acquisition of data for this evaluation with support from Laboratory Scientists from CDC Atlanta. The author contributed to data analysis in determining the main outcomes of the study. Statistical support was received from SACEMA. The author contributed to interpreting the results and she wrote the thesis.

Stellenbosch University https://scholar.sun.ac.za

Contents

1.0 SIGNIFICANCE OF THIS EVALUATION 1

2.0 INTRODUCTION 3

3.0 LITERATURE REVIEW 6

3.1 Modelling of HIV incidence trends using prevalence data 6
3.2 Measuring incidence using prospective cohort data in combination with modelling 8
3.3 Laboratory based estimates of HIV incidence 10
3.4 Utility of multi-assay algorithms in improving measurements from single laboratory assays 15
3.5 Measuring HIV Incidence using multiple methods 16

4.0 DESCRIPTION OF CANDIDATE LABORATORY ASSAYS 17

4.1 BED-Capture Enzyme Immune Assay (Sedia HIV-1 BED Incidence EIA, Cat. No. 1000) 17
4.2 Sedia™ HIV-1 Limiting Antigen Avidity Enzyme Immuno Assay (LAg-Avidity EIA) 17
4.3 Modified BioRad Genetic Systems HIV-1/HIV-2 plus O avidity-based assay (BRAI) 18

5.0 METHODS 19

5.1 Description of ZVITAMBO cohort samples and historical tests conducted 19
5.2 HIV tests and VL tests conducted on ZVITAMBO samples 20
5.3 Characterisation of BED using ZVITAMBO samples 20
5.4 Ethical considerations for ZVITAMBO samples 21

6.0 METHODS IN THE EVALUATION OF LAG-AVIDITY EIA AND BRAI AVIDITY ASSAY 22

6.1 Sample viability test 22
6.2 Laboratory testing using the Sedia™ HIV-1 LAg-Avidity EIA (Catalog No. 1002) 24
6.2.1 Use of viral load testing with LAg assay 25
6.2.2 Retesting HIV serology for samples with LAg ODn < 0.400 25
6.3 Laboratory testing using BioRad Genetic Systems HIV-1/HIV-2 plus O EIA (BRAI Avidity Assay) 25
6.3.1 BRAI Assay Invalid Results 26
6.4 Quality Control of Laboratory Assays 27
6.5 Sample selection for evaluation 27
6.6 Statistical analysis 29
6.6.1 Samples in MDRI analyses 29
6.6.2 Statistical analysis of MDRI 29
6.6.3 Samples in FRR analyses 31
6.6.4 Statistical analysis of FRR 32
6.6.5 Incidence calculations 33
6.6.6 Follow-up incidence 34
6.6.7 Coefficient of Variation and Confidence Intervals 35

7.0 ANALYSIS OF HIV SEROLOGY IN THE EVALUATION OF BED, LAG AND BRAI 37

7.1 Sample Viability Test 38
7.2 Review of HIV serology results 39
7.2.1 Retesting of cases providing an "invalid" BRAI test result at Baseline 42
7.2.2 Retesting of cases providing an "invalid" BRAI test result at Visit 5 45
7.3 Discussion on the implications of misclassification of HIV result in the BED, BRAI and LAg evaluation 48
7.3.1 General considerations 48
7.3.2 Specific implications for the ZVITAMBO study 49
7.4 Optical density patterns among seroconverters as observed using BED, LAg and BRAI 53
7.5 Discussion and conclusion on variation in optical density readings for BED, LAg and BRAI 57

8.0 DETERMINATION OF MEAN DURATION OF RECENT INFECTION (MDRI) FOR BED, LAG AND BRAI ASSAYS 59

8.1.1 Introduction 60
8.1.2 Effects of varying the value of Time (T) and Number of Samples (ns) on estimated values of MDRI 61
8.1.3 Methods 63

Stellenbosch University https://scholar.sun.ac.za

8.2 Results of MDRI Calculations 68
8.2.1 Changes in biomarker optical density as a function of time since seroconversion 69
8.2.2 MDRI based on Turnbull Survival Analysis (SA) 72
8.2.3 MDRI Using Non Linear Mixed Methods (NLMM) 76
8.2.4 MDRI Estimates Calculated When T=1 or 2 years 80
8.2.5 Comparison of MDRI based on NLMM, SA, Binomial Regression and r/s for the three assays 82
8.2.6a Impact of adding VL to MDRI using Survival Analysis 86
8.2.6b Comparison of MDRI derived for assay and assay with VL based on Binomial regression, Survival Analysis and R/S Method 88
8.2.7 Exploration of low MDRI for BED, LAg and BRAI arising from ZVITAMBO postpartum cohort 90
8.2.7 a MDRI for BED, LAg and BRAI by stage of seroconversion 91
8.2.7b Comparison of MDRI by Vitamin A arm in original ZVITAMBO trial 93
8.2.8 Comparison of MDRI for Clade C and Clade B using BRAI Assay 94
8.3 Discussion of results of MDRI for BED, LAg and BRAI in this evaluation 95

9.0 FALSE RECENCY RATE (FRR) OF HIV INFECTION 99

9.1 Distribution of ODn readings for BED and LAg and AI for BRAI 99
9.2 FRR for BED, LAg and BRAI when Original and New Data are used with or without viral load 101
9.2.1 Viral loads for cases testing HIV positive at Baseline and Visit 5 101
9.3 FRR by Age Group of Mother 108
9.4 Comparison of MDRI and FRR evaluation results to other studies 109
9.5 Discussion of the evaluation FRR for BED, LAg and BRAI assays 110

10.0 INCIDENCE ESTIMATES DERIVED USING THE THREE BIOMARKERS 112

10.1 Preamble 112
10.2 HIV incidence rates over the first 12-months postpartum as estimated by follow-up and calculated using BED, LAg or BRAI. 115
10.2.1 Estimates using only serological biomarkers: original data. 116
10.2.2 Estimates using serological biomarkers plus viral load. 119
10.3 Baseline HIV Incidence using BED, LAg and BRAI with or without viral load 125
10.3.1 Using serological biomarkers only 125
10.3.2 Using viral load in conjunction with a serological biomarker for baseline estimates of incidence 134
10.4: HIV incidence as a function of age or parity 136
10.5 Comparison of HIV incidence estimated from samples at Baseline and at 12-months Postpartum 138
10.6 Summary HIV Incidence at Baseline and 12 Months Postpartum 140
10.7 Discussion of HIV incidence estimates 142
10.7.1 Using serological biomarkers only to identify recent infections 142
10.7.2 Using serological biomarkers together with viral load to identify recent infections 144
10.7.3 High HIV incidence measured at Baseline 145

11.0 LIMITATIONS OF THE STUDY 146

12.0 CONCLUSION 147

13.0 RECOMMENDATIONS FROM EVALUATION 151

15.0 GLOSSARY 152

Stellenbosch University https://scholar.sun.ac.za

TABLE OF FIGURES

Figure 5-1: Schematic diagram of the ZVITAMBO study enrolment phases 19
Figure 6-1: Schematic diagram of the BED, LAg and BRAI evaluation using ZVITAMBO project samples 23
Figure 6-2: Actual ZVITAMBO samples tested in the evaluation of BED, LAg and BRAI .. 28
Figure 7-1: Results of ZVITAMBO sample viability test 38
Figure 7-2: Distribution of optical density (ODn) readings (BED and Lag) and AI (BRAI) for seroconverting mothers 53
Figure 7-3: Distribution of ODn for BED and LAg and AI for BRAI by estimated days since seroconversion 54
Figure 7-4: Natural logarithm of optical density (BED, LAg) and avidity index (BRAI) readings against days since seroconversion 56
Figure 8-1: Natural logs of BED optical density readings against days since seroconversion 69
Figure 8-2: Natural logs of LAg optical density readings against days since seroconversion for selected cases where the seroconverting mother produced at least six HIV positive blood samples after seroconversion 70
Figure 8-3: Natural logs of BRAI log optical density readings against days since seroconversion for selected cases 71
Figure 8-4: MDRI for BED at different cut-offs and varying t_0 using survival methods 73
Figure 8-5: MDRI for BRAI and LAg at different cut-off using Turnbull SA methods 74
Figure 8-6: MDRI Estimates using NLMM for BED and LAg 77
Figure 8-7: Comparison of MDRI estimates for BED and LAg obtained using SA and NLMM 78
Figure 8-8: Comparison of MDRI for LAg using variations of Sweeting's Methods 79
Figure 8-9a: Comparison of MDRI estimates for LAg when T=1 and T=2 80
Figure 8-10: Comparison of MDRI based on NLMM, SA, Binomial Regression and the Ratio (R/S) for BED, LAg and BRAI [ns>=2; t_0 <=120 days; T = 1 year] 84
Figure 8-11: MDRI for BED, LAg and BRAI and VL MAA using Turnbull's survival analysis (ns=2 and t_0=120) 87
Figure 8 -12: Comparison of MDRI for BED, LAg, BRAI with VL estimated using Binomial regression, SA or R/S. [ns>=2; t_0<=120 days; T = 1 year] 89
Figure 8-13: MDRI for BED, LAg, and BRAI by stage of seroconversion and compared to published results 92
Figure 8-14: Comparison of MDRI by Vitamin A status for LAg 93
Figure 8- 15: MDRI obtained using BRAI on Clade C (ZVITAMBO) compared to MDRI on Clade B (USA) 94
Figure 9-1: Distribution of optical density (BED, LAg) or avidity index (BRAI) readings by assay 100
Figure 9-2a: Distributions of viral loads for samples taken at A. Baseline, for cases testing HIV testing positive at that time 102
Figure 9-2b: Visit 5 for cases testing HIV testing positive both at Baseline and at Visit 5 .. 102
Figure 9-3: FRR by different cut-offs for BED, LAg and BRAI assay either used alone or in combination with VL 105
Figure 9-4: Summary of FRR by cut-off for BED, LAg and BRAI 106
Figure 9-5: Summary FRR vs MDRI for BED, LAg and BRAI 107
Figure 9-6: FRR by maternal age 108
Figure 10-1: HIV Incidence over the first 12 months postpartum derived using ZVITAMBO original data. 117

Stellenbosch University https://scholar.sun.ac.za

Figure 10-2: Adjusted HIV incidence estimates for BED, LAg and BRAI comparing Original and New Data ... 118
Figure 10-3: Annual risk of infection as calculated using LAg with or without VL and using either the original or the new HIV diagnoses. ... 120
Figure 10-4: Annual risk of infection as calculated using BRAI with or without VL and using the original and new HIV diagnoses. ... 122
Figure 10-5: Annual risk of infection for BED, LAg and BRAI with VL using SA methods for MDRI and Original Data ... 123
Figure 10-6: Annual risk of infection for BED, LAg and BRAI with VL using SA methods for MDRI and New Data ... 124
Figure 10-7: Baseline (V0) HIV incidence for BED by cut-off using original and new data 126
Figure 10-8: Baseline (V0) HIV incidence for LAg by cut-off using original and new data 128
Figure 10-9: Baseline (V0) incidence for BRAI by cut-off using original and new data ... 129
Figure 10-10: Baseline incidence estimates for BED, LAg (NLMM) and BRAI (Survival) unadjusted and adjusted estimates (original data) ... 131
Figure 10-11: Baseline incidence estimates for BED, LAg (NLMM) and BRAI (Survival) unadjusted and adjusted estimates (New data) ... 133
Figure 10-12: Comparison of baseline HIV incidence for BED, LAg and BRAI assays used either alone or in combination with VL ... 135
Figure 10-13: Baseline incidence estimates and maternal age for BED, LAg, and BRAI ... 136
Figure 10-14: Baseline incidence estimates by parity for BED, LAg and BRAI ... 137
Figure 10-15: Comparison of adjusted baseline and postpartum incidence estimates obtained by application of the BED, LAg and BRAI assays to the ZVITAMBO original data. ... 139

Stellenbosch University https://scholar.sun.ac.za

LIST OF TABLES

Table 7-1: Results for 20 case originally diagnosed as HIV positive at baseline – but testing "invalid" by BRAI 41
Table 7-2: Retesting of samples collected at Baseline (Visit 0) 43
Table 7-3: Retesting of samples collected at Visit 5... 46
Table 8-1: Summary MDRI for LAg and BRAI using Turnbull SA methods 75
Table 9-1: Distribution of cases testing as "recent" or "long term" infections (using LAg with a cut-off of 0.8) as function of log(10) viral load 103
Table 9-2: Distribution of cases testing as "recent" or "long term" infections (using LAg with a cut-off of 0.8) as function of log(10) viral load 104
Table 9-3: Estimates of MDRI and FRR values for clade C HIV made in this study compared to studies of Kassanjee et al. (2014) and Duong et al. (2015). 109
Table 10-1: Summary of estimates for BED, LAg and BRAI serology (Original data) 140
Table 10-2: Summary of estimates for BED, LAg and BRAI when VL is included (Original data).. 140

Stellenbosch University https://scholar.sun.ac.za

Acronyms

AI	Avidity Index
ART	Antiretroviral Therapy
BED	BED Capture Enzyme Immunoassay assay
BRAI	Bio-Rad HIV ½ Plus O Avidity Incidence Assay
CEPHIA	Consortium for the Evaluation and Performance of HIV Incidence Assays
CV	Coefficient of Variation
FRR	False Recency Rate
Gp	Gamma-protein
HAART	Highly active antiretroviral therapy
HPTN	HIV Prevention Trials Network
IDE	Immuno-dominant enzyme
Lag	Limiting Antigen Avidity Enzyme Immuno-Assay
LMM	Linear Mixed Model
Ln	Natural logarithm (log to the base e)
MAA	Multi-assay algorithm
MDRI	Mean Duration of Recent Infection
NLMM	Non Linear Mixed Model
NPV	Negative Predictive Value
PPV	Positive Predictive Value
OD	Optical Density
ODn	Normalised Optical Density
RITA	Recent Infection Testing Algorithms
SA	Survival Analysis
STD	Sexually Transmitted Disease
t_0	Time between the last negative and first positive HIV test
TRI	Test of Recent Infection
USA	United States of America
VL	Viral Load
ZVITAMBO	Zimbabwe Vitamin A for Mother and Baby

Stellenbosch University https://scholar.sun.ac.za

Stellenbosch University https://scholar.sun.ac.za

1.0 Significance of this evaluation

Zimbabwe participated in the characterization of the BED laboratory assay, using samples collected in the follow-up study of postpartum women in the Zimbabwe Vitamin A for Mother and Baby (ZVITAMBO) project.[1,2] Using the BED analysis, researchers reported a mean duration of recent infection (MDRI) of 196 (188-204) days for the ZVITAMBO cohort at a cut-off of 0.8. They used this to calculate an unadjusted annual incidence of 7.6% at 12 months postpartum, very much higher than the follow-up estimate of HIV incidence of 3.4% (95% CI 3.0-3.8). The researchers attributed this discrepancy to a high false recency rate (FRR). That is to say, of women known to have been HIV positive for more than 12 months, 5.2% (4.4-6.1) tested recent by BED. When they accounted for this FRR, the adjusted BED estimate of incidence matched closely the follow-up estimate.

In the current evaluation of the LAg and BRAI assays, we used the identical ZVITAMBO samples used to evaluate the BED assay, thus providing an opportunity for characterising MDRI and FRR values of the new assays and comparing their performance with the previously characterised BED assay. The study is unique in that it utilises samples collected from women within 72 hours of giving birth. This group of women had a different physiological make-up that is associated with period of pregnancy and are different from women in the general population. Follow-up samples collected from the same women during the postpartum period uniquely relate to the period during which there is active immune reconstitution. Moreover, the samples are highly unusual in that the research was carried out in a setting where anti-retroviral therapy (ART) was not generally available: there was, therefore, no confusing effect of ART on the probability that a case tested "recent".

This report is a comparison of the performance of the three assays based on the MDRI and FRR metric. For each assay, we compare the calculation of MDRI using different statistical methods. We varied the minimum number (*ns*) of samples required for a case to be included in the analysis, the pre-set cut-off (*C*), and the maximum time allowable between the last negative and first positive sample (t_0). The emphasis on the CV in this work (particularly for MDRI estimates) stems from the fact that the MDRI point estimates are generally similar, regardless of the methods used to estimate them. Therefore, CV is used to highlight

Stellenbosch University https://scholar.sun.ac.za

differences in precision and we use the 95% confidence interval to show any significant differences between the point estimates.

We explored the impact of physiological differences on the MDRI by calculating the MDRI for women who seroconverted at differing times postpartum. The ZVITAMBO trial randomly assigned women at baseline to two study arms; the Vitamin A group and Placebo group in a clinical trial to study the effect of Vitamin A on the following medical outcomes for mothers and their infants: acquisition of HIV infections; survival of HIV positive case; survival of HIV negative cases. We compared MDRI by exposure to Vitamin A.

A small proportion of samples that required retesting of HIV serology and remained discordant necessitated that we explore analysis in which we used the "Original" HIV diagnoses and where we used a "New" dataset incorporating changed HIV diagnoses. We calculated MDRIs, FRRs and HIV incidence estimates for the two datasets. Finally, this evaluation provided us an opportunity to compare observed HIV incidence from a follow-up cohort with calculated HIV incidence in which we applied the calculated test properties either in an assay using only a serological biomarker or a multi-assay algorithm where we used viral load as an additional screening tool.

Stellenbosch University https://scholar.sun.ac.za

2.0 Introduction

The evaluation of laboratory assays for measuring HIV incidence has become a priority because of the complexity of HIV epidemics and the need to measure the impact of public health interventions that are currently being rolled out globally.[1, 2,3, 4,5] Although follow-up cohort studies provide a direct measure of HIV incidence, their high costs, and biases due selection and loss to follow-up are problematic.[3] Moreover, ethical considerations demand that one should be making every effort to ensure that persons found to be HIV negative remain uninfected. The researcher is thus constrained to attempt to alter the course of infection and thus the incidence that s/he is trying to measure. All of these effects make follow-up increasingly less desirable as a method for estimating HIV incidence. Mathematical modelling using prevalence and mortality data from population based surveys and antenatal clinic surveillance have been used to estimate incidence.[4,5,6,7,8] Model outputs are, however, only as good as the data and parameters that are used to fit the models.[9]

Laboratory assays that use samples collected in population-based surveys, offer an attractive alternative option for estimating HIV incidence.[2,10,11,12] The idea behind such an approach is that we should use clinical tests not only to decide whether a given sample has been taken from a person who is either HIV positive or HIV negative, but also whether HIV positive cases have either been infected recently, or are long-term infections.

Globally, there are efforts towards improving such laboratory-based assays to measure trends in HIV incidence using cross-sectional survey samples. [13,14,15,16,17] Two main parameters, the mean duration recent infection (MDRI) and false recency rate (FRR) are required for calculating HIV incidence from such tests for recency. They are therefore critical in defining a laboratory assay's utility in the estimation of HIV incidence, either as a single assay or in combination with other biomarkers in a multi-assay algorithms (MAA).[2]

The MDRI is the mean time that an infected person remains in a state of recent infection while infected for less than a predefined time (T). During this period, an HIV infected person's blood sample returns an optical density (ODn), or avidity index (AI), below some pre-selected level, referred to as the cut-off value C.[2]

Stellenbosch University https://scholar.sun.ac.za

The FRR is defined as the proportion of individuals infected for greater than time (*T*, commonly set to 365 or 730 days) who are misclassified as recently infected. The ideal laboratory biomarker would have FRR=0; failing that, the FRR should be as low as possible, in order to minimise the number of long-term infections incorrectly diagnosed as recent. The FRR parameter is known to be sensitive to variations in HIV types i.e., viral diversity, disease stage as measured by CD4 cell count or viral load (VL), and use of antiretroviral therapy (ART).[18,19]

Both the MDRI and FRR values will increase monotonically with increasing *C*: the higher the value of *C*, the longer it will take the biomarker level to exceed this mark and, thus, the longer the MDRI and the greater the number of case testing as recent at any time. Equally, the higher the value of *C*, the larger the proportion of cases that could still test recent at time *T* after seroconversion.

The requirements of a biomarker make competing demands of the FRR and the MDRI. We want a biomarker where we can select a value of *C* that is low enough to minimise the FRR – i.e., where almost all patients with long-term infections have biomarker levels $> C$. Lowering *C* will mean, however, a reduction in the numbers of HIV positive cases that test recent and thus an increase in the sample size required for accurate estimation of incidence over the predefined time *T*. [20,21] The ultimate value of a laboratory assay lies in its ability to provide an accurate estimate of HIV incidence and the right balance in its MDRI and FRR critically determines this utility.[20]

Our evaluation characterised two avidity based assays – the Limiting Antigen Avidity Enzyme Immuno-Assay (LAg)[22,23] and the Bio-Rad Genetic Systems HIV-1/HIV-2 plus O Enzyme-linked Avidity Incidence Assay (BRAI).[24] We estimated the MDRI and FRR, and calculated laboratory-based incidence rates, using samples from a well-characterised cohort of postpartum women infected predominantly with Clade C HIV. We used samples collected during the Zimbabwe Vitamin A for Mother and Baby (ZVITAMBO) Trial, 2002-2007.[25,26] These samples had been used in an earlier evaluation of the BED-Capture Enzyme Immune Assay (BED-CEIA, simply BED).[27] This evaluation is particularly interesting because the ZVITAMBO samples are all from women who were not on ART and who provided blood and milk samples at baseline, within 96 hours of delivery, and at subsequent follow-up visits up to a period of two years postpartum. This study provides critical data on the performance

Stellenbosch University https://scholar.sun.ac.za

of the new avidity assays and contributes to the body of knowledge on how we can use these assays for HIV incidence surveillance purposes on samples collected in cross-sectional surveys.

Stellenbosch University https://scholar.sun.ac.za

3.0 Literature Review

In this section, we highlight the methods used in calculating HIV incidence over-time and we explore the merits and de-merits of each method. We present the ongoing efforts to track new HIV infections in response to the growing need to control HIV epidemics and sub-epidemics.

3.1 Modelling of HIV incidence trends using prevalence data

Modelling of HIV prevalence data has made important contributions to the estimation and prediction of HIV incidence trends. Two main modelling approaches have been used i.e., the static (or steady state) and dynamic models. The latter accounts for time-dependent changes in the state of the system, while a static model calculates the system in equilibrium, and thus is time-invariant.

Static models have used HIV prevalence among pregnant 15-24 year-old women, attending antenatal clinics, as a proxy for HIV incidence [28,4, 29] Over the years, countries have used the *Epidemic Projections Package* (EPP) and *Spectrum* models to create national estimates of HIV-1 incidence using time series HIV prevalence data from antenatal clinic surveillance, calibrated using population level trend data.[30,31] An important advantage of the EPP Spectrum suite of models is that they generate HIV incidence estimates for earlier years of the epidemic. The main disadvantage, however, is that changes in HIV incidence are detected using the retrospective prevalence data and therefore are limited in their utility for tracking epidemics. In addition, these models do not make use of age-specific HIV prevalence data and cannot thus provide age-disaggregated HIV incidence estimates that are critical in targeting interventions.

The use of static models for determining HIV incidence is further complicated by the growing influence of ART programs. The dynamics of increased survival of patients on ART, variations in individual patient clinical factors such as CD4 cell counts, viral loads, and the timing of ART initiation, all make it difficult to build mathematical models and to interpret these results.[32] Ideally, trends in HIV incidence should measure the impact of treatment and prevention programs in preventing new infection and halting the perpetuation of epidemics.[5] Using data from Zimbabwe, Hallett et al. (2009) developed a dynamic model that incorporated the natural changes in an HIV epidemic as it matures with ART to model HIV

Stellenbosch University https://scholar.sun.ac.za

transmission.[33] The model applied Bayesian Melding in which the model and observed trends in prevalence were compared in relation to sexual behaviour change.[33]

Recently, Mahiane et al. (2012) suggested that using individual level data versus aggregate data results in improved age- and time-specific incidence measures.[34] Building on the UNAIDS Modes of Transmission model (MOT), Borquez et al. (2016) proposed the Incidence Patterns Model for Sub-Saharan Africa.[35] Using Bayesian uncertainty incidence estimation, the model accounts for marital status, sexual activity and belonging to a key population, geographical distribution and observational incidence and population level prevalence data to provide HIV incidence estimation.[35] Although dynamic models are more robust than static models, their requirement for data elements and assumptions that may not be readily available, reduces their desirability for use in estimation of HIV incidence.

Other models have used multiple data sources including AIDS case reports, mortality and selected risk behaviours to calculate HIV incidence among young people. [10] A back-calculation of mortality data by Lopman et al. (2008) concluded that HIV incidence peaked in Zimbabwe in the period 1988 to 1990.[6] Using data from household surveys conducted in Sub-Saharan African countries, Hallett et al. (2009) and Rehle et al. (2010) have successfully demonstrated the utility of serial measures of HIV prevalence in determining HIV incidence rates.[7,36] The method is based on a calculation of change in prevalence among individuals age α in the first survey, and $\alpha + \tau$ in the second survey, carried out at time τ later.[7] The change in prevalence was attributed to incident infections and AIDS deaths. Although these models provide useful estimates of age specific distribution of new infections, they rely on the availability of accurate mortality data.

Stellenbosch University https://scholar.sun.ac.za

3.2 Measuring incidence using prospective cohort data in combination with modelling

The HIV Prevention Trials Network (HPTN) has conducted multi-country clinical trials that have been useful for gauging the efficacy of treating sexually transmitted diseases (STDs), use of microbicides and antiretroviral therapy in reducing HIV transmission. These clinical trials, HPTN034, HPTN035 and HPTN 052 have provided measures of HIV incidence in cohorts attending STD clinics.[37,38,39]

Very few cohort studies have been conducted to measure HIV incidence at population level, and these have been confined mostly to small geographic settings, thus making it difficult to generalise results to the rest of the population. Notable follow-up studies have been conducted in Uganda.[40] Studies have also been conducted to measure HIV incidence among commercial sex workers in Thailand.[41,42]

In Sub-Saharan Africa, Zimbabwe, has documented a few follow-up studies that provided measures of HIV incidence. Mbizvo et al. (1998) measured HIV incidence in a cohort of 2,833 male factory workers (age 17-61years at last birthday).[43,44] A follow-up data analysis compared HIV incidence obtained using prospective cohort methods and those obtained by modelling cross-sectional prevalence data.[45] The HIV incidence obtained from age-standardization was 2.02 (95% CI 1.57- 2.47) per 100 person-years compared to 1.98 - 2.74 per 100 person-years obtained using cross-sectional methods. The variation in HIV incidence followed a similar pattern to the age categorized HIV prevalence. This analysis showed that data on cumulative incidence and survival are important in determining HIV incidence trends.[45] A later factory worker study using cluster random sampling methodology, in which the experimental arm received vouchers to receive voluntary testing and counselling showed an HIV incidence of 1.21 per 100 person-years and there was no significant difference between the two arms.[46]

The Manicaland cohort study initially enrolled 1,627 HIV negative adult males and 2,465 HIV negative adult females recruited between 1998 and 2000.[47] Subsequent cohorts were enrolled every two years. Informed consent at each visit allowed the researchers to collect individual level behavioural and biomarker data. The study estimated that incidence of new

Stellenbosch University https://scholar.sun.ac.za

HIV infections was as high as 19.9 (95% CI 16.3-24.2) per 1000 person years in men and 15.7 (95% CI 13.0 -18.9) per 1000 person years for women in the study period.[47] Other notable cohorts that have provided HIV incidence data include the ZVITAMBO Project.[25] This study reported a cumulative HIV incidence of 3.4 per year (95% CI 3.0 -3.8) among 9,562 postpartum women who were HIV negative at enrolment in antenatal clinics in greater Harare, the capital city of Zimbabwe.[25]

While prospective cohort studies have the capacity to provide reliable estimates of HIV incidence, the cost of conducting such studies, loss to follow-up, selection biases and the difficulty of following a nationally representative sample, make them less attractive. A review of current methods in HIV incidence estimation by Brookmeyer (2010) noted that changing methods present challenges of reproducibility of results and this makes it difficult to compare HIV incidence trends.[3] He concluded that there is an urgent need for providing simpler and more reproducible methods for estimating HIV incidence trends such as application of laboratory assays to samples obtained in cross-sectional surveys.[3]

Stellenbosch University https://scholar.sun.ac.za

3.3 Laboratory based estimates of HIV incidence

Laboratory-based assays for estimating HIV incidence have evolved from dilution-based detuned assays that measure low HIV antibody titre, to complex and expensive synthetic-based synthetic oligopeptides and recombinant antigen assays that can be used to measure HIV antigens.[13] One of the early laboratory assays for estimation of HIV incidence measures the prevalence of the HIV protein, p24, in the absence of HIV antibody, as a marker for the transient status of recent infection.[16] The shortcoming of this p24-based assay was the short window (usually 1-2 weeks). This characteristic renders the assay unsuitable for application to samples collected in a cross-sectional survey, where few if any persons are likely to have been infected within the narrow time-frame. Other variants of this assay include a combined anti-p24 and immune-gamma globulin 3 (IgG3) assay that measures a narrow and temporary response to p24 in the subclass of IgG. Immuno-dominant enzyme (IDE) assays such as IDE-V3 (V3 refers to region) assay measures total response to selected gamma-protein (gp) 41 and gp120 epitopes that are commonly found in most antibody responses and so is therefore non-specific and of very low sensitivity. Critical developments by Janssen et al. (1998) provided a basis for moving from a detuned assay to one of the first serological incidence-testing algorithm.[48] This two-assay algorithm used a sensitive Enzyme-linked Immunosorbent Assay (EIA) diagnostic test to identify HIV-1 seropositive persons and a Less Sensitive (LS) EIA to distinguish recent from long-term infections.[48] These assays have been found to have a high misclassification rate among persons with long-term infections and those on ART, and have not been evaluated beyond subtype B infections.[48]

Further advances in laboratory-based assays include the use of avidity assays to determine HIV incidence.[15,49,50,51] Avidity refers to the accumulated strength of multiple affinities between the viral protein (antigens) and HIV specific antibodies. Avidity assays use the properties that:

(i) In response to exposure to HIV-1 virus, the immune system initially produces low avidity HIV-1 antibodies but, due to B-lymphocyte evolution and selection, the strength of the binding between the HIV antigens and human antibodies increases with time since infection;

(ii) An Avidity Index (AI) is measured as a ratio of the OD of a denaturing well compared to the OD of the control well (with wash buffer) expressed as a

Stellenbosch University https://scholar.sun.ac.za

> percentage in a two well avidity assay or as optical density readings that are normalised (ODn) by comparing readings to an external calibrator. This is done by adding a chaotropic (denaturing) agent to the antigen/antibody mixture that breaks hydrogen bonds: an optical density reading is then taken.[15]

In early infection, weak binding results in the level of antibodies in the treated sample being lower than that in the control, and the AI then takes values less than one. For more established infection, antibody levels in the two samples are similar and the AI approaches a value of one. The subtype of virus can lead to weaker binding (low-avidity); therefore using a multi-subtype avidity assay can improve performance. The cut-off values at which these measurements are performed are critical in the characterisation of a biomarker. [10,51]

A laboratory assay with the biochemical capacity to qualitatively detect the presence of the HIV antigen, even before antibodies are detectable in the blood, would provide an accurate indication of the recency of infection. Researcher have evaluated INNO-LIA HIV I/II Score Blot Assay (Fujirebio, Zwignaard, Belgium).[52] This assay uses the enzyme immuno-assay principle. The assay has five HIV-1 antigen bands: sgp120 and gp41 that detect HIV-1 specific antibodies; and p31, p24 and p17 that may cross react with HIV-2. The antigens gp36 and sgp105 are applied to detect antibodies to HIV-2. When a test sample is incubated with sequential addition of multi-antigen strip, goat antigen IgG, alkaline phosphate, enzyme substrate and reaction stopped by sulfuric acid this results in colorimetric identification of HIV specific antibodies. This line-based measurement of the reactivity of synthetic oligopeptides and recombinant antigens with HIV antigens in the blood makes the assay more sensitive. The assay currently has had limited use in surveillance, however, because of its high cost.

When applying laboratory assays to samples derived in cross-sectional surveys the calculation of HIV incidence requires three main inputs. Firstly, we require the number of samples that return an optical reading below a pre-set cut-off (C) such that we classify them as recently infected, while we classify those with a reading above C as long-term for a the HIV positive population size (N). Secondly, we need to estimate parameters specific for the assay. These are, the MDRI, which is the average time spent in a state of recency, while infected for less than some specified time T; and the FRR, which is the proportion of samples that continue to be misclassified as recent when the person has been infected for a time longer

Stellenbosch University https://scholar.sun.ac.za

than *T*. The MDRI and FRR are key indicators of the performance of a laboratory assay.[2] Evaluations comparing longitudinal data and laboratory based assay data have been critical in characterising laboratory assays.

The BED assay has been applied to cross-sectional and longitudinal studies. Studies have established the FRR and MDRI in different HIV populations. [27,53] This assay misclassifies individuals as recent infections even when they have clearly been HIV positive for some years,[27,54,55,56,57] and the rate at which this misclassification occurs varies with geographic location, age and duration between infection and seroconversion.[56,58]

In the ZVITAMBO study, the BED HIV incidence estimate, unadjusted for the FRR, was 7.6% per 100 person years. This was 2.2 times higher than the observed follow-up HIV incidence.[27] However, once account was taken of the proportion of cases that tested as recent infections, despite being infected for more than one year, the adjusted BED estimate of incidence matched closely the follow-up estimate of 3.4% (95% CI 3.0 - 3.8). This gave some hope that the BED, or a similar assay with a lower false-recent rate, might be used to provide acceptably accurate estimates of HIV incidence from cross-sectional HIV surveys.[27]

The synthetic peptide of the BED measures the increasing proportion (optical density) of HIV-IgG to total IgG after seroconversion. Highly active antiretroviral therapy (HAART) is reported to repair cell- mediated immunity. In the case of HIV infection, HAART suppresses replication of the virus and therefore this result in significantly reduced levels of HIV-IgG. When laboratory assays are used, these reduced HIV-IgG levels result in significantly reduced proportion of HIV IgG/Total IgG and misclassification of long-term infections as recent infections. It is therefore critical to rule out ART exposure for all individuals testing as recent infections.[56,59] To obtain reasonable estimates of HIV incidence, McDougal et al. (2009) proposed an estimator for adjustment under steady state assumptions to account for the long-term specificity.[60] Hargrove et al. (2008) suggested a new estimator that depended only on the mean window period and false recent rate[2] and McWalter and Welte (2009) proposed a similar estimator, based on a full mathematical analysis that accounts for a dynamic epidemic and provides a weighted incidence which can be applied to all assays.[61] More recently two independent theoretical approaches have suggested that, where HIV incidence is estimated using biomarker methods, the incidence should be estimated over a finite time *T*.[61,63] In these formulations the MDRI must be estimated among cases that have

Stellenbosch University https://scholar.sun.ac.za

been HIV positive for at most time *T*, not over their whole lifetime.[63] The two key parameters are modified to MDRI (Ω_T), the average time spent in the recent state while infected for less than Time (*T*) and the FRR (ε_T) as the proportion of cases testing as recent infections among those known to have been HIV positive for more than time *T*.[2]

Using longitudinal cohort samples, the MDRI and FRR of BRAI avidity assay have been compared to those of BED assay.[64,65] The BRAI assay has been evaluated using the HIV-NET- United States of America (USA) cohort (89 subjects provided 349 observation samples), VAX003 (105 subject, 95 observations) Thailand, VAX04- USA 962 subjects, 274 observations), Reach- Nigeria (14 subjects, 131 observations), SIPP- USA (11 subjects, 95 observations) cohorts.[65,66] These cohorts included different HIV-1 subtypes but B was the most prevalent. Recent results of the BRAI evaluation show that the assay has a lower FRR (<1%) compared to >5% for BED. Recent data from Botswana confirm improvements in the BRAI assay compared to BED.[67]

The LAg has also been evaluated against BED.[49,50,68] While the manufactures of LAg recommend excluding people who are on ART, elite controllers and those with CD4 cell count less than 200 cells/µl, Longosz et al.. (2014) argued that this does not completely eliminate misclassification.[68] They stated that the misclassification of cases by LAg was mainly due to viral load (VL) suppression because of antiretroviral therapy and lower CD4 cell count. While other workers have observed these factors, they highlighted that it was important to determine the appropriate laboratory assay for the population of interest. Of concern in this evaluation was the definition of elite controllers who were classified on the basis of VL count less than 400 copies; despite the current threshold for classification as an elite controller, which is set at < 50 copies/ml. Notwithstanding this anomaly, other large scale studies have shown improvements in the performance of LAg assay[50]

The manufacturers of LAg avidity assay have recently set a predetermined MDRI value for the calibrator at 141 days (95% CI 119-150) at a cut-off of 1.0. An analysis by Duong et al. (2015) now recommends a cut off of 1.5 which has a corresponding MDRI of 130 days (118-142) for all subtypes and 152 days for subtype C.[69] The differences in MDRI calculated using seven different methods were minimal at each cut-off. Based on this analysis, an optimal cut-off was determined based on the trade-off between a high MDRI and a low FRR.[4] Regardless of the method used in calculating the MDRI, the main focus is on the

Stellenbosch University https://scholar.sun.ac.za

development of a laboratory assay whose test properties of MDRI and FRR will allow calculation of precise HIV incidence estimates.[69]

Stellenbosch University https://scholar.sun.ac.za

3.4 Utility of multi-assay algorithms in improving measurements from single laboratory assays

In attempting to improve the performance of laboratory assays, researchers have used multi-assay algorithms. Brookmeyer et al. (2013) and Laeyendecker et al. (2013) used BED to measure incidence in Clade B samples in the United States of America, in an MAA which included HIV-VL and CD4 cell count and concluded that MAAs can be used to provide more accurate estimates of HIV incidence.[70,71]

An earlier evaluation by Konikoff et al. (2013) assessed the performance of LAg as a single assay and compared it with the performance of MAAs.[50] In this evaluation, the MDRI was 119 days for a three assay MAA (CD4, LAg and BRAI) and 146 days for a four assay MAA (VL, CD4, LAg and BRAI). They concluded that, given the costs of four-assay algorithm, an optimised two assay MAA (LAg and BRAI) was reliable in providing HIV incidence consistent with follow-up and therefore could reliably be used for estimating HIV incidence.[50] Recent work by Serhir et al. (2016) proposed that using BRAI first, then LAg in serial, was a more sensitive algorithm in screening out false recent cases and estimation of incidence.[72]

Very few studies have compared three laboratory assays using the same specimens. Using well-characterised Sub-type B, German seroconverter cohort panel, Hauser et al. (2014), compared the FRR of BED, BRAI and LAg.[64] The researchers found that the two avidity assays had a lower FRR (2%) than the BED (7%) among long term ART naïve patients. Among 14 patients on ART and 5 Slow Progressors, LAg misclassified 1/14 and 0/5 while BRAI misclassified 2/14 and 1/5 respectively. For recently infected individuals, BRAI correctly classified 88% compared to 48% by LAg.[64] Based on these results, the researchers could potentially infer that there was a great improvement in the assays and the effects of ART were minimal: however, the small samples size does not warrant making such conclusions.

Another multi-assay evaluation by Kassanjee et al. (2014) reported that, for all specimens in the evaluation, the FRR for LAg was 1.3% compared to 6.2% for BRAI.[73] FRR was higher among patients infected in the past 2-3 years (2.5% LAg and 12.5% BRAI), and even higher, for both assays, among those on treatment (58.8% LAg and 50.0% BRAI). The researchers

Stellenbosch University https://scholar.sun.ac.za

concluded that the large proportion of false recent results in ARV-treated individuals affects assays' performances and therefore assays required further optimisation. [73]

Following this large-scale evaluation, Kassanjee et al. (2013) concluded that, despite an individual assay's shortcoming, the merits of an assay lie more in its ability to provide a precise measurement of HIV incidence.[20] Based on these evaluations, there is a need to find an optimal trade-off between a sufficiently large MDRI and a sufficiently small FRR in order to achieve precise measurements of HIV incidence.[20]

3.5 Measuring HIV Incidence using multiple methods

Early in the race to find a suitable method for incidence estimation, Brookmeyer (2010) reported that the greatest challenge in estimation of HIV incidence is that different methods and analysis often produce different estimates for the same time points.[3] Currently, no single method on its own provides a universally accepted measure of HIV incidence and advances in laboratory assays should not therefore preclude the more conservative approach of data triangulation. There are growing calls to use multiple methods and data sources to gain more understanding of the complex dynamics of HIV transmission.[3] Rutherford et al. (2010) suggested using secondary data from multiple sources, for purposes of interpreting data sets that cannot be included in meta-analysis, as a useful method for assessing the impact of interventions.[9] Kim et al. (2011) concluded that triangulation of methods is a useful way of determining trends in HIV incidence estimates.[74] They recommended further systematic evaluation of new and existing laboratory assays to determine the reliability of national HIV incidence trends.[74]

While the suggestions to triangulate methods in order to understand HIV epidemics are certainly valid, there is a need to have reliable methods that are both efficient and reproducible for use in routine surveillance. Laboratory assays may provide this solution, if they have characteristics that lead to increased accuracy (smaller bias) in measurement of HIV incidence. Our evaluation of the LAg and BRAI assays is critical in providing evidence in this regard.

4.0 Description of candidate laboratory assays

4.1 BED-Capture Enzyme Immune Assay (Sedia HIV-1 BED Incidence EIA, Cat. No. 1000)

Researchers in the United States Centres for Disease Control and Prevention (CDC) developed the BED one of the earliest incident laboratory assays.[75] This assay uses a synthetic antigen containing sequences from multiple subtypes to measure the proportion of anti HIV-1 immuno gamma globulin (IgG) present in total IgG following seroconversion. Persons are classified as 'recent' seroconverters if their blood samples test positive by a standard HIV-1 ELISA and have an ODn below a pre-set cut-off on the BED assay. This assay has been extensively evaluated against observational data, mathematical models and recently against avidity based laboratory assays. It was shown to provide unreliable estimates of HIV incidence due to its high FRR a factor which is sensitive to geographic variance in population and reaction to disease stage.[17, 76]

4.2 Sedia™ HIV-1 Limiting Antigen Avidity Enzyme Immuno Assay (LAg-Avidity EIA)

The LAg-Avidity EIA [22] is an in vitro quantitative limiting antigen assay used to determine recent and long-term HIV-1 infection status.[27,49,50] The assay uses a 96 well plate coated with multi-subtype gp41 recombinant protein (rIDR-M). The assay plate is run with four controls, Negative Control (NC), Low Positive Control (LPC), High Positive Control (HPC) and Calibrator in the first four wells. The optical density readings values are optimised by dividing the specimen OD by the Calibrator OD. This way all 92 samples run on the same plate utilise the same calibrator reference point, thereby minimising plate-to-plate variability. All specimens with ODn>1.0 are classified as long-term infections and no further tests are required. In order to rule out misclassification of samples as recent due to viral suppression in patients on ART and in cases of elite controllers, the manufactures recommend that all samples with an ODn ≤ 1.0, be tested for HIV-1 viral load. Specimens with a LAg assay ODn ≤ 1.0 and a VL ≥ 1000 copies/ml are classified as recent HIV infection. Samples with an undetectable VL (< 1000) and ODn ≤0.400 require HIV-1 serology retest in order to rule out false HIV+ serology.

Stellenbosch University https://scholar.sun.ac.za

4.3 Modified BioRad Genetic Systems HIV-1/HIV-2 plus O avidity-based assay (BRAI)

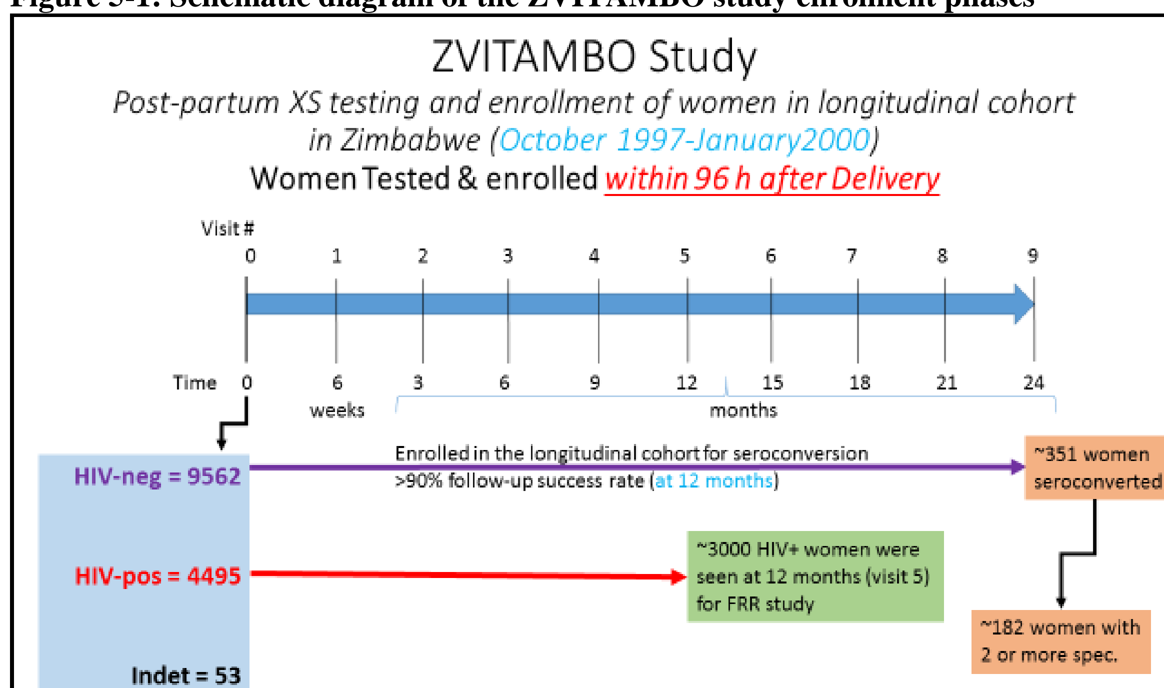
The BRAI avidity assay was developed by modifying the BioRad Genetic System HIV-1/HIV-2 Plus O (BRAI) protocol, by Centers for Disease Control and Prevention, Atlanta, GA. USA.[24] The immunoassay is an IgG/IgM (3rd generation) enzyme immune-assay that uses recombinant proteins and synthetic peptides to detect antibodies to HIV-1/HIV-2.[77] The modified assay is based on the avidity principle and can be run on plasma, serum and DBS eluate. The assay was modified to include a sample dilution of 1 in 10 using cold washing buffer and a first incubation at 4^{0}C. During the second incubation, one well is treated with 0.1M Diethylamine (DEA), a chaotropic agent, and the other (reference) well is treated with wash buffer. The first incubation at low temperature and the use of a chaotropic agent allows the differentiation between low- and high-affinity HIV-1 antibodies. An avidity index (AI) is calculated for each sample by dividing the OD of the well containing DEA by the OD of the reference well and multiplied by 100 only if the OD in presence of wash buffer for the samples is equal or higher than the run cut-off (mean of Negative OD + 0.250). If the OD in presence of wash buffer is below the run cut-off the sample is invalid. Specimens with an AI value in the range of 20-50% should be repeated in duplicate and the final interpretation is determined by the mean of the duplicate results. Specimens with an AI below or equal to a predetermined cut-off, e.g. less than 30%, are classified as recent infections.[24]

5.0 Methods

5.1 Description of ZVITAMBO cohort samples and historical tests conducted

The ZVITAMBO Trial was a randomized, controlled clinical trial that measured incidence of HIV in mothers administered Vitamin A versus those given placebo. The study enrolled, within 96-ours post-delivery, 14,110 mother-infant pairs recruited from maternity clinics and hospitals in Greater Harare. The mothers provided written informed consent, and were recruited in the period November 1997 to January 2000.(Figure 5-1: Schematic diagram of the ZVITAMBO study enrolment phases).[25] Mother-baby pairs were followed-up at 6 weeks, 3 months and 3-monthly thereafter for at least 1 year, and diminishing subsets at 3-mothly intervals for up to 2 years.[25] Patients provided samples of blood and breast milk at enrolment (baseline) that were tested for HIV-1, viral load and CD4 cell count. Blood and breast milk samples were taken at each follow-up visit.

Figure 5-1: Schematic diagram of the ZVITAMBO study enrolment phases



Stellenbosch University https://scholar.sun.ac.za

5.2 HIV tests and VL tests conducted on ZVITAMBO samples

At delivery, mothers were tested for HIV antibodies using two ELISA tests run in parallel: *Genescreen Diagnostics Pasteur, Johannesburg, South Africa (Genescreen HIV1/2*) and *Murex HIV 1.0.2 ICE (Murex HIV 1/2), Murex Diagnostics, Eden Vale, South Africa.* Discordant results were confirmed using *Western Blot (HIV 2.2, Genelabs Diagnostics SA, and Geneva Switzerland).* From the different blood draws, 9562 mothers tested HIV negative result at baseline, while 4495 mothers were HIV positive at baseline and subsequent blood draws. The ZVITAMBO study excluded fifty-three (53) mothers with baseline HIV indeterminate results. From the 9562 HIV negative at baseline, 353 women seroconverted during follow-up and subsequent samples were taken at each follow-up visit.

Plasma samples were also tested for HIV viral load by quantitative HIV RNA testing (Roche *Amplicor* HIV-1 Monitor test) which had an ultra-sensitive detection limit of less than 400 copies/ml.[26,76]

5.3 Characterisation of BED using ZVITAMBO samples

Of the women tested at baseline, 4495 women tested HIV positive and the samples they provided were archived: 3010 of these women were subsequently seen at 12 months postpartum and 2749 of them provided a sample that was later tested using BED, in order to estimate the False Recent Rate (FRR). (Figure 5-1:Schematic diagram of the ZVITAMBO study enrolment phases).[27]

At baseline (Figure 5-1), 9562 women tested HIV negative and 353 of these women were initially thought to have seroconverted (based on HIV serology tests performed at clinic visit) during follow-up. When we retested these samples, two women were found never to have seroconverted, thus leaving 351 women. We tested all samples from seroconverting women using BED and used the patterns of increase in optical density, with estimated time since seroconversion, in order to determine MDRI at various cut-offs[2]. Of these, 234 provided at least two (2) HIV positive samples and only 186 women were seen at visit 5.

Stellenbosch University https://scholar.sun.ac.za

5.4 Ethical considerations for ZVITAMBO samples

All specimens were bar-coded and no personally identifying information was linked to the specimen, thereby assuring the anonymity of the participant. Participants in the ZVITAMBO cohort study gave permission to store samples and to use them for future additional tests.

The original ZVITAMBO cohort study received ethical approval from Johns Hopkins University and the Medical Research Council of Zimbabwe (MRCZ). The current evaluation of laboratory incidence assays received approval from MRCZ, Research Council of Zimbabwe, Centers for Disease Control and Prevention, Atlanta, and Stellenbosch University.

Stellenbosch University https://scholar.sun.ac.za

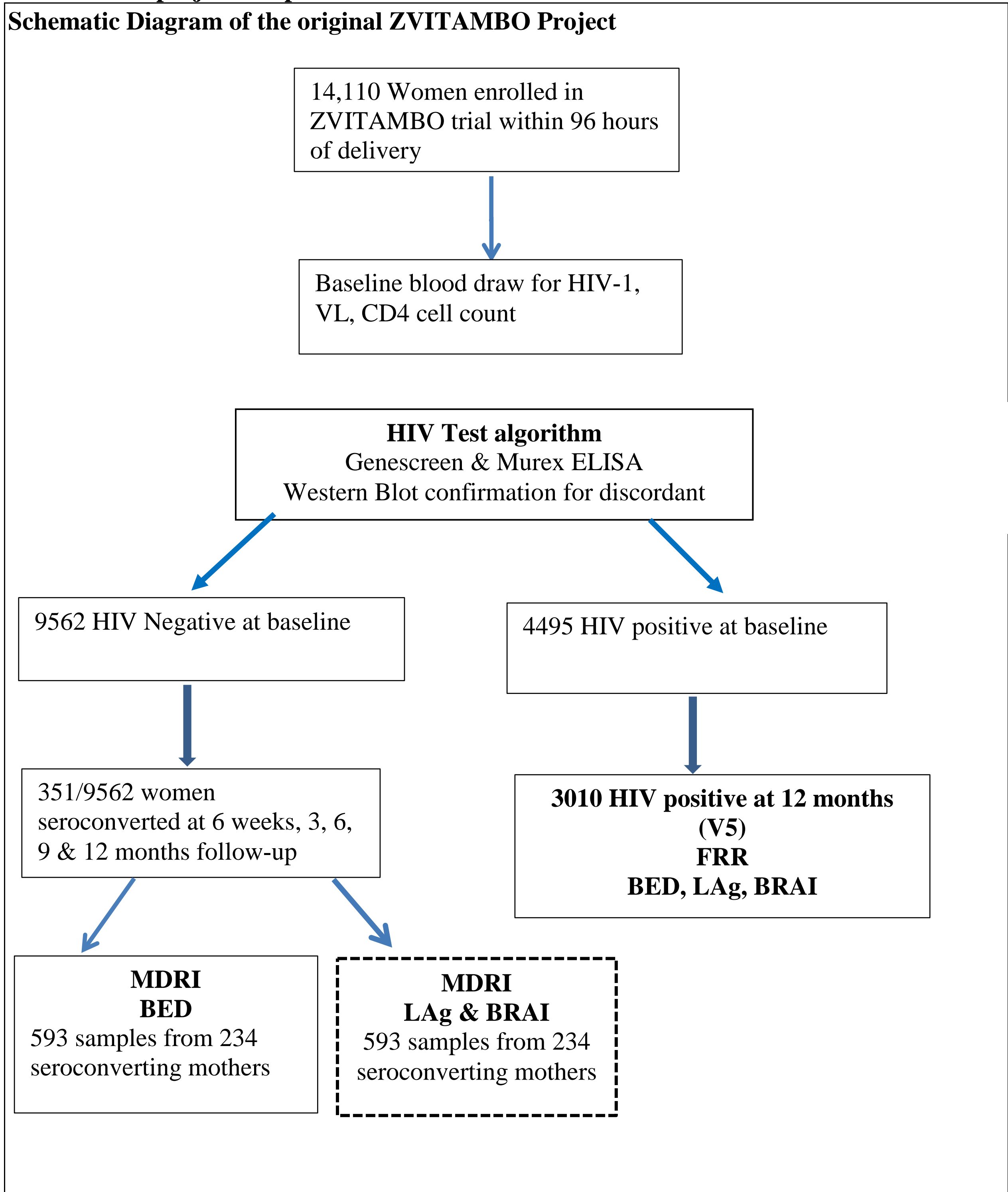
6.0 Methods in the Evaluation of LAg-Avidity EIA and BRAI Avidity Assay

6.1 Sample viability test

As a precursor to this study we conducted preliminary work to establish the viability of ZVITAMBO samples. A total of 224 randomly selected ZVITAMBO samples were retested using BED and this step showed that the samples, which had been collected over a decade previously, were still viable and could be used to evaluate the performance of the LAg and BRAI assays for FRR, MDRI and estimation of HIV incidence. We were confident in proceeding to tests the available samples with LAg and BRAI assays. We present a schematic diagram of the current ZVITAMBO enrolment, BED, LAg and BRAI evaluations below (Figure 6-1).

Stellenbosch University https://scholar.sun.ac.za

Figure 6-1: Schematic diagram of the BED, LAg and BRAI evaluation using ZVITAMBO project samples



Stellenbosch University https://scholar.sun.ac.za

6.2 Laboratory testing using the Sedia™ HIV-1 LAg-Avidity EIA (Catalog No. 1002)

We used the single well protocol as described previously by Wei et al. (2010) and as directed by the manuafacturers.[15,22] A plate map was prepared for the 96 well plate, to include commercially supplied negative controls (in duplicate), calibrators (triplicate), low positive controls (triplicate) and high positive controls (triplicate) followed by samples in the remaining88 wells single wells for each plate. We increased the number of controls and calibrator (See section 4.2 Manufacture recommendation) in order to improve quality assurance for each plate. We diluted samples 1:10 with the propylene sample diluent and transferred 100µl of the diluted sample into the appropriate well (as per plate map) of the avidity plate. We washed each micro well four times using wash buffer, then added 200µl dissociation buffer at pH 3.0 for 15 minutes at 37^{0}C. We repeated the washing procedure. We added 100 µL of diluted and freshly prepared (1:1001) Goat Anti-Human IgG-HRP Conjugate to each micro-well, sealed microplate and incubated for 30 minutes at 37^{0}C followed by the wash procedure. We then added 100µl of Tetramethylbenzidine (TMB) solution, incubated for 15minutes at 25^{0}C, and then stopped the reaction using the stop solution.

We obtained optical density (OD) readings using a spectrophotometer at 450nm wavelength and reference filter of 620-650 nm. We entered all OD readings on the CDC supplied spreadsheet. We calculated the average for the negative control, the median optical density of the low and high positive and the median value for the calibrator and compared the range of values with the manufacture's values in order to accept (valid) or reject (invalid) the plate values. We calculated the normalized optical density for each control, calibrator and specimen by dividing the OD value by the median OD of the Calibrator. The process of OD normalization by an internal calibrator decreases run-to-run variability and increases reproducibility.[22]

We accepted all specimens with ODn > 1.0 as long-term infection and conducted no further tests. We retested samples with ODn ≤ 1.0 using LAg in duplicate We retested samples that returned an ODn <0.4 for HIV-1 and 2 serology status using *Alere Determine ™ HIV-1/2 Ag/Ab* (Determine) then *Biolytical Laboratories, INSTI™ HIV-1.2*(Insti) antibody test in serial and confirmed using confirmed using *Calypte HIV-1 Western Blot* (WB).

Stellenbosch University https://scholar.sun.ac.za

6.2.1 Use of viral load testing with LAg assay

The presence of HIV-1/2 ribonucleic acid (RNA) confirms the presence of virus and in this regard can be taken as a confirmation of HIV serological status. Viral load is useful in resolving serologic status and simultaneous differentiation of recent and long term infections. Early infection is associated with rapid initial increase in VL count, followed by a decline as a result of virus selective pressure, the intervention of the host immune system and natural growth and death of virus.

We retrieved all available results for viral load analyses from the ZVITAMBO database. Where there were missing data we attempted to retrieve samples but, in all cases, the samples were depleted and could not therefore be retested in this current evaluation (See results in Section 7.2).

6.2.2 Retesting HIV serology for samples with LAg ODn < 0.400

In accordance with the manufacturer protocol, we retested for HIV-1 antibodies those samples that returned a LAg ODn below 0.400. This was based on the premise that the sample could mistakenly have been diagnosed as HIV positive, when it was in fact sero-negative. We therefore carried out the further serological tests as a confirmation of the HIV status. We used rapid HIV test kits, *Determine*™ *HIV-1/2 Ag/Ab,* and *Insti* in serial, confirmed using Western Blot. The full analysis of results are presented in Section 7.2.

6.3 Laboratory testing using BioRad Genetic Systems HIV-1/HIV-2 plus O EIA (BRAI Avidity Assay)

We prepared a plate map according to the BRAI avidity assay protocol (*Genetic Systems HIV-1/HIV-2 plus O EIA) (Centers for Disease Control and Prevention, BRAI Laboratories, Atlanta GA, USA*). The protocol map reserves the first two strips for controls, 3 Negative controls (NC) from the kit for the run's cut-off calculations, Incident control (IC), HIV-positive kit controls (HIV-1, HIV-2 and HIV-O) and Positive Control (PC) provided by CDC. We prepared reagents as per protocol instructions.[24]

We diluted each specimen and the CDC Incident and Prevalent 1:10 with cold specimen diluent and loaded them in two wells following the plate map. We also loaded the negative controls in 3:4 dilution, sealed the plate and incubated at 4^0C for 60 minutes. We washed the

Stellenbosch University https://scholar.sun.ac.za

plate six times then added the second set of controls from the kit, NC, HIV-1, HIV-2 and HIV-O controls in single well. For each specimen, NC, IC and PC, we added 100 µl of wash buffer to the first well, wash buffer (WB) wells. We then added 100 µl of the 0.1 M DEA to the second well (DEA wells) for each sample, IC and PC. We sealed the plate and incubated at 37^{0}C for 30 minutes. We repeated the wash procedure, sealed and incubated for 40 minutes at 37^{0}C and added 100 µl of Working Conjugate Solution to all wells. We added 100 µl of the Working TMB Solution to all wells then incubated in the dark for 30 minutes at temperature 24^{0}C - 25^{0}C. We stopped the reaction using stopping solution and read the absorbance in a plate reader at wavelength of 450 nm and a filter 630 nm as reference. We recorded the OD for the WB and DEA wells for each sample on the work sheet (CDC Atlanta provided) and transferred the results to the database. We accepted a run as valid following the evaluation criteria in the protocol (run's cut-off, IC and PC, kit's HIV-positive controls within the recommended values). We calculated an avidity index (AI) when the OD in the presence of wash buffer was higher than or equal to the run's cut-off (mean of NC+0.250); otherwise, the sample was termed "invalid". We calculated the AI by dividing the OD of the DEA well by the OD of the wash buffer well and multiplying by 100. We retested, in duplicate, samples with AI in the grey zone (20-50%) as previously described and the result was the mean of the duplicate values. Samples with AI values >30% were classified as prevalent (long-term) infection and ≤ 30% were incident infections.

6.3.1 BRAI Assay Invalid Results

Specimens whose wash buffer ODn fell below the negative cut-off value of the assay returned an "invalid" result because the basis of the BRAI assay is a comparison of the antibody binding difference between the wash buffer and the DEA wells. A sample which is HIV serology negative, or is a very early infection when antibody titers are very low, may return an "invalid" result on the BRAI assay because there is no antibody binding with which to compare. We retested all invalid samples in duplicate using BRAI assay. For samples which remained invalid, we retested the HIV serology using rapid HIV test kits, *Determine*™ *HIV-1/2 Ag/Ab,* and *Insti* in serial, confirmed using Western Blot. We also retrieved the VL result from the ZVITAMBO database and used them for further determination of HIV status. The full analysis of these results are presented in Section 7.2.

Stellenbosch University https://scholar.sun.ac.za

6.4 Quality Control of Laboratory Assays

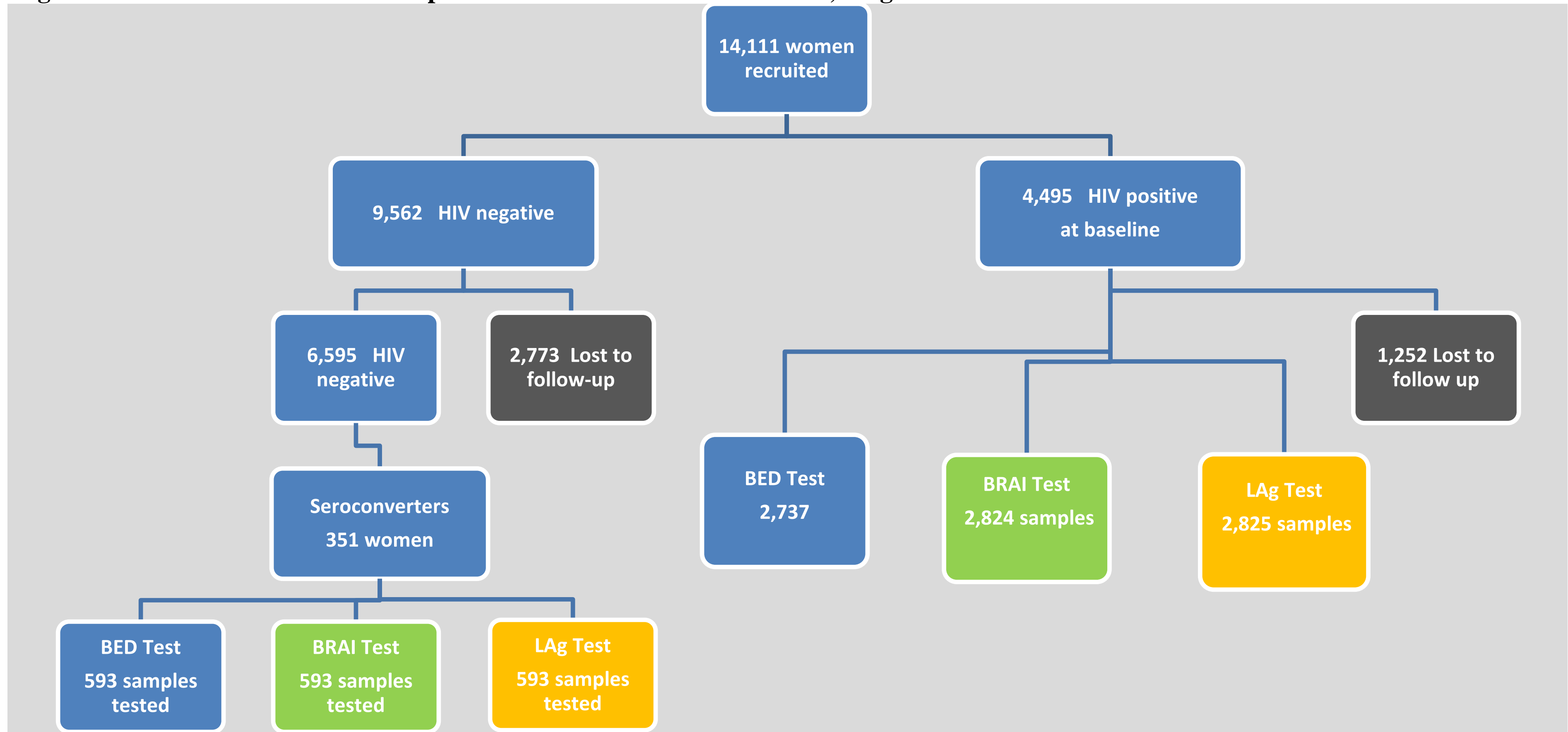
We conducted internal quality assurance of laboratory assays by including manufacture-supplied controls as well as CDC supplied QC samples with a predetermined optical density. We considered all plates with controls outside of pre-set values as invalid runs and repeated the tests.

6.5 Sample selection for evaluation

We have provide a schematic diagram of the actual number of samples tested using each assay in this evaluation. (Figure 6-2)

Stellenbosch University https://scholar.sun.ac.za

Figure 6-2: Actual ZVITAMBO samples tested in the evaluation of BED, LAg and BRAI



Stellenbosch University https://scholar.sun.ac.za

6.6 Statistical analysis

6.6.1 Samples in MDRI analyses

Out of 9565 HIV-1 negative women, 353 women were initially classified as having seroconverted during follow-up (Figure 6-2). Of these 353 women, we reclassified two (2) as always being HIV negative, leaving 351 women who seroconverted and accounting for 593 HIV positive samples that contributed to MDRI calculation. Out of these 351 seroconverting women, 234 provided more than one HIV positive sample. This allowed us to conduct analysis of MDRI when the number of samples (*ns*), collected in subsequent visits after the HIV sero-status changed from HIV negative to HIV positive (posterior measurements) was varied between minima of *ns* =1 and *ns*=4. Our evaluation is one of the few studies that has been able to explore the impact of varying the number of samples included in MDRI calculation.

6.6.2 Statistical analysis of MDRI

We analysed the pattern of increase in ODn, or AI, with time since seroconversion, in the series of positive samples collected from seroconverting cases. Since the shortest period between visits was 6 weeks, and was otherwise at least 3 months; and could be longer than this if a mother missed one or more visits; we never knew exactly when a case had seroconverted. We only knew that on a particular visit a case tested HIV negative for the last time and, on subsequent visits, always tested HIV positive. By definition, seroconversion occurred at some time in the interval (t_0) between the time of the last HIV negative and first HIV positive tests. Clearly, the longer the time lag period t_0, the less sure we can be about the timing of seroconversion. Accordingly, in our analyses, we investigated the effect of varying the maximum allowable value of t_0 for a case to be included in an analysis.

We defined the MDRI for a biomarker as the time that the biomarker level stays below the value of a pre-set cut-off *C*, during some pre-defined time *T*. The cut-off is a pre-set, user-specified, optical density reading or avidity index. We classify samples returning values below *C* as recent (assay positive) infections: samples with higher values are classified as long-term infection (assay negative).

Stellenbosch University https://scholar.sun.ac.za

We explored these relationships (varying *C*, *ns* and t_0) using several statistical methods. We wanted to be sure that whatever differences we observed in calculating MDRI for LAg and BRAI would not be attributable to the type of statistical analysis, and accordingly compared the estimates derived using a number of different methods. We used non-linear mixed modelling (NLMM) for BED and LAg assays, and variations of models by Sweeting et al. (2010) and Hargrove et al. (2012).[21,63,77]

Since avidity index does not increase according to a parametric form, we used survival analysis and binomial regression analysis to calculate MDRI for BRAI.[73] In addition to these methods, we also explored the use of a ratio method in which we determined the proportion of all seroconverters testing recent according to defined cut-off among seroconverters (r/s). We describe the methods in more detail in Section 8.

The ZVITAMBO trial measured HIV incidence, by following-up and retesting women and babies who were HIV negative at baseline, and compared the incidence among mothers and babies administered Vitamin A, compared to those who received a placebo. Accordingly, we also explored the differences in MDRI by study arm. Furthermore, we compared MDRI for women who seroconverted during the first nine months (Visits 1 to 4) to those who seroconverted at least 12 months postpartum (Visits 5 to 8).

Stellenbosch University https://scholar.sun.ac.za

6.6.3 Samples in FRR analyses

A total 4495 women tested HIV positive at baseline in the ZVITAMBO trial (Figure 6-2). Of these, 3010 women were seen at Visit 5 (V5), 12 months postpartum. These women were known to have been HIV positive for at least one year (365 days), based on the original ZVITAMBO parallel rapid HIV testing using *Genescreen HIV-1/2* and *Murex HIV-1/2,* with confirmation using *Western Blot.* We retrieved and tested 2825 of these samples using LAg avidity assay and 2824 using BRAI avidity assay in the evaluation of the FRR for the two-biomarker systems.

We focused on samples collected at baseline (Visit 0 or V0) and 12 months postpartum (Visit 5 or V5) for three reasons: (i) All women in the study were tested at baseline. At Visit 5, the ZVITAMBO team made every effort to locate, and test, every person in the study – particularly because financial constraints meant that it was necessary to drop people from the study, in increasing proportions, after V5. Visit 5 thus produced the largest proportion of cases followed up out of all visits after baseline. (ii) Financial constraints also meant that it was not possible to test all samples from Visits 1, 2, 3 and 4 using BED, LAg and BRAI. (iii) Similar considerations meant that not all samples collected after baseline were tested for HIV: once a case had tested HIV positive on two different visits it was deemed unnecessary to test further samples for HIV.

Infection occurred at different time points during the 24-month follow-up period and, accordingly, any case where a mother had ever previously tested HIV negative at any previous visit was tested for HIV and, where found to be HIV positive, tested also with BED, BRAI and LAg. In this way, we used all eligible samples from HIV positive results from seroconverters from all visits in the estimation of MDRI.

Stellenbosch University https://scholar.sun.ac.za

6.6.4 Statistical analysis of FRR

We estimated the FRR for each marker as the proportion of HIV positive samples that return a recent result out of the total number of people known to be HIV positive for more than 1 year. We calculated FRR at different cut-offs for each assay and plotted graphs against varying C. In order to have comparable estimates, we also explored the relationship between the MDRI and FRR for different assay systems.

Stellenbosch University https://scholar.sun.ac.za

6.6.5 Incidence calculations

We used the formulae proposed by Hargrove et al. (2012b) and Kassanjee et al. (2012) to estimate incidence, taking into account the MDRI and FRR parameters derived in the evaluation. We calculated the HIV incidence from the cross-sectional survey data obtained at baseline, and at 12 months postpartum, in the ZVITAMBO study[63,62]

The annual risk of infection is the probability of becoming infected within a period of one year. Formally the adjusted annual risk of HIV infection (J_T) is calculated from:

$$\hat{J}_T = \frac{R - \varepsilon P}{R + \varpi_T N - \varepsilon(N + P)} \quad (1)$$

Alternatively, the adjusted instantaneous incidence rate (I_T) is given by:

$$\hat{I}_T = \frac{R - \varepsilon P}{(\Omega_T - \varepsilon T)N} \quad (2)$$

where R is the number of recent cases among P testing HIV positive, N is the number testing HIV negative and T is the time (one year in our case) over which the MDRI (Ω_T) and the FRR (ε) are defined. The unadjusted values (j_T and i_T) are found by setting $\varepsilon = 0$ in Equations (1) and (2) to give:

$$\hat{j}_T = \frac{R}{R + \Omega_T N} \quad (3)$$

$$\hat{i}_T = \frac{R}{\Omega_T N} \quad (4)$$

This weighted mean incidence is ideally suited to calculate incidence for any biomarker without any interference from biomarker dynamics, epidemiological and demographic history. The South African Centre for Epidemiological Modelling and Analysis (SACEMA) online resource for calculation of incidence is found at; http://www.incidence-estimation.org/page/spreadsheet-tools-for-biomarker-incidence-surveys

6.6.6 Follow-up incidence

The follow-up HIV incidence in the original ZVITAMBO cohort study (1998 -2000) was based on Turnbull Survival analysis at 24 months with 95% CI calculated by the bootstrap method and censoring women at 24 months or time of last negative test.[25, 27] These follow-up estimates represent a cumulative probability of HIV infection over a specified period of time usually a year or greater as in the case of calculation by Hargrove et al. 2008.[27] Kassanjee et al. (2012) proposed an equation for calculating a uniformly weighted average HIV incidence that refers to a period of time (T) preceding the cross-sectional survey.[62] This formula accounts for the fact that there is continuous population dynamic such that incidence at time (t) contributes to the I_T with a weight proportional to (i), in the susceptible population.[62] This method, allows for the aggregation and averaging of substantial information arising from a large susceptible population. Unlike probability or proportion, the hazard rate of infection can be greater than 1.[62]

In our evaluation, we based our HIV incidence estimates, using biomarker testing, on the adjusted instantaneous incidence rate method (Equation 2). Follow-up and biomarker-based incidence estimates each have their own problems. The biases inherent in follow-up study include the higher loss to follow-up for individuals that are at a higher risk of infection compared to those who remain under observation. On the other hand, biomarker-based incidence estimates are affected by the test properties of the biomarker, the MDRI (Ω_T) and FRR (ε) of the biomarker assay.

Our comparison of the follow-up incidence to the biomarker-based estimates does not provide a validation of one approach or the other. This is obviously true for our study, where we used the same data to estimate both the follow-up and biomarker estimates of incidence. Even if we make the estimates using independent data sets, however, there are different biases and weightings involved in the different approaches. If, however, we find significant differences between the estimates calculated using the two approaches this can draw attention to problems with at least one of the estimates.

Stellenbosch University https://scholar.sun.ac.za

6.6.7 Coefficient of Variation and Confidence Intervals

The *coefficient of variation* (CV), also referred to as the *relative standard error*, is a statistical measure of the dispersion (variability) of data points in a data series around the mean. We calculate the CV by dividing the standard error by the mean and the statistic is useful for comparing the degree of variation from one data series to another, even if the means are drastically different from one another. In the case of our evaluation, the CV provides an indication of the precision of the MDRI, FRR or incidence estimates for each of the assays. The emphasis on the CV in this work (particularly for MDRI estimates) stems from the fact that the MDRI point estimates may in general be similar regardless of the methods used, therefore CV is useful in highlighting these differences.

Confidence intervals for means are intervals constructed using the delta method approximation, such that for a specified proportion of the time, typically either 95% or 99% of the time, they contain the population mean. We did not make any formal comparisons between credible and confidence intervals. In this thesis, we were interested only in situations where either:

(i) There are major differences between estimates, such that there is no overlap of the 95% confidence/credible intervals, or

(ii) The differences between the point estimates are small relative to the size of the confidence/credible intervals to the extent that the intervals concerned overlap the point estimate with which comparison is being made.

Although we did not conduct the formal hypothesis testing, the visual representations of our comparisons were useful in highlighting differences in the MDRI and incidence. We acknowledge, however, that a 95% CI and point estimate can overlap and yet the difference would be significant in a hypothesis test with a significance threshold of 5% if the estimates were highly positively correlated. Many of the estimates in this thesis are not independent, and probably highly correlated.

The Bayesian concept of a credible interval is sometimes put forward as a more practical concept than the confidence interval. For a 95% credible interval, the value of interest (e.g. size of treatment effect) lies with a 95% probability in the interval.

Stellenbosch University https://scholar.sun.ac.za

In this evaluation, we need the CV of the FRR order to calculate the CI around the incidence estimate, using Equation e7 in the Appendix to Kassanjee et al (2012) [62].

Stellenbosch University https://scholar.sun.ac.za

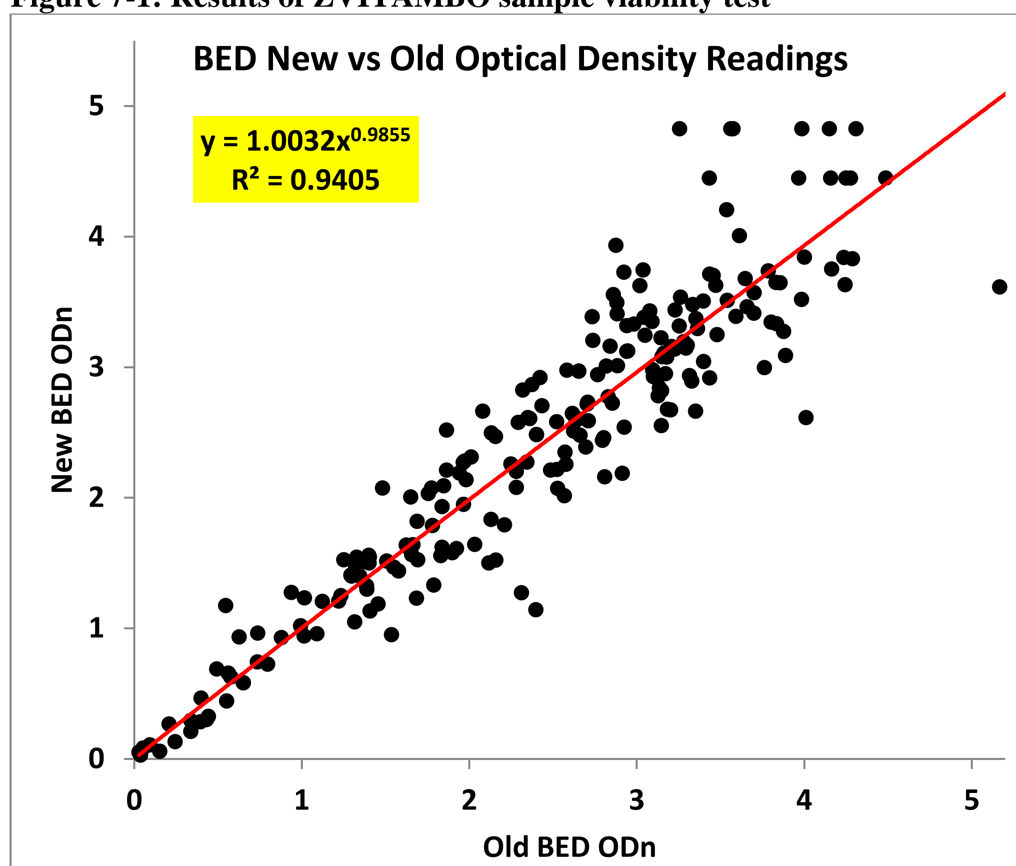
7.0 Analysis of HIV serology in the evaluation of BED, LAg and BRAI

Stellenbosch University https://scholar.sun.ac.za

7.1 Sample Viability Test

Before the new assays could be applied to the ZVITAMBO samples we needed to ensure that the samples were still viable, in the sense that they provided very similar optical densities when analysed using the BED method that was used previously by Hargrove et al. (2008).[27] We retested 224 randomly selected ZVITAMBO samples with the BED assay. Figure 7-1 shows that there was a high correlation between the optical densities originally observed in 2008 and those found in the new analysis. Moreover, the deviations observed appeared to be random with respect to the absolute value of the ODn reading, with no suggestion of systematic bias in the observed errors. We concluded that we could use these samples to evaluate the LAg and BRAI avidity assays.

Figure 7-1: Results of ZVITAMBO sample viability test



Stellenbosch University https://scholar.sun.ac.za

7.2 Review of HIV serology results

There have been significant advances in the development of laboratory tests to detect HIV infection. These tests can be classified into operational categories of assay that detect:

1) The viral antigen (mainly HIV-1 p24)
2) Antibodies against the virus (HIV-1 (Group M and O), and HIV-2)
3) Viral nucleic acid and quantify load (Viral Load tests)
4) T-Lymphocyte cell counts.

Rapid HIV test kits that detect the presence/absence of antibodies have been widely used in screening for HIV in health programs. Two characteristics are important in describing the performance of a Rapid Test Kit (RTK): (i) Sensitivity; the ability of a test to identify correctly those with disease; (ii) Specificity; the ability of a test to identify correctly those without disease. Evaluations of RTKs show that none of them has been able to achieve 100% for both sensitivity and specificity; a small percentage of patients will thus always be misdiagnosed.[78,79]

In early stages of infection, HIV-1 p24 viral antigens are the first markers for HIV infection, followed by the production of antibodies. Third and fourth generation RTKs have been developed that allow the simultaneous detection of HIV-1 p24 antigen and HIV-1/2 antibodies.[78] This simultaneous detection of antigen/antibody is instrumental in reducing the period between the time of infection and the time when antibodies become qualitatively detectable. In this evaluation, we define HIV positive infection/ serology status as the presence of HIV-1/2 antibodies as well as HIV-1 p24 antigens in acute infection.

The availability of antiretroviral therapy for suppression of viral replication supports the need for early detection of HIV infection. Beyond identifying the presence of HIV infection, the ability to provide a time variable to the infection by classifying an infection as either recent or long term is of importance in the control of the epidemic. The goal of a laboratory Test for Recent Infection (TRI) is to classify HIV positive people as either recent or long-term infections. The correct identification of the presence of HIV infection (HIV serology status) is therefore a critical first step

Stellenbosch University https://scholar.sun.ac.za

required for the application of a laboratory TRI to estimate the proportion of samples that test as recently infected with HIV. We have a two-step process in which we first classify a sample as either HIV positive or negative. We then classify the HIV positive cases as either recent or long term. Based on this, there is no pre-requirement that a laboratory assay should be able to detect HIV infection and simultaneously classify them as either recently infected or as long-term infection.

In this evaluation, when we applied the LAg and the BRAI avidity assays to the ZVITAMBO prospective cohort samples, we found that we needed to retest HIV serology of some samples, where the OD/AI results suggested that their HIV serological status might have been misclassified.

We used the manufacturer recommended criteria for resting HIV serology (see Section 6.0 Methods). For LAg assay, HIV serology retest was based on ODn < 0.400 and VL< 1000 copies/ml: for BRAI it was based on those returning an “invalid” results following retesting in duplicate.

We tested 113 plates (4,764 samples; baseline and seroconverters including repeats) using the BRAI avidity assay. We reviewed the raw data output for each plate. When the plate had 10% of the results showing an invalid result we considered this plate as having a high return on invalids and therefore required re-run of the entire plate. Eleven plates (236 /4,764 samples tested) required repeat testing due to high return of invalid results by the BRAI assay. This repeat testing resolved the status of 192/236 samples. Only 44/236 cases required testing in duplicate using BRAI assay, after which 20 cases originally diagnosed as HIV positive at baseline remained invalid by the BRAI. In Table 7-1, we show the results of this re-testing exercise, together with the results of all three biomarker tests carried out at Baseline (V0) and Visit 5 (V5) for each of the 20 cases.

Table 7-1: Results for 20 case originally diagnosed as HIV positive at baseline – but testing "invalid" by BRAI

	HIV status			BED		Lag		BRAI	CD4	Viral load	
ID	V0(O)	V0(N)	V5(O)	V0	V5	V0	V5	V5	V0	V0	V5
11413A	1	2	1	0.171	1.618	0.072	4.011	100	*	1	2201
13006C	1	2	0	0.075	*	0.330	*	*	*	1	*
13655D	1	1	1	0.043	1.979	0.438	3.812	100	559	5188	14586
16058D	1	2	0	0.031	*	0.106	*	*	*	1	*
16072G	1	2	0	0.037	.	0.139	*	*	*	1	*
16190C	1	2	2	0.029	0.034	0.198	0.242	999	*	1	1
16236D	1	2	0	0.033	*	0.122	*	*	*	1	*
16310C	1	2	1	0.030	0.906	0.167	6.033	100	619	8411	5532
16499Z	1	2	1	0.031	*	0.145	*	*	*	1	*
19552A	1	2	1	0.038	*	0.127	*	*	*	1	*
19891D	1	2	2	0.026	0.026	0.141	0.347	999	*	1	1
20195F	1	2	0	0.035	*	0.151	*	*	1246	1	*
20424C	1	2	2	0.216	0.037	0.143	0.224	999	*	1	1
20609D	1	2	2	0.028	0.038	*	0.313	999	*	1	1
21891Z	1	1	1	0.031	1.288	0.253	3.435	100	638	179025	1819
22068C	1	2	1	0.148	2.771	0.191	4.629	100	*	1	90289
22120D	1	2	0	0.085	*	0.071	*	*	992	1	*
22150N	1	2	0	0.151	*	0.090	*	*	*	1	*
22412F	1	2	1	0.041	*	0.366	*	*	*	1	*
23012X	1	2	1	0.046	3.416	0.144	5.336	95	*	1	3769

Key for column headings in Table 7-1

ID = Unique patient identifier; V0 (O) = Original Baseline HIV diagnosis; V0 (N) = New Baseline HIV diagnosis; V5 (O) = Original Visit 5 HIV diagnosis; V0= Baseline viral load; V5= Visit 5 viral load

Key for body of Table 7-1

HIV status: 1= HIV positive; 2 = HIV negative; 0 = not tested for HIV. For BRAI, 999 = "invalid result"; Viral load = 1 implies that no virus could be detected; * in any cell indicates a missing value.

Highlighting:

Blue = Original HIV positive diagnosis not changed because LAg ODn > 0.400 and/or high viral load (>5000).

Green = New V0 diagnosis is HIV negative, but clearly HIV positive at V5 [i.e., seroconverted].

Purple = New V0 diagnosis is HIV negative, but originally HIV positive at V5 – but no biomarker tests [i.e., might have seroconverted].

Yellow = New V0 diagnosis is HIV negative, and, indeed, appears clearly HIV negative at V5

Stellenbosch University https://scholar.sun.ac.za

7.2.1 Retesting of cases providing an “invalid” BRAI test result at Baseline

Our main interest here is in cases where the original test of the baseline sample produced a diagnosis of HIV positive. In one case (21356C), however, that originally tested HIV *negative* at baseline, and tested “invalid” by BRAI, the new serology indicated that the case was HIV positive. This was consistent with a LAg ODn at baseline >0.4 and a very high viral load (>800,000). This appears now to have been a very early infection: very low antibody titres can cause problems for the BRAI analysis and, as in this case, an “invalid” result in a case clearly infected with HIV. A further case (10634D) tested HIV negative at baseline but appears, in retrospect to have been HIV positive: the LAg ODn was 3.38, the BRAI AI was 92 and the viral load >100,000.

Twenty (20) cases originally tested HIV positive at baseline but returned a result of “invalid” by BRAI. In two of these cases it was decided that there was insufficient evidence to consider changing the diagnosis: in one case (13655D) the LAg ODn = 0.438 > 0.4 and the viral load was 5188; in the other (21891Z), whereas the LAg ODn = 0.253 < 0.4, the viral load was very high ($>10^5$), indicating that the case was certainly HIV infected.

In the other 18 cases (Table 7-2) one should consider changing the baseline HIV diagnosis to HIV negative, for the following reasons: (i) new HIV serology produces a negative result; (ii) LAg ODn<0.4 in all cases; (iii) very low BED ODn (<0.25) in all cases. In all cases except one (16310C) there was also no detectible virus; for the exceptional case, however, the extremely low BED and LAg values suggest that, if the case was truly infected with HIV, the infection was so new that the serology did indeed provide a correct diagnosis of negative. One further baseline cases (12156Z) was considered for reclassification in the absence of a BRAI test, on the grounds of: (i) new HIV serology produces a negative result; (ii) LAg ODn<0.2; (iii) very low BED ODn (<0.05). Table 7-2 shows the results for the cases considered for reclassification.

Stellenbosch University https://scholar.sun.ac.za

Table 7-2: Retesting of samples collected at Baseline (Visit 0)

Patient	Results string		HIV status				BED		Lag		BRAI		Viral load	
ID	Resall(old)	Resall(new)	V0(O)	V0(N)	V5(O)	V5(N)	BED0	BED5	LAg0	LAg5	BRAI0	BRAI5	VL0	VL5
11413A	1001010000	2001010000	1	2	1	1	0.171	1.618	0.072	4.0114	999	100	1	2201
12156Z	2100000000	2100000000	1	2	0	0	0.032	*	0.165	*	*	*	2575	*
13006C	1000000000	2000000000	1	2	0	0	0.075	*	0.330	*	999	*	1	*
16058D	1051000000	2051000000	1	2	0	0	0.031	*	0.106	*	999	*	1	*
16072G	1100000000	2100000000	1	2	0	0	0.037	*	0.139	*	999	*	1	*
16190C	1155550000	2255520000	1	2	2	2	0.029	0.034	0.198	0.2424	999	999	1	1
16236D	1155500000	2155500000	1	2	0	0	0.033	*	0.122	*	999	*	1	*
16310C	2155555000	2155515000	1	2	1	1	0.030	0.906	0.167	6.0328	999	100	8411	5532
16499Z	1015505000	2015515000	1	2	1	1	0.031	*	0.145	*	999	*	1	*
19552A	1110000000	2110010000	1	2	1	1	0.038	*	0.127	*	999	*	1	*
19891D	1155550000	2255520000	1	2	2	2	0.026	0.026	0.141	0.3467	999	999	1	1
20195F	1000000000	2000000000	1	2	0	0	0.035	*	0.151	*	999	*	1	*
20424C	1155550000	2255520000	1	2	2	2	0.216	0.037	0.143	0.2243	999	999	1	1
20609D	1355550000	2255525555	1	2	2	2	0.028	0.038	.	0.3130	999	999	1	1
22068C	1001151110	2001111110	1	2	1	1	0.148	2.771	0.191	4.6293	999	100	1	90289
22120D	1000000000	2000000000	1	2	0	0	0.085	*	0.071	*	999	*	1	*
22150N	1115000000	2115000000	1	2	0	0	0.151	*	0.090	*	999	*	1	*
22412F	2000030010	2000010010	1	2	1	1	0.041	*	0.366	*	999	*	1	*
23012X	1001010000	2001010000	1	2	1	1	0.046	3.416	0.144	5.3361	999	95	1	3769

Stellenbosch University https://scholar.sun.ac.za

Key for Table 7-2:

Column headings:

Patient ID – Unique patient identifier;

V0(O) = Original HIV status at Baseline   V0(N) = New HIV status at Baseline

V5(O) = Original HIV status at Visit 5   V5(N) = New HIV status at visit 5;

BED/LAg/BRAI (0) = Biomarker readings at Baseline

BED/LAg/BRAI (5) = Biomarker readings at visit 5 respectively

VL0 = Viral load at Baseline   VL5 = Viral load at Visit 5 .

Body of Table:

HIV status: 0=No sample; 1= HIV positive; 2 = HIV negative; 3=Indeterminate; 5=Sample available but not analysed.

For BRAI, 999 = "invalid result".

Viral load = 1 implies that no virus could be detected; * in any cell indicates a missing value.

Stellenbosch University https://scholar.sun.ac.za

7.2.2 Retesting of cases providing an "invalid" BRAI test result at Visit 5

Among 3,250 women originally classified as HIV positive at Visit 5, 17 cases returned an "invalid" result based on the BRAI assay, and/or had no detectible viral load and/or a very low LAg (<0.4). We retested these cases for their HIV status and all tested HIV negative (Table 7-3). They could thus potentially be reclassified as HIV negative at Visit 5. We made an exception for case 22967F, which we did not retest for HIV status, because it returned a LAg ODn slightly above 0.400 and we therefore classified the case as recent by LAg and invalid by BRAI.

A further nine cases were anomalous as they had originally tested HIV positive at Baseline and HIV negative at Visit 5. All of the Visit 5 samples for these cases gave an "invalid" result when tested with BRAI, had very low viral loads (<200), and very low LAg (<0.4). Among the baseline samples, four cases had the same undetectable viral load and low LAg profile as the group just mentioned, the BRAI result was "invalid", and the cases all tested HIV negative with the new serological test. These cases also appear in Figure 7-2 as cases that we could reclassify as HIV negative at Baseline. The final five cases continued to be anomalous: the new serology indicated that they were HIV positive at baseline but all indications were that they were HIV negative at baseline.

In Table 7-3 we show results for samples that were retested to confirm their HIV status because;

(A) samples, originally diagnosed as HIV positive; (i) gave an "invalid" test with BRAI and had no detectible viral load and/or LAg ODn<0.4 (not highlighted); (ii) had no detectible viral load and/or LAg ODn<0.4 (highlighted green):

(B) originally tested HIV positive at baseline but HIV negative at Visit 5 and (i) had no detectible viral load at either visit and LAg ODn<0.4 at both visits (highlighted green); (ii) at baseline had high viral load (>2000) and LAg (>2.5) but, at Visit 5, very low viral load (<200) and LAg (<0.4).

Table 7-3: Retesting of samples collected at Visit 5

Patient	Results string		HIV status				BED		Lag		BRAI		Viral load	
ID	Resall(old)	Resall(new)	V0(O)	V0(N)	V5(O)	V5(N)	BED0	BED5	LAg0	LAg5	BRAI0	BRAI5	VL0	VL5
10740Z	2002010000	2002020000	2	2	1	2	*	0.027	*	0.184	*	999	*	1
11107K	2002010000	2002020000	2	2	1	2	*	0.031	*	0.105	*	999	*	1
11289G	2000210000	2000220000	2	2	1	2	*	0.029	*	0.134	*	999	*	1
11436Z	2002010000	2002020000	2	2	1	2	*	0.029	*	0.126	*	999	*	1
12989A	2002010000	2002020000	2	2	1	2	*	0.038	*	0.099	*	999	*	1
13653G	2002010000	2002020000	2	2	1	2	*	0.036	*	0.089	*	999	*	1
13978N	2222212220	2222222220	2	2	1	2	*	0.037	*	0.068	*	999	*	.
14741X	2002010000	2002020000	2	2	1	2	*	0.038	*	0.135	*	999	*	1
19652K	2002010000	2332320000	2	2	1	2	*	0.031	*	0.240	*	999	*	1
22967F	2000010000	2300020000	2	2	1	2	*	0.029	*	0.442	*	999	1	1
23238G	2002010000	2002020000	2	2	1	2	*	0.029	*	0.123	*	999	*	1
23263F	2002010000	2002020000	2	2	1	2	*	0.029	*	0.143	*	999	*	1
23514F	2002010000	2002020000	2	2	1	2	*	0.029	*	0.131	*	999	*	1
14312P	2002010000	2332320000	2	2	1	2	*	0.031	*	0.182	*	*	*	1
14069C	2002010000	2332020000	2	2	1	2	*	0.039	*	0.130	*	*	*	1
17449N	2000010000	2330020000	2	2	1	2	*	0.048	*	0.343	*	*	1	1
17509Z	2000010000	2000020000	2	2	1	2	*	0.057	*	0.197	*	*	1	1
16190C	1155550000	2255520000	1	2	2	2	0.029	0.034	0.198	0.242	999	999	1	1
19891D	1155550000	2255520000	1	2	2	2	0.026	0.026	0.141	0.367	999	999	1	1
20424C	1155550000	2255520000	1	2	2	2	0.216	0.037	0.143	0.224	999	999	1	1
20609D	1355550000	2255525555	1	2	2	2	0.028	0.038	*	0.313	999	999	1	1
15080C	1015555500	1015525500	1	1	2	2	3.510	0.031	3.305	0.146	100	999	2709	1
16769A	1051555500	1051525500	1	1	2	2	1.334	0.030	2.968	0.140	100	999	79239	1
17136X	1155555000	1155525000	1	1	2	2	3.344	0.028	3.397	0.111	100	999	2171	1
22097G	1005555555	1005525555	1	1	2	2	1.520	0.028	2.765	0.130	100	999	9582	1
23847K	1155555555	1155525555	1	1	2	2	1.350	0.039	3.283	0.100	100	999	44114	157

Stellenbosch University https://scholar.sun.ac.za

Key for Table 7-3:

Column headings:

Patient ID – Unique patient identifier;

V0 (O) = Original HIV status at Baseline   V0 (N) = New HIV status at Baseline

V5 (O) = Original HIV status at Visit 5   V5 (N) = New HIV status at visit 5;

BED/LAg/BRAI (0) = Biomarker readings at Baseline

BED/LAg/BRAI (5) = Biomarker readings at visit 5 respectively

VL0 = Viral load at Baseline   VL5 = Viral load at Visit 5 .

Body of Table:

HIV status: 0=No sample; 1= HIV positive; 2 = HIV negative; 3=Indeterminate; 5=Sample available but not analysed.

For BRAI, 999 = "invalid result".

Viral load = 1 implies that no virus could be detected; * in any cell indicates a missing value.

Highlighting

Green: Original and New diagnosis is HIV negative, but clearly HIV positive at V5 (O) [i.e., seroconverted] and negative at V5 (New)

Yellow: Original HIV diagnosis is positive but V0(N), V5(O/N) are all negative

Blue: Original and New HIV are both positive while V5 (O/N) are both negative

Stellenbosch University https://scholar.sun.ac.za

7.3 Discussion on the implications of misclassification of HIV result in the BED, BRAI and LAg evaluation

7.3.1 General considerations

This evaluation highlighted ongoing challenges with HIV-1 misdiagnosis due to factors ranging from assay sensitivity and specificity to external factors.[78,79,80,81] The ZVITAMBO study used *Murex HIV-1/2* in parallel with *Genescreen HIV1/2* and Western Blot as the tiebreaker for discordancy. Researchers have reported that *Murex HIV-1/2* had a sensitivity of 100% and specificity of 99.3 (97.8-100), while *Genescreen* HIV1/2 had an equally high sensitivity of 100% and slightly lower specificity of 94.9% (91.0-98.7).[78] However, the specificity of Murex may be much lower in African setting where there is a high prevalence of schistosomiasis.[82] In the original ZVITAMBO study, only 53/14,110 cases remained indeterminate.[26]

An evaluation of BRAI to determine its ability to detect p24 and HIV1/2 antibodies showed a high detection rate of 100% for both p24 and HIV1/2 antibodies.[78, 84] Although there appeared to be a very good correlation between the results of the 18 Visit 5 samples that were retested following criteria of LAg<0.400 and a subsequent invalid result on BRAI, a decision as to whether to reclassify a sample as either negative or positive remains a subject of discussion. The observed results are consistent with the idea that the HIV testing algorithm used in the original ZVITAMBO trial may have had a higher sensitivity than the (re)testing algorithm that we used in the current project. Additionally, both a HIV negative case, and an HIV positive case that is very recently infected, will both have very low BED and LAg ODs and will in all likelihood test as "invalid" by BRAI. This is because, for the BRAI assay, an extremely low antibody titre in the sample gives a negative result in the wash buffer well and thus the AI value obtained is not valid. The basis of the BRAI assay is a comparison of the antibody binding difference between the two wells; the wash buffer and sample well. If the wash buffer well is negative, one is technically saying there is no antibody binding with which to compare, and we therefore get an "invalid" result. During early infection, p24 antigen rises rapidly until antibodies appear, but the antibody avidity is very weak at this time and, and the chance of an invalid result remains high.[85] In this regard, the "invalid" result returned by BRAI can be argued that these results reflect a recent infection and therefore do

Stellenbosch University https://scholar.sun.ac.za

not warrant, reanalysis of HIV serological status.

Addition of VL in an MAA for estimation incidence allows us to identify additional cases that may appear to be recent infections, given the serological biomarker analyses, but where the viral load suggest that they are really long-term infections. However, if the limit of detection is set at a high level, e. g. ~1000 copies/ml, this may not entirely remove false recent specimens. This means that those who are on ART and elite controllers will remain in the numerator as false assay positives (recent cases). In the original ZVITAMBO trial, viral load testing was conducted using the *Ultra-Sensitive Roche Cobas Amplicor HIV-1* monitor that had detection limits of 50 copies in milk and less than 400 copies in plasma.[25,76] These levels of detection may explain why 291 (6.6%) of the 4,391 baseline HIV positive cases tested, had undetectable VLs.

7.3.2 Specific implications for the ZVITAMBO study

Reclassifications of Baseline diagnoses

There are various implications of making changes in the original ZVITAMBO HIV diagnoses from HIV positive to HIV negative. Where we make these changes for the baseline diagnoses, these cases are, of course, no longer eligible to be included in the estimation of the False Recent Rate (FRR).

On the other hand, if reclassified as HIV negative, these cases become eligible for consideration as seroconverters. We ask, therefore, how many of the 18 cases reclassified did seroconvert? From Table 7-1, we see that in four of the cases the viral load at Visit 5 was undetectable, consistent with the idea that the case was never HIV positive. A further three cases were never seen again after baseline.

The other 11 cases appear to have seroconverted after baseline. Thus, four cases that tested HIV positive at Visit 5 had high viral loads ranging between 2000 and 90,000, and were clearly seroconverters. Seven (7) other cases tested HIV positive prior to, or after Visit 5, though we could not assess viral loads, or BED, BRAI and LAg levels, for these cases. It would appear, therefore, that at least four, and as many as 11, cases

Stellenbosch University https://scholar.sun.ac.za

out of 18 newly classified as HIV negative, seroconverted after baseline. Moreover, given that three cases were never actually seen after baseline, the proportion seroconverting could be viewed as anywhere between 4/15 and 11/15. Given that the HIV incidence, estimated by follow-up in the ZVITAMBO cohort, was of the order of 3% per annum, the expected number of seroconversions out of 18 cases is <1 in a year. It is thus intuitively obvious that the probability of observing 11 seroconversions out of such a small samples is extremely small. In fact, a simple calculation shows the probability is of the order of 10^{-13} (Hargrove, personal communication). Moreover, even the probability of observing four seroconversions is ≈0.001.

Reclassifications of Visit 5 diagnoses

At Visit 5 there are 17 cases, originally diagnosed as HIV positive that could be potentially reclassified as HIV negative (Table 7-3). The effect of doing so has an important impact, as we shall see below, on the *biomarker* estimates of the HIV incidence over the first 12-months postpartum. The reasons for this are that:
(i) In all 17 cases, the mother was originally diagnosed as HIV negative at baseline.
(ii) If diagnosed as HIV positive at Visit 5 she was thus, by definition, a seroconverter. (iii) For all 17 cases, the BED and LAg ODn values measured at Visit 5 were much lower than any feasible OD cut-off that might be chosen. (iv) It thus follows that, if diagnosed as HIV positive at Visit 5, all 17 cases would also be classified as recent seroconversions. (v) Reclassifying all 17 as HIV negative at Visit 5 would thus reduce the number of recent cases by 17 in every incidence calculation – and the incidence estimate is very sensitive to the numbers of infection classified as "recent".

The reclassifications would also influence the follow-up estimates of incidence, but the impact is smaller, because the reduction is now a proportion of all seroconverters over the period between birth and 12-months postpartum – and not simply among those judged as *recent* infections at Visit 5.

The reclassifications could also affect the estimates of the MDRI – because, again, we would no longer view the 17 cases as seroconverters. Inspection of Table 7- 2 shows, however, that for virtually all cases, Visit 5 was the first and only time that these

Stellenbosch University https://scholar.sun.ac.za

women tested HIV positive. Given that, for LAg and BED, we always required at least two positive samples to be provided for a case to be included in MDRI analyses, the reclassifications have no effect on the MDRI estimation procedures.

Impact of reclassifications on analyses: Differential impact depending on the biomarker(s) used to identify "recent" seroconverters

As noted above, reclassifications of some cases as HIV negative affects FRR and (some) MDRI estimates and thus affects the resulting HIV incidence estimates. The implications are, however, different for analyses using, on the one hand, BED and LAg and, on the other, BRAI. In the former cases we have to choose which diagnoses we are going to use.

For BRAI there is no such choice, because all of the disputed cases always test "invalid" and should not be included in the analyses – unless we take the decision that the "invalid" result is incorrect and we simply decide to classify the samples as recent infections. If we take this view, however, there would then be no justification for any reclassifications and we would effectively just be using the old classifications.

When we use viral load, together with the serological biomarkers, to identify recent infections, these problems largely fall away. The reason for this is that, when we use viral load, a case is *only* classified as "recent" if, and only if, the serological biomarker suggests that the infection is "recent" *and* the viral load is >1000. Since, however, we only ever reclassified a case as HIV negative if the viral load was <1000, it follows that the ambiguous cases are either classified as "long-term", if we accept the original HIV diagnoses, or as HIV negative if we accept the new diagnoses. In the ZVITAMBO Trial, >65% of all cases were HIV negative, and ~90% of HIV positive cases test as long-term infections at recommended values of the pre-set cut-off. Reclassifying a small number of "recent" cases as either "long-term" or "HIV negative" results in a relatively small change to a large number. The effect on ensuing parameter estimates is accordingly negligible. We discuss this matter further in Section 10.2.2 and Section 10.6**.**

Stellenbosch University https://scholar.sun.ac.za

***Which classifications should we use*?**

The very low LAg and BED ODn values and the general absence of detectable virus for cases that return an "invalid" BRAI test, coupled with an HIV negative diagnosis from the new serological tests, are all consistent with the idea that the cases were originally misdiagnosed as HIV positive.

Conversely, the vanishingly low probability of observing such large proportions of seroconverters among cases reclassified as HIV negative at Baseline (see above) raises the possibility that the real reason for the problem arises from the higher sensitivity of the original HIV tests. The results are consistent with some or all of the disputed cases being extremely recent HIV infections, consistent with their very low BED and LAg ODn values – and with the "invalid" classification from the BRAI analyses. What is unexpected and finds no ready explanation is the absence of detectable virus.

We are unable to provide an unequivocal solution to this problem. Accordingly, given the uncertainty of the results, we explored analyses using both the "Original data" and revised data classified as "New data" in this report. Following data cleaning of datasets, the BRAI dataset had 4,466/4,495 (99.4%) baseline records and 2,824/3010 (93.8%) V5 HIV positive, the LAg dataset had 4,468/4,495 (99.8%) baseline records and 2,825/3010 (93.8) V5-HIV positive and the BED dataset had 4,418/4495(98.6%) and 2735/3010 (90.9%) HIV positive at V5 (Figure 6-2). The numerator refers to the actual number of samples tested and the denominators refer to the total sample that was in storage at ZVITAMBO laboratory and were available for testing in this current evaluation.

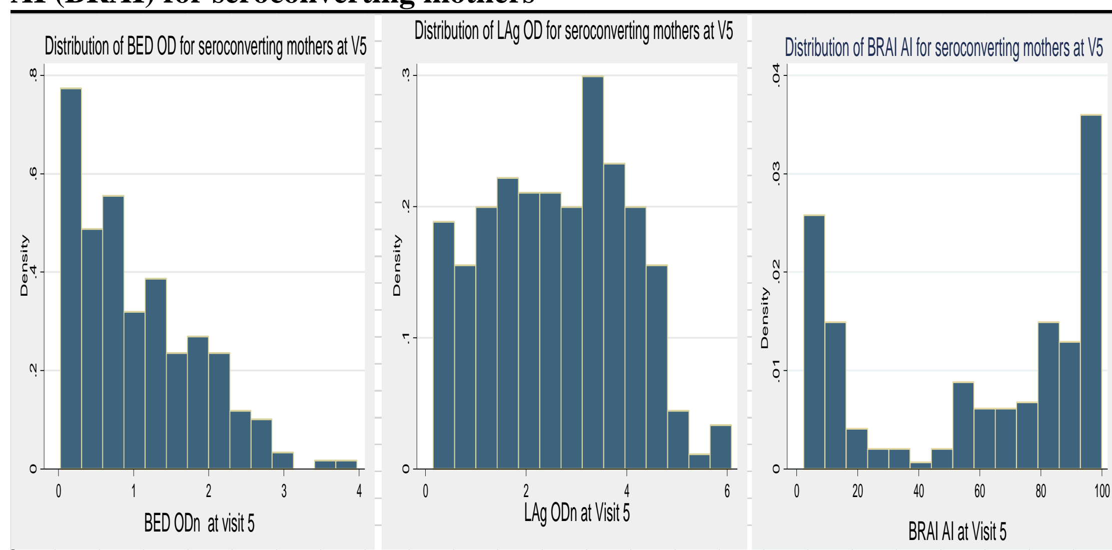
When the new avidity assays are applied to cross-sectional surveys samples, there is a likelihood that misclassification of incident vs long term could result because of HIV-1 misdiagnosis, and this has potential to cause imprecise HIV incidence measures. This can be mitigated by continuous development of test kits that have dual capacity to detect presence of HIV and concurrently classify the infection as either recent or long term as well as use of MAA.[83,84,86,86]

Stellenbosch University https://scholar.sun.ac.za

7.4 Optical density patterns among seroconverters as observed using BED, LAg and BRAI

We plotted the distribution of ODn and AI readings for all mothers who were HIV negative at baseline and seroconverted during the first 12 months of follow-up (V5) (Figure 7-2).

Figure 7-2: Distribution of optical density (ODn) readings (BED and Lag) and AI (BRAI) for seroconverting mothers

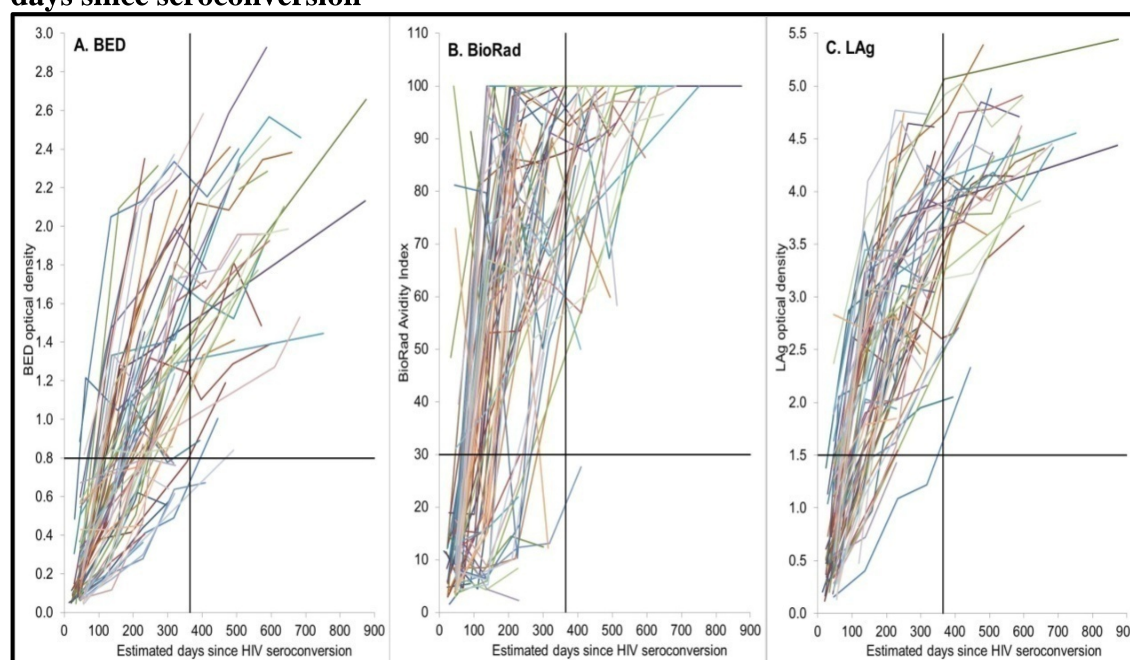


Visual inspection of normalised optical density (ODn) and Avidity Index (AI) patterns shows more evenly distributed ODn readings in LAg than in BED, and wider variation in AI for BRAI.

We plotted graphs for ODn and AI against days since seroconversion for cases among 182 seroconverters where t_0 (the time between the last negative and first positive result) was at most 120 days, and where we took blood samples from the mother on at least three independent occasions after she seroconverted (Figure 7-3).

Stellenbosch University https://scholar.sun.ac.za

Figure 7-3: Distribution of ODn for BED and LAg and AI for BRAI by estimated days since seroconversion



The increase in optical density, which is a measure of increase in antibody maturation for both the LAg and BRAI, showed that there was less variability for the two avidity assays than for BED. This visual depiction of the optical density by days since seroconversion provides an analytical picture of the biochemical interactions between the sample and assay in relation to days since initial infection.

For the BRAI assay, the AI is set to increase to a maximum value of 100%, so the graph for BRAI shows a plateau at the 100% mark (Figure 7-3). It is possible that cases will fail to provide ODn/AI above the cut-off because they are elite controllers. However, Figure 7-3 shows that, for LAg, virtually all cases progressed above cut-off if seen for a sufficient length of time post-seroconversion. Indeed, no case tested at a time >365 days since seroconversion had an ODn < 1.5, the recommended cut-off. We did, however, see non-progression in a few cases for BED and BRAI assay, suggesting that non-progression may be more a function of assay kinetics rather than the presence of elite controllers. The rates of change in optical density with time since seroconversion varied between individual women, and this variation between mothers was greatest for the BED assay and smallest for LAg.

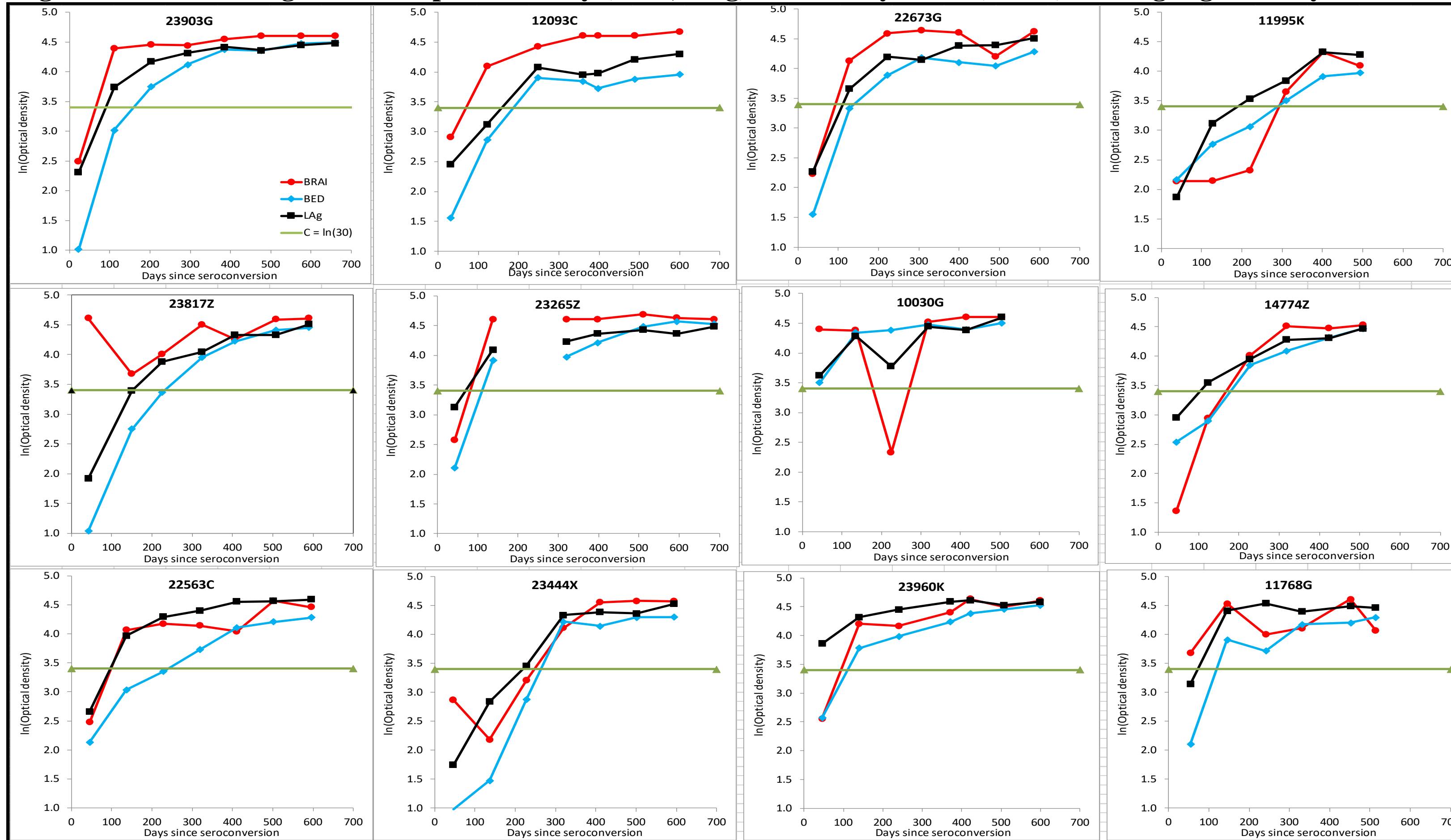
Stellenbosch University https://scholar.sun.ac.za

We plotted the log optical density readings against the days since seroconversion for 12 mothers who provided at least six samples following seroconversion (Figure 7-4). For purposes of comparison only, we multiplied each ODn and AI values by different constant factors for each biomarker.

For all three assays, there was a general increase in ODn with increase in time since seroconversion (Figure 7-4). While for BED and LAg, the pattern of increase followed a somewhat regular pattern, for the BRAI the curve was less regular with some cases showing reversal in log ODn readings to below the cut-off (e.g. 10030G).

Stellenbosch University https://scholar.sun.ac.za

Figure 7-4: Natural logarithm of optical density (BED, LAg) and avidity index (BRAI) readings against days since seroconversion



Stellenbosch University https://scholar.sun.ac.za

7.5 Discussion and conclusion on variation in optical density readings for BED, LAg and BRAI

Measuring optical density or avidity index for seroconverting samples is the operational method by which samples are classified either as recent or long term based on whether the readings lie below or above a predefined cut-off. The plots of optical density readings are critical in showing how an assay performs in relation to the amount of antibody or antigen present as a function of time since infection; i.e., it is a direct measure of the size or magnitude of a positive EIA result, expressed as optical density (ODn).[87] By plotting the distribution of the optical density readings of samples from seroconverting individual persons, we have shown the variations in kinetics of BED, LAg and BRAI.

In our current evaluation conducted on sub-type C samples, BED had the widest variation in ODn plots with time since seroconversion. Changes in BRAI AI with estimated time since seroconversion showed markedly more variation than observed in the change in LAg normalised optical density (ODn). There is a rapid change in AI for recent infection as a result of BRAI's ability to measure IgM. This can be explained by the fact that although IgM antibodies have low binding affinity each antibody has 10 antigen binding sites relative to two for IgG; therefore IgM detection may result in a high avidity reading early on compared to the two for IgG in later infection. Alternatively, this steep rise in BRAI may actually show a high avidity in cases with low total antibody levels and these cases may show low LAg values. In contrast, for the LAg assay we suggest that the lower values in early infection could be because of a concentration too low to saturate all antigen binding sites (limiting antigen in each well); or a true low avidity that result in a low value during early infection. We observed this in actual values for poorly correlated results, where LAg and BED had much lower ODn readings. These findings are consistent with changes observed in Fiebig's acute HIV staging system.[88]

Stellenbosch University https://scholar.sun.ac.za

Based on the assays' biochemical performance, very recently infected cases can be incorrectly classified as "invalid" by BRAI, or as long-term infections, consequent on low antibody titres in the reference well of the two-well assay. This problem does not occur for LAg, but cases incorrectly diagnosed as HIV positive will be (incorrectly) diagnosed as recent infections by LAg, whereas they will (appropriately) be classified as "invalid" by the BRAI assay. The manufacturer of LAg have mitigated this by requiring that all samples returning an ODn less than 0. 400, be retested for HIV-1-2 serology using an algorithm with high sensitivity and specificity.

A quantitative analysis of the increase in HIV Type I antibodies after seroconversion for BED assay showed wide variability of individual ODn plots that was attributed to subtype differences in the biology of infection leading to variability in antibody response.[14,87] This analysis suggest that measurement accuracy is higher for BRAI and LAg than for BED, but particularly for the LAg assay. In the evaluation by Parekh et al. (2002), of 259 seroconverters using the LAg assay, results showed that the assay has an increased avidity for gp41 specific antibodies post-seroconversion that levels off after 500 days.[53] In a few cases involving elite controllers, the avidity remained low and sub-optimal panels showed wide variations. This suggests the need to identify and remove such samples when calculating MDRI for selected populations in order to get better accuracy of the estimates. On the other hand, the BRAI assay is capable of measuring the IgM so that the AI rises rapidly in early infection. The two-well calculation of the avidity index has a maximum of 100%. This poses a challenge when this proportion is beyond 100% such that AI variability plots assume a plateau. A cause for concern with BRAI is the increased variability beyond one year as shown by the return of AI below the cut-off value for some cases. This occurrence may be associated with inherent assay variability or immune escape variants arising with resulting new low affinity antibody responses to new antigen epitopes. Further work is required to elucidate the cause of these problems.

Stellenbosch University https://scholar.sun.ac.za

8.0 Determination of Mean Duration of Recent Infection (MDRI) for BED, LAg and BRAI assays

Stellenbosch University https://scholar.sun.ac.za

8.1.1 Introduction

We define the MDRI for an assay as the mean time that a person remains in a state of recent infection, while infected for less than a pre-defined period T. During this period, an HIV infected person's blood sample returns an optical density reading below a pre-selected ODn or AI, referred to as the cut-off value C. The laboratory test property of MDRI would require a C low enough to minimise the FRR (such that almost all patients with long-term infections would have values higher than C). At the same time, it must still be long enough not to require a very large sample size to estimate incidence for the predefined time T. [62,63,]

The estimation of MDRI depends on assay interactions with antibody kinetics, which vary between individuals, by HIV subtype and geographical region. In general, the MDRI increases with increasing cut-off value C. [21, 54,69] An increase in C results in the monotonic increase in absolute number of individuals categorised as recent and, may also result an increase in the of proportion of false recent cases. In general, the exact time point at which seroconversion occurs is not known. We assume that it must occur at some point during the period (t_0), i.e., between the last negative and the first positive HIV tests. Similarly, we do not know the exact time point at which a person leaves the state of recency i.e., when the ODn or AI value first exceeds C.

Accordingly we estimate the time (τ) say, between seroconversion and leaving the recent state, and estimate the MDRI to be either the observed value of τ, if $\tau < T$, or we estimate the MDRI as T, if $\tau \geq T$. In this study we generally set T= 1 year (365 days). Researchers suggested that extending T to a period of 2 years yields better estimation of HIV incidence derived using MDRI and FRR. [20, 65,73] The use of $T = 2$ years seemed inappropriate for our study since only a small minority of cases were followed up for 2 years, and none for longer than that period. Nonetheless, we did compare MDRIs estimated using T as either 1 or 2 years.

Stellenbosch University https://scholar.sun.ac.za

8.1.2 Effects of varying the value of Time (*T*) and Number of Samples (ns) on estimated values of MDRI

Workers have proposed several statistical methods for calculating MDRI. Sweeting et al. (2010) explored two statistical methods for measuring the window period of an assay. [77] The first method uses survival analysis to derive a non-parametric maximum likelihood estimate of the window period.[77] This analysis used the date of last negative and the first positive test results and assumes knowledge of defined interval where a person exists out of recent state α. They dismissed this method based on its failure to take into account repeated biomarker measurements and failure to account for errors in the biomarker process. The second method uses a mixed-effects model to describe the growth of an antibody assay while incorporating uncertainty associated with the seroconversion time. We found this method to be more plausible in estimation of MDRI.

In applying the BED assay to prospective cohort samples, one of the major issues that became known was the over-estimation of HIV incidence estimates.[27] One of the theories proposed to explain this discrepancy was the existence of a proportion individuals misclassifying as recent (ε now called FRR) despite being infected for periods greater than order one year. Another explanation was that the window period, the total time an HIV case spends as a recent in their entire life after infection was an under-estimate of the population MDRI.[21] It was possible, therefore, that estimates of incidence derived using these parameters differed from estimates derived from the follow-up of cohort of initially HIV cases, as in the ZVITAMBO study, because the MDRI was being under-estimated. As a result of this, it has been proposed that both MDRI and FRR, and incidence estimates, should be calculated with reference to a fixed time period T.[21] This T has been arbitrarily set at one year (365 days), but because of the observed increase in antibody beyond this period, 2 years has been proposed as an alternative option.[20,63,62,73] Using T= 2 years did not seem appropriate for this present study as most women were not seen beyond $T > 1$. Nonetheless, we explored the effects of extending T beyond the 1 year, examining plots of optical density/AI versus time since seroconversion for BED, LAg and BRAI using all samples in Section 7.

Stellenbosch University https://scholar.sun.ac.za

We found that there was a rapid antibody maturation in the period <400 days (Figure 7-3). In this section, we took a closer look at the pattern of increases in ODn/AI versus time since seroconversion using only samples that seroconverted in the first 12 months of follow-up (Section 8- 2.1) in order to see if extending the period to two years would yield any meaningful difference. We describe this sub-analysis in Section 8.2.4.

Recall that we define the variable t_0 as the time between the last HIV serology negative, and the first positive, test results. We need to decide on acceptable values of t_0 in selecting the cases we use in estimating the MDRI. As t_0 increases, we become less and less certain about the time of seroconversion. On the other hand, as t_0 decreases, fewer cases qualify for inclusion in the analysis. Similarly, there is a question about the minimum required number (*ns*) of independent HIV tests made on each HIV positive case before the case is included in the analysis. If *ns* is large we will get a good picture about the pattern of increase in ODn, or AI, values, but reduced numbers of cases will qualify to be included in the analysis. When we reduce the required minimum for *ns*, the sample size increases but we are less certain about the pattern of increase. Clearly there is also a relationship between t_0, the minimum number of samples per case (*ns*) and the resulting number (*N*) of cases (mothers) contributing to the test.[21]

An important output from the calculation of MDRI is the coefficient of variation (CV), which is a measure of precision (variability) and calculated as the standard deviation (error) of the estimate divided the mean, and usually expressed as a percentage. The lower the value, the more precise the measurement and vice versa. In the Hargrove et al. (2012a) analysis, when the number of samples each person/case contributes was kept constant and t_0 was increased from 80 to 160 days, the CV of the estimate declined.[21]

Stellenbosch University https://scholar.sun.ac.za

8.1.3 Methods

Building on the work of Sweeting et al. (2010), Hargrove et al. (2012a) described five methods for fitting ODn data in order to derive the MDRI for BED.[21,77] Although these five methods provided comparable estimates of MDRI, the Non Linear Mixed Model (NLMM), with a slight modification of the method suggested by Sweeting *et al.* (2010), provided the best fit for BED ODn data with the smallest variance among all the methods. While the method of Survival analysis (SA) (Turnbull, 1976) and Linear Mixed Model (LMM) had similar estimates, the coefficient of variation for SA estimates was at least 3 times that of NLMM when applied to BED data.[89] Recently, Duong et al.(2015) conducted a recalibration of the LAg MDRI calculation.[69] This exploration of the calculation of MDRI and FRR concluded that although there was coherence in different statistical methods, there was a need to balance a long MDRI and small FRR. They suggested the use of a cut-off of 1.5 for estimation of LAg MDRI as the optimum cut-off: they also suggested that binomial regression analysis as the preferred method for estimating the MDRI.

Each method for calculating MDRI has its own assumptions about the exit from state of recent to non-recent. Some methods assume a single exit from the state of recent to non-recent while others assume multiple transitions between states. The feasibility of each method depends on the available data such as number of subjects and follow-up data points per subject. We give a brief description of some of the methods below:

1. **Graphical method**: Consecutive ODn are plotted against the time since seroconversion. This method uses linear interpolation between two data points and produces wide variations, it was therefore not considered suitable for estimation of MDRI.[21]

Stellenbosch University https://scholar.sun.ac.za

2. **Proportion of recent infections among seroconverters (*r/s*).**

Assuming uniformly distributed seroconversion events over the time-period [0, *T*] the MDRI is estimated from the proportion of all seroconverters testing recent according to a pre-set ODn/AI cut-off. It has been argued that, if seroconversions are uniformly distributed, the ratio *r*/*s* (*r* and *s*, defined below) should provide an estimate of the MDRI close to the estimate required to ensure equality between follow-up and BED estimates of incidence.[21, 62] This method may be useful where there are limited data points, even when only one follow-up data point is available. The MDRI (Ω_{Trs}), defined for a pre-set cut-off *C*, is estimated by;

$$\widehat{\Omega}_{Trs}(C) = r/s \qquad (1)$$

where *s* = Number of HIV positive samples at time *T* among cases that were HIV negative at time 0

r = Number among *s* testing recent at cut-off *C*

Hargrove et al (2011) showed that the ratio r/s should thus provide an estimate of the MDRI that is close to the estimate required to ensure equality between follow-up and biomarker estimates of incidence.

3. **Linear Mixed Model (LMM)**: This method uses a LMM with fixed and random effects to model a linear relationship between the ODn and time since infection (transformed according to the recommendation of the developer). This model yields a straight line for each woman, from which we estimated the time spent in the recent state, using an inverse prediction technique, with the upper limit restricted to *T*. The equation for modelling the change in optical density reading is:

$$\sqrt{OD_{ij}} = A_i + B_i\ In(t_{ij}^{0}) + e_{ij} \qquad (2)$$

Where A_i and B_i are constants containing fixed and random effects

t_{ij}^{0} = time at observation j since last HIV negative test and

e_{ij} = Independent and identically distributed normal errors

Stellenbosch University https://scholar.sun.ac.za

Bootstrap techniques are applied to these individual estimates to obtain the final estimate of the MDRI as well as the associated confidence interval.[21]

4. **Non-linear Mixed Model (NLMM):** This method models the relationship between assay ODn and time since infection. Sweeting et al. (2010) modelled the BED change in optical density reading from *i* to *j* as;

$$OD_{ij} = a_i + (b_i - a_i)\exp(-c_i t_{ij}) + e_{ij} \qquad (3)$$

Where a_i, b_i and c_i are constants, t_{ij} is the time since infection and e_{ij} are independent and identically normally distributed errors.

Unlike the LMM, this function approaches finite asymptotes for short and long times since infection. The NLMM uses a Bayesian approach where, for this model, time of seroconversion is assumed to be uniformly distributed between the dates of last negative and first positive HIV tests. In this method, we used Markov Chain Monte Carlo (MCMC) methods to obtain the distribution of individual posterior recency duration up to a maximum period *T*. The method also assumes that the biomarker processes increase monotonically. So the MDRI and CI are obtained from this distribution of individual MDRIs.[21] Hargrove et al. (2012) introduced a variant of this method that provided a better fit to the data for individual cases. [21,62] Using this method, we modelled increases in optical density with time (*t*) since HIV seroconversion as:

$$\log(OD_{ij}) = a_i + (b_i - a_i)\exp(-c_i t_{ij}) + e_{ij} \qquad (4)$$

where $c_i>0$, $a_i>b_i$ and e_{ij} are independent and identically distributed normal errors.

In this model, the function approaches an asymptote for large values of *t* and most importantly, it goes to zero as *t* approaches minus infinity. This method is biologically more plausible than the LMM.[21]

Stellenbosch University https://scholar.sun.ac.za

5. Survival Analysis: The exact time of seroconversion and time to reach a predefined cut-off are unknown; therefore, there we need to assume the data are double interval censored. We obtain the intervals from longitudinal measurements of AI. The estimated dates of seroconversion lie the between dates of the last HIV negative and first HIV positive tests. We obtain the interval for the time to reach the predefined cut-off by using the date of the last ODn below cut-off and the first date with ODn above the cut-off. Because this creates intervals of shortest and longest MDRI, for each individual Sweeting et al. (2010) used this to calculate the lower and upper bounds for MDRI. The method was discarded because it resulted in an incorrect likelihood function.[78] Hargrove et al. (2012a) proposed an alternative approach. They approximated seroconversion to have occurred at midpoint between the last negative and first HIV positive test result.[24] For the ZVITAMBO data, the average time was 83 days and a maximum of 120 days, therefore margin of error was likely to be minimal. This method uses single interval censored data and Turnbull's modification of Product-Limit Estimator to obtain a survival function. When this is integrated over $[0,T]$, it provides an estimate of MDRI (Ω_{TS}) and its corresponding confidence interval.[2] The main advantage of this method is that it does not have parametric assumptions so that the precision of the MDRI estimate is not affected by independent and varying parameters.[24] However, some limitations are that it does not use the shape of ODn/AI with time, because it assumes a continuous sojourn in a state and so it is not possible to measure multiple transitional states such as are consistent with TRI assays. More detailed analysis of methods 1-5 are provided in Hargrove et al. (2012b)[24]

6. Binomial Regression analysis: Regression techniques can be used to model the evolution of a quantitative biomarker with time since infection. This method estimates the probability of testing recent as a function of time since estimated date of detectable infection $P_R(t)$[69,73]

$$g(P_R(t) = f(t) \tag{5}$$

Stellenbosch University https://scholar.sun.ac.za

A number of parametric forms are possible, but we fit a linear binomial regression model for P_R, as a cubic polynomial in *t*. MDRI is then the integral of $P_R(t)$ from zero to *T*. This has been arbitrarily set at 365 days or *T*= 1 year. We can model of $g()$ using the logit function and $f(t)$ a cubic polynomial in t and the model is presented as;

$$\ln(P_R(t)/(1-P_R(t)) = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 \qquad (6)$$

In this evaluation, we explored the estimation of MDRI using Turnbull SA, NLMM and binomial regression methods for BED, LAg and BRAI using the original HIV serology classification.

Stellenbosch University https://scholar.sun.ac.za

8.2 Results of MDRI Calculations

MDRI calculation were conducted using samples from 351 women who tested HIV serology negative at baseline and seroconverted during the follow-up period (postpartum). We used samples obtained from 182 women who provided at least two (2) HIV positive samples after seroconversion, resulting in 593 samples that were included in the MDRI calculation.

We present graphical plots of the optical density/avidity index reading by time since seroconversion for the three assays to highlight assay dynamics in response to antibody maturation.

We present results of MDRI in which all samples use the original HIV serology classification, by itself, to diagnose cases as “recent”, if the serological level is $<C$ – the pre-set optical density (OD*n*) or avidity index (AI) cut-off, or “long-term” otherwise. Then we present results of MDRI when we use both viral load (VL) and a serological biomarker in an MAA algorithm. We then classify a case as “recent” if, and only if, the measured level of the serological biomarker is $<C$ *and* the viral load is $\geq$1000. Otherwise, we define the case as “long-term”. Lastly, we present results for the data set in which we changed HIV diagnoses in the “Original data” from HIV positive to negative in a small number of cases, to produce what we call the “New Data”.

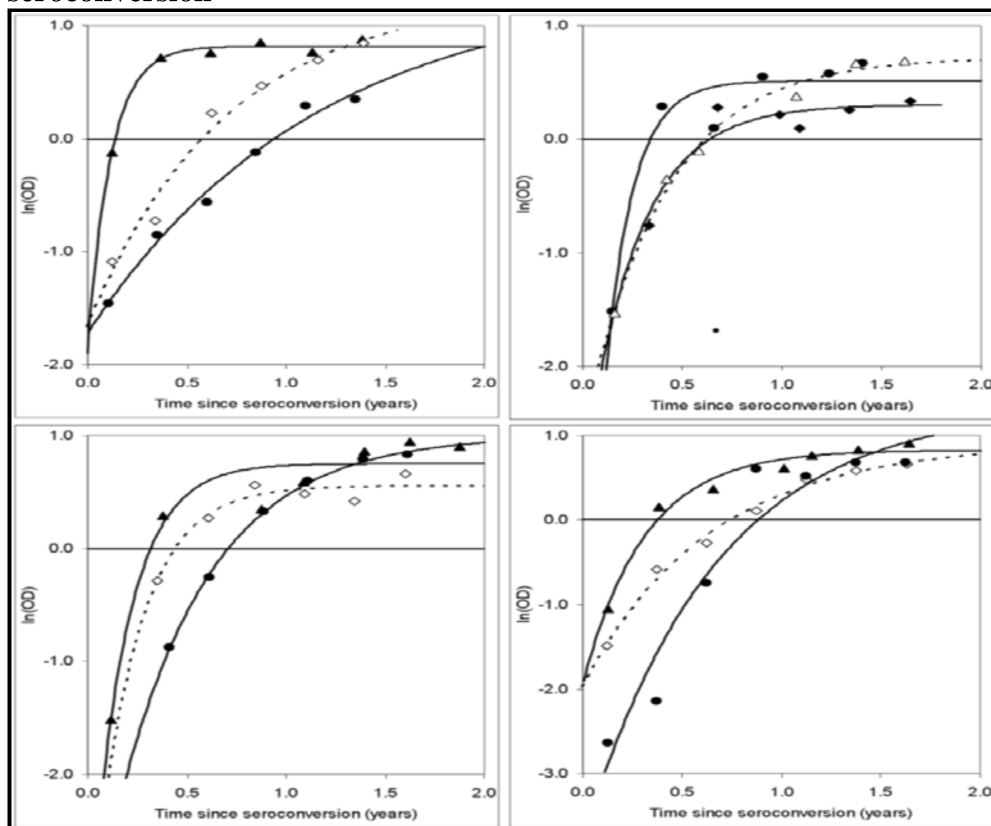
Notice that the sample size is always smaller when we use viral load (VL) and a serological biomarker together, because there are some missing values for viral load and this can cause the number of samples per case to drop below the minimum required value of *ns*.

Stellenbosch University https://scholar.sun.ac.za

8.2.1 Changes in biomarker optical density as a function of time since seroconversion

Hargrove et al. (2012a) plotted the ln(ODn) and ln(AI) readings against time since seroconversion for women who provided at least six samples following the first HIV positive result (Figure 8-1)

Figure 8-1: Natural logs of BED optical density readings against days since seroconversion



***Used with permission from Hargrove et al. 2012a.**[21]

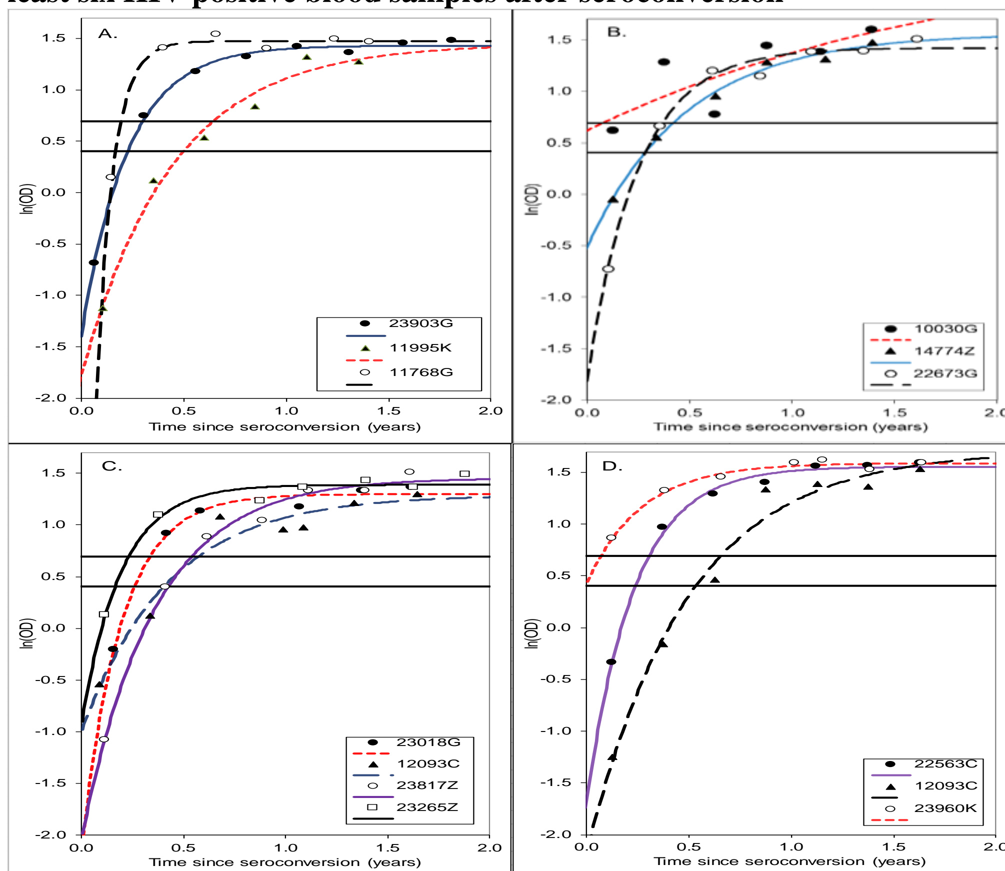
For the BED assay (Figure 8-1), the pattern of increases in ODn with time since seroconversion for case *i* seen at visit *j,* is described by the function:

$\log_e(OD_{i,j}) = a_i + (b_i - a_i)\exp(-c_i t_{i,j})$ as described in Hargrove et al. 2012a.[21]

In this evaluation, we plotted natural log of LAg optical density against the days since seroconversion for selected cases that had provided at least six samples after seroconversion (Figure 8-2).

Stellenbosch University https://scholar.sun.ac.za

Figure 8-2: Natural logs of LAg optical density readings against days since seroconversion for selected cases where the seroconverting mother produced at least six HIV positive blood samples after seroconversion

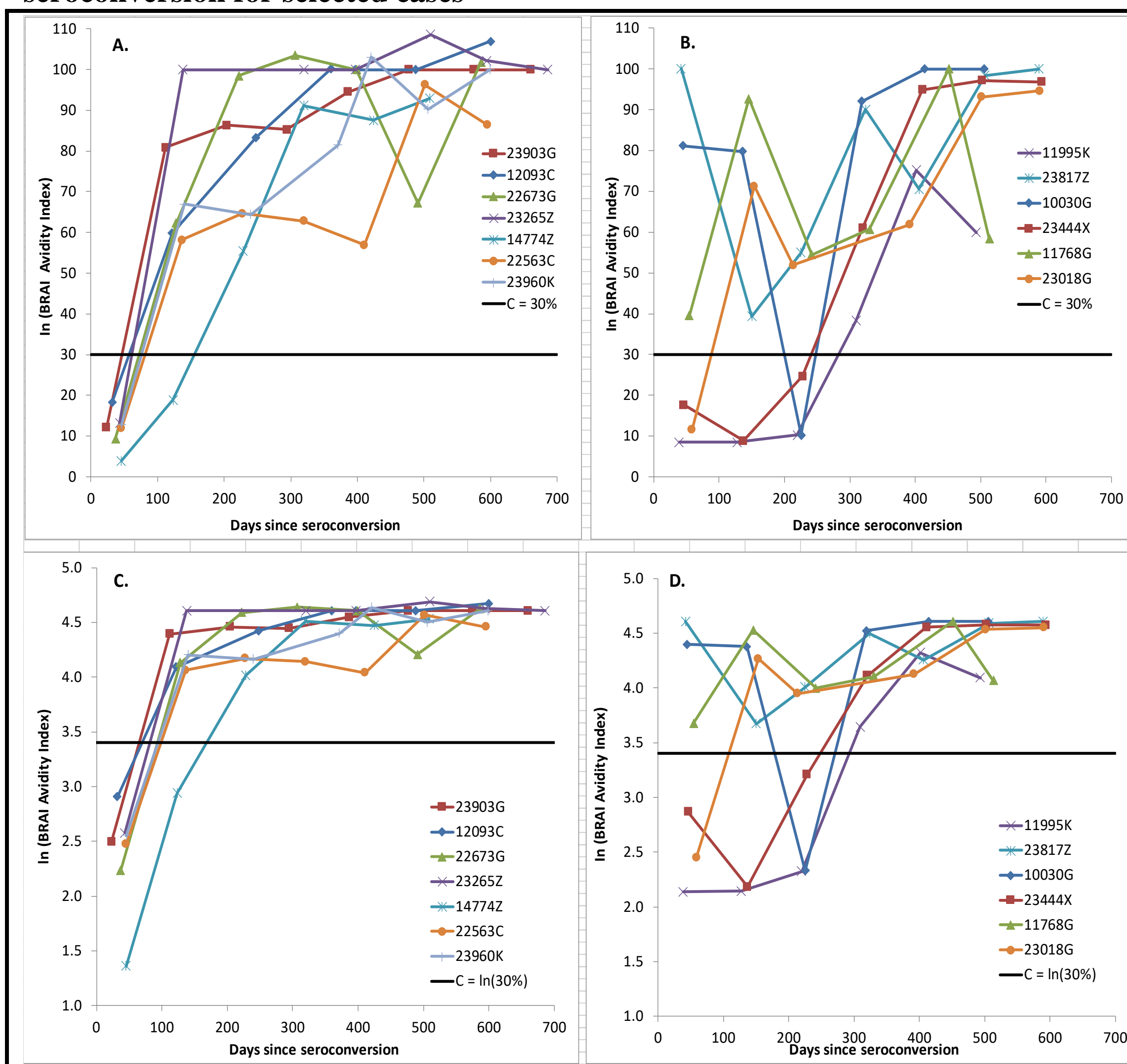


For the LAg assay the pattern of increases in ODn with time since seroconversion (Figure 8-2) shows a similar mathematical form to the increases seen in BED.[21] The rate of increase is much higher in LAg and follows an even closer adherence to a parametric model than the BED, suggesting the possibility of achieving better precision in the MDRI estimates.

Stellenbosch University https://scholar.sun.ac.za

In contrast, the development of the BRAI AI with time (t) since seroconversion-conversion is less consistent than for BED and LAg (Figure 8-3). Some cases (A) show steady increase in AI with t, but in others (B) there are major trend reversals, sometimes with the cases re-entering the recent state (Figure 8-3). This causes departures from the kind of parametric model that could provide a good description of AI increases with time since infection. We therefore focused the analysis of BRAI MDRI on Turnbull SA methods and binomial regression.

Figure 8-3: Natural logs of BRAI log optical density readings against days since seroconversion for selected cases



Note: Seroconverting mother produced at least six HIV positive blood samples

Stellenbosch University https://scholar.sun.ac.za

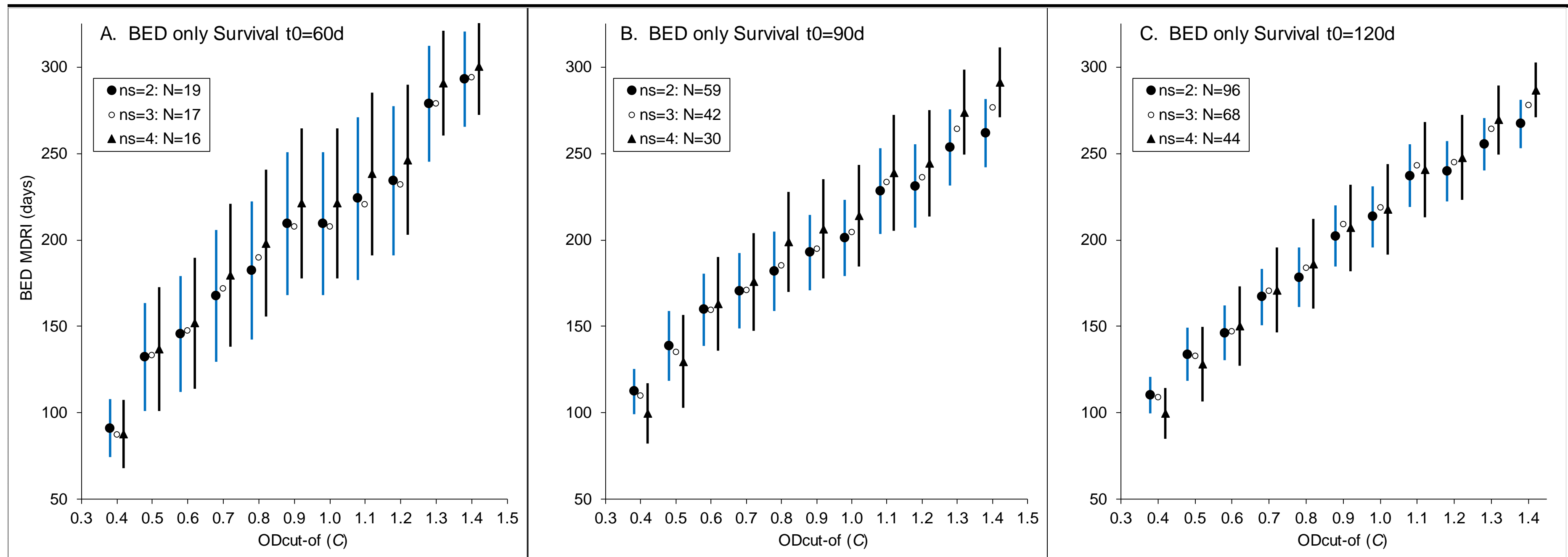
8.2.2 MDRI based on Turnbull Survival Analysis (SA)

We varied t_0, cut-off and calculated the corresponding MDRI and CI around these estimates for BED (Figure 8-4), BRAI and LAg (Figure 8-5).

Regardless of the cut-off, the greatest precision, as defined by the smallest CV of the MDRI estimates was achieved with t_0 = 120 days. The estimate with narrowest CI was achieved when we set t_0 = 120 days and ns=2.

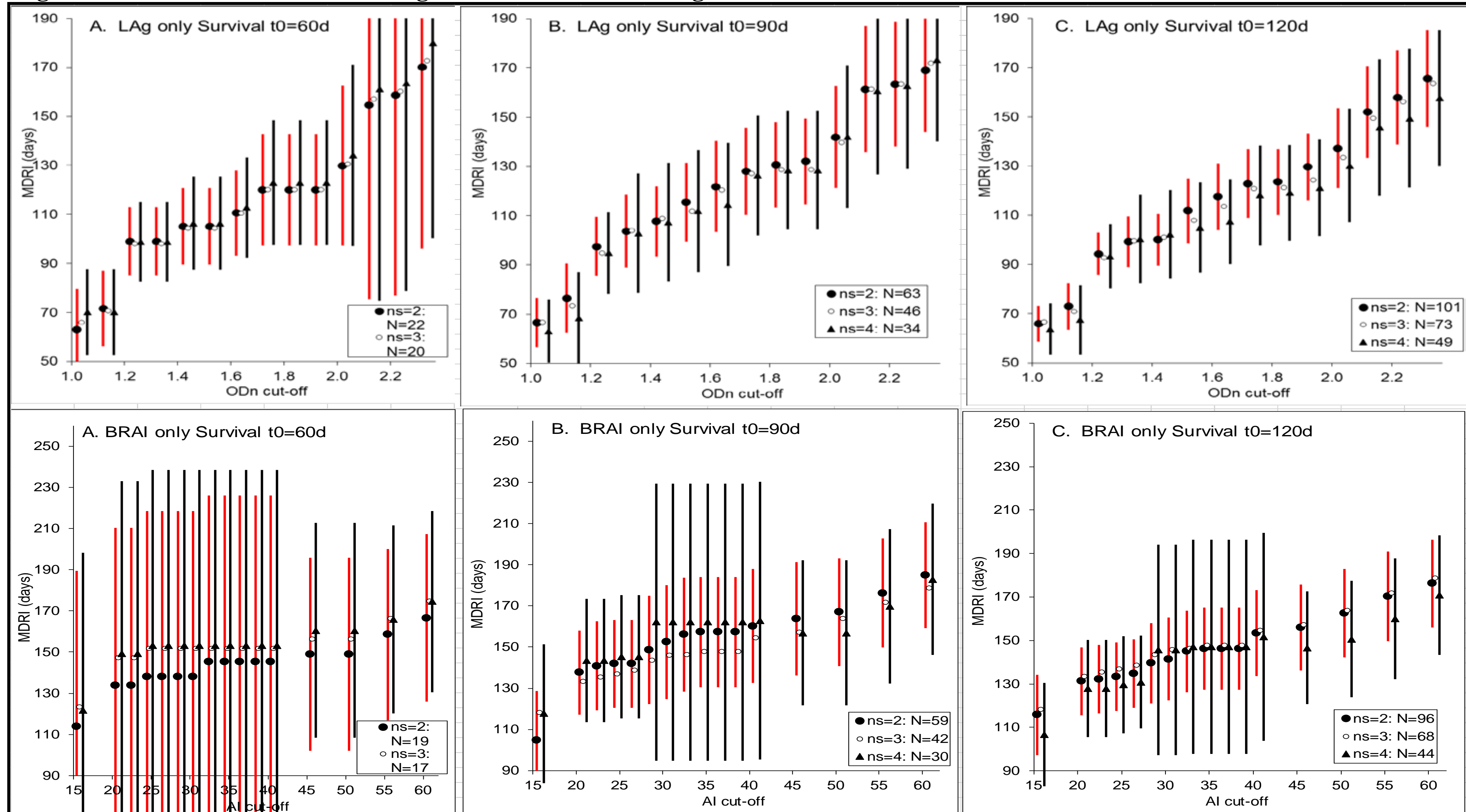
Stellenbosch University https://scholar.sun.ac.za

Figure 8-4: MDRI for BED at different cut-offs and varying t_0 using survival methods



Stellenbosch University https://scholar.sun.ac.za

Figure 8-5: MDRI for BRAI and LAg at different cut-off using Turnbull SA methods



Stellenbosch University https://scholar.sun.ac.za

Table 8-1: Summary MDRI for LAg and BRAI using Turnbull SA methods

ns=2		t_0=60	t_0=90	t_0=120
LAg	Cut-off= 1.5	105 (90-121) CV 7.69%	115 (100 -128) CV 6.4%	109 (98-125) CV 6.0%
	Cut-off =2.0	130 (97- 163) CV 12.8%	142 (121 - 163) CV 7.5%	136 (120 -152) CV 6.0%
BRAI	Cut-off 30%	120 (93 -147) CV 11.7%	144 (125-180) CV 7.6%	135 (120-151) CV 6.0%
	Cut-off 40%	127 (100- 153) CV 10.7%	153 (130-188) CV 7.7%	144 (128-160) CV 5.6%

With constant values of *ns*=2 and C = 1.5, reproducibility – as defined by the CV – decreased with increasing t_0; we saw the same trend for C = 2.0, but for each value of t_0 the CV was higher for C = 2.0 than for C = 1.5 (Table 8-1). For BRAI, similarly, the CV decreased with increasing t_0, so that and the most precise (narrow CI) MDRI estimate was achieved at t_0=120 for either cut-off 30% or 40%. There was very little difference, however, between the BRAI MDRIs estimated using C = 30% or 40%. The small increase of BRAI MDRI with C is consistent with the very rapid increase in AI in the early stage of infection that also increases with cut-off (Figure 8-5).

We note several issues with the SA analysis for the three assays. Firstly, CIs for these measurements are wide, with CVs always exceeding 5%. We note that the SA model is best designed to measure a single and continuous sojourn in a single state. For the biomarkers tested here, particularly BRAI, we observed multiple transitional states – i.e., cases were observed to leave and to re-enter the recent state. Re-entries into the recent state will result in under-estimation of the true MDRI using survival analysis. The real implications of this problem are investigated and discusses further in Sections 8.2.6a and Section 9.2.1.

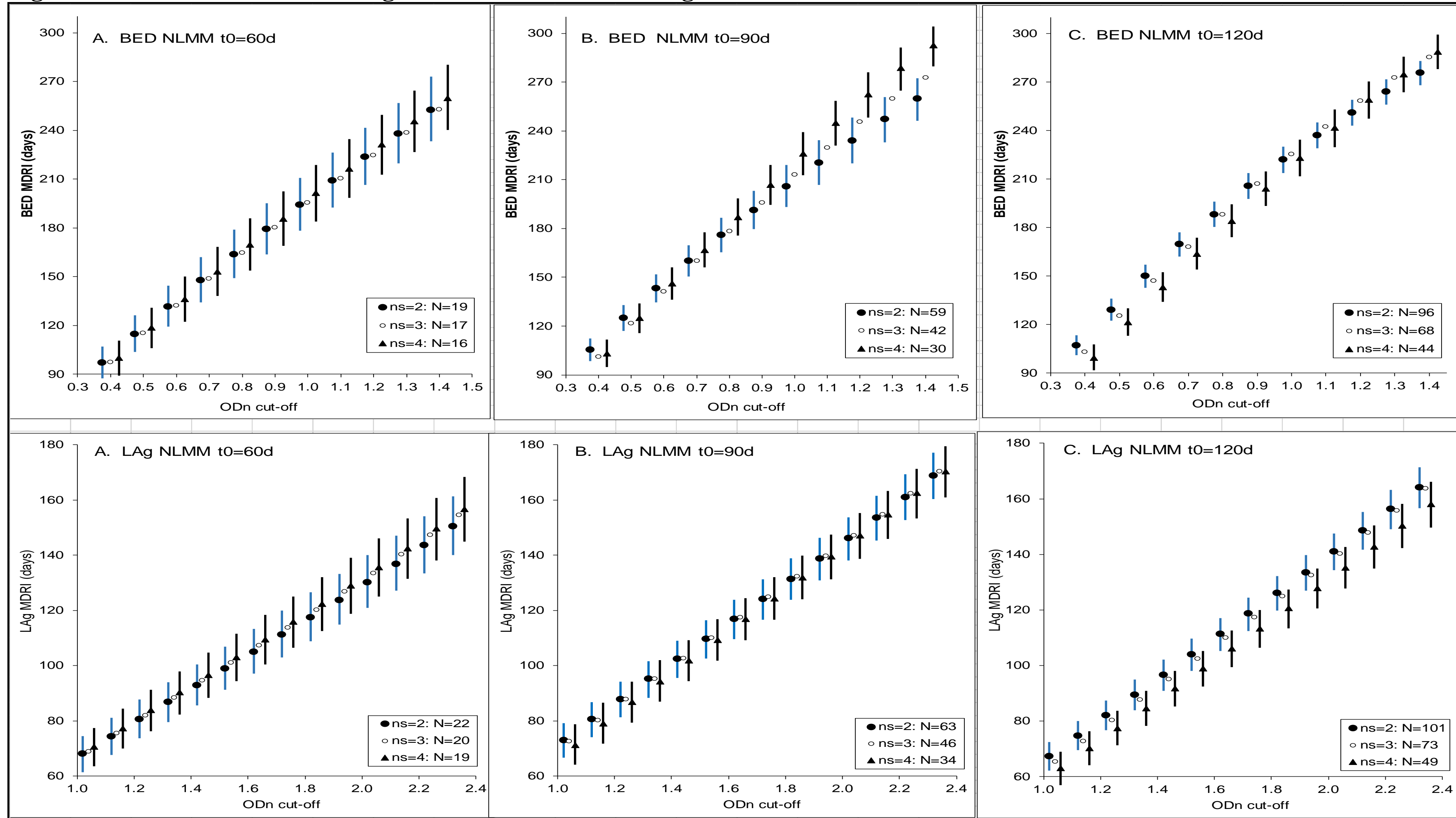
Stellenbosch University https://scholar.sun.ac.za

8.2.3 MDRI Using Non Linear Mixed Methods (NLMM)

The NLMM is not applicable to the BRAI method, since the avidity index (AI) does not increase according to a parametric form (Figure 8-3) and BRAI has a fixed upper-limit of 100%, unlike the BED and LAg. Accordingly, we estimated the MDRI for the BRAI method using survival analysis as described above (Section 8.2.2). We used NLMM to calculate MDRI for the BED and LAg assays (Figure 8-6). Using the NLMM method for both LAg and BED assays, the MDRI increased steadily with increase in cut-off, as expected (Figure 8- 6).

Stellenbosch University https://scholar.sun.ac.za

Figure 8-6: MDRI Estimates using NLMM for BED and LAg

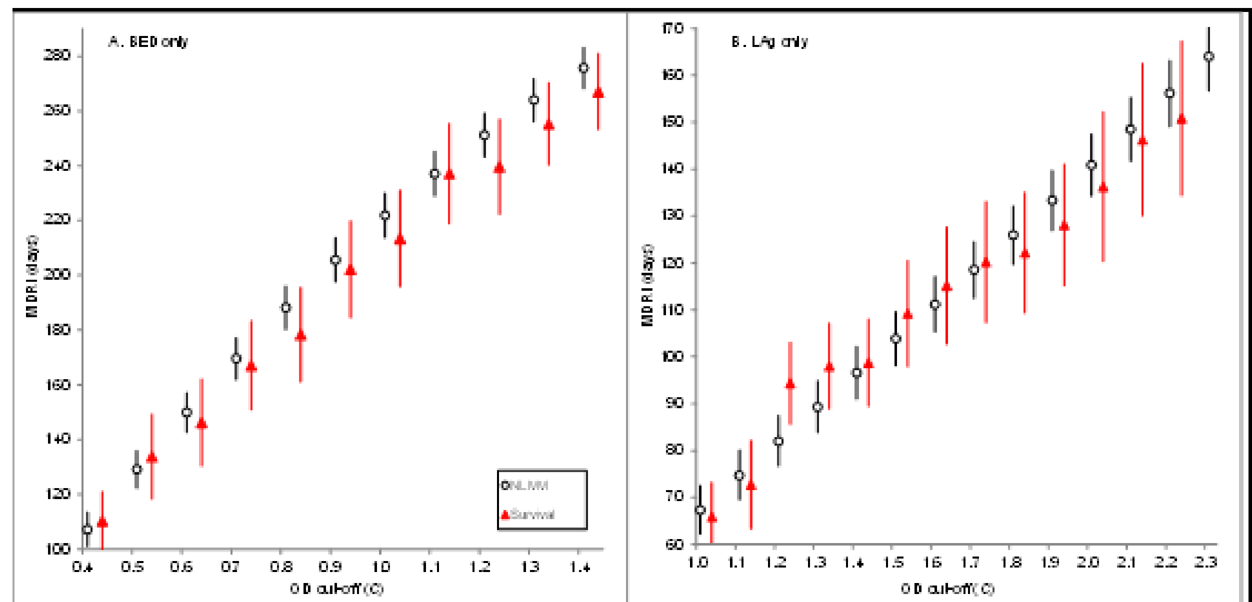


Stellenbosch University https://scholar.sun.ac.za

In the case of BED and LAg where we were able to use the SA and NLMM methods, we compared the estimates derived using each method by varying cut-off.

The MDRI estimates obtained using the NLMM had a lower variance than those obtained using SA for both LAg and BED (Figure 8-7). This is because the NLMM method utilises a well-defined mathematical function that provides a good fit to the data describing the increase in ODn values with time since seroconversion. Survival analysis makes no assumptions about the form of the increase with time.

Figure 8-7: Comparison of MDRI estimates for BED and LAg obtained using SA and NLMM



Assay	Survival Analysis	NLMM
LAg cut-off 1.5	109 (98-121) CV 6.0%	104 (98 -110) CV 2.9%
LAg cut-off 2.0	136 (120-152) CV 6.0%	141 (134-148) CV 3.4%
BED cut-off 0.8	178 (161 -196) CV 4.9%	188 (180-196) CV 2.1%

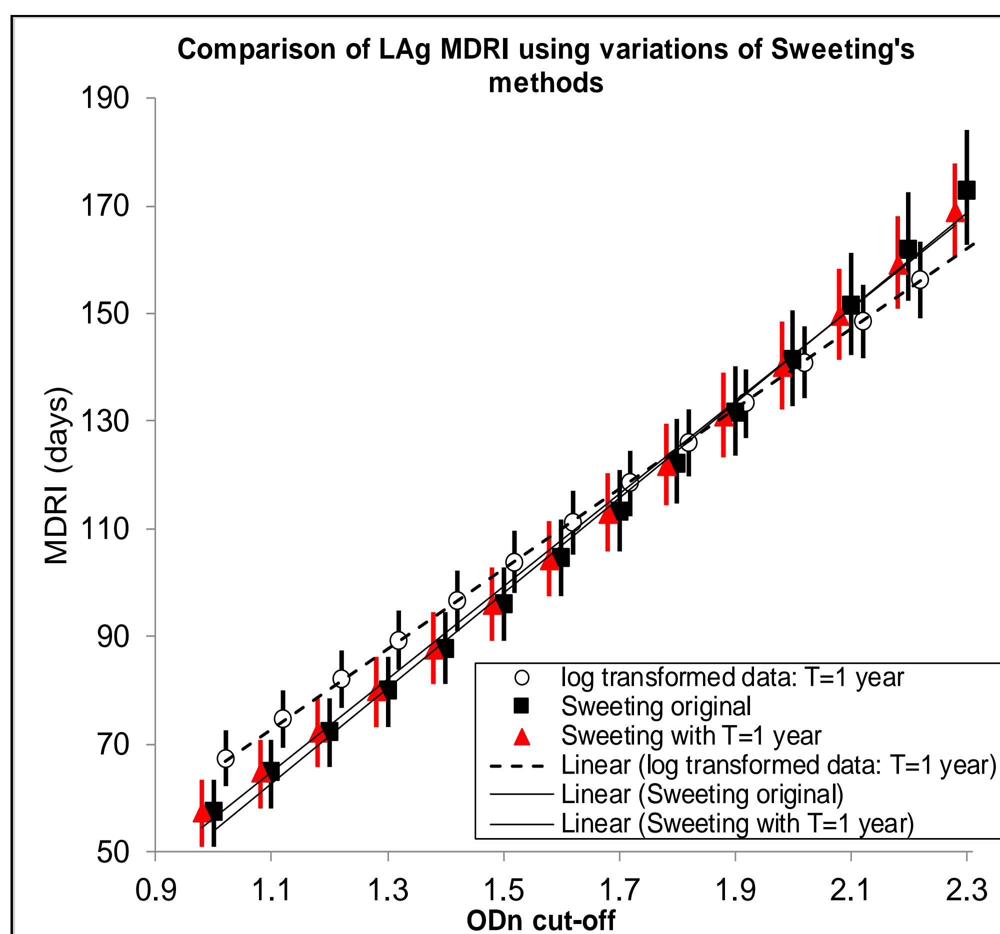
Our comparison of the NLMM method and SA is appropriate in that the former uses a Bayesian approach and so the interval we are quoting is a credible interval. This is denoted by the 2.5th and 97.5th percentiles of the posterior distribution of the mean MDRI. We did not make any formal comparisons between credible and confidence intervals. In this thesis, we were interested only in situations where, there are major

Stellenbosch University https://scholar.sun.ac.za

differences between estimates, such that there is no overlap of the 95% confidence/credible intervals. Alternatively, that the differences between the point estimates are small relative to the size of the confidence/credible intervals to the extent that the intervals concerned overlap the point estimate with which comparison is being made. Although we did not conduct the formal hypothesis testing, the visual representations of our comparisons were useful in highlighting situations where the means are markedly different from each other.

We compared the MDRI estimates arising from the use of the original code provided by Sweeting et al. (2010) with our modifications where we either: (i) introduced a time limit T for the maximum duration of recent infection; (ii) log-transformed the ODn data prior to analysis (Figure 8-8).

Figure 8-8: Comparison of MDRI for LAg using variations of Sweeting's Methods



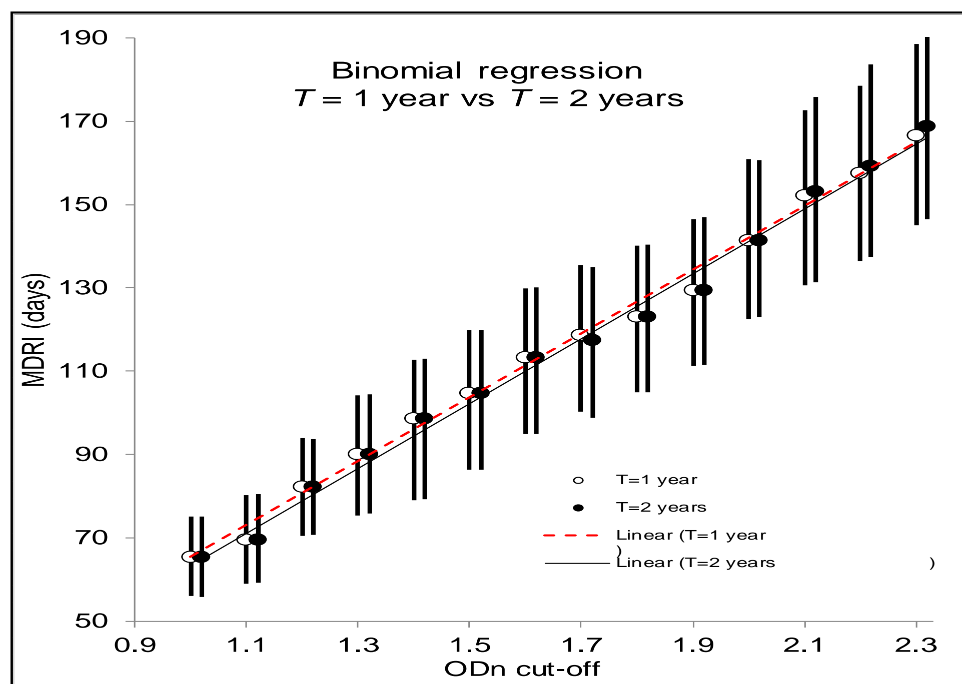
Stellenbosch University https://scholar.sun.ac.za

At a cut-off of 1.5 for the LAg assay, there was no difference in MDRI estimated using the modified Sweeting et al. (2010) [77] method or the NLMM. For BED and LAg assays, we noted that SA and NLMM methods give approximately the same answers when *T*=1 year. The only difference is that NLMM produces estimates with smaller variance than SA and this is similar to Sweeting's original method.

8.2.4 MDRI Estimates Calculated When *T*=1 or 2 years

We used linear binomial regression to estimate the LAg MDRI and also to test the hypothesis that extending the period *T* from 1 to 2 years can increase the MDRI, with the potential to improve the accuracy of the estimate.[20,62, 65,73] We found no difference in LAg MDRI whether we set *T* equal to 1 year or 2 years: the point estimates differed by <0.1% (Figure 8-9a). The reason for this result appears to be the consistently rapid increase in LAg ODn in the ZVITAMBO situation. Thus, for example in the data plotted in Figure 7-3C, there was no case observed where a sample taken at >1 year post-seroconversion had a LAg ODn < 1.5. Obviously, therefore, extending *T* beyond 1 year will make little difference to the MDRI.

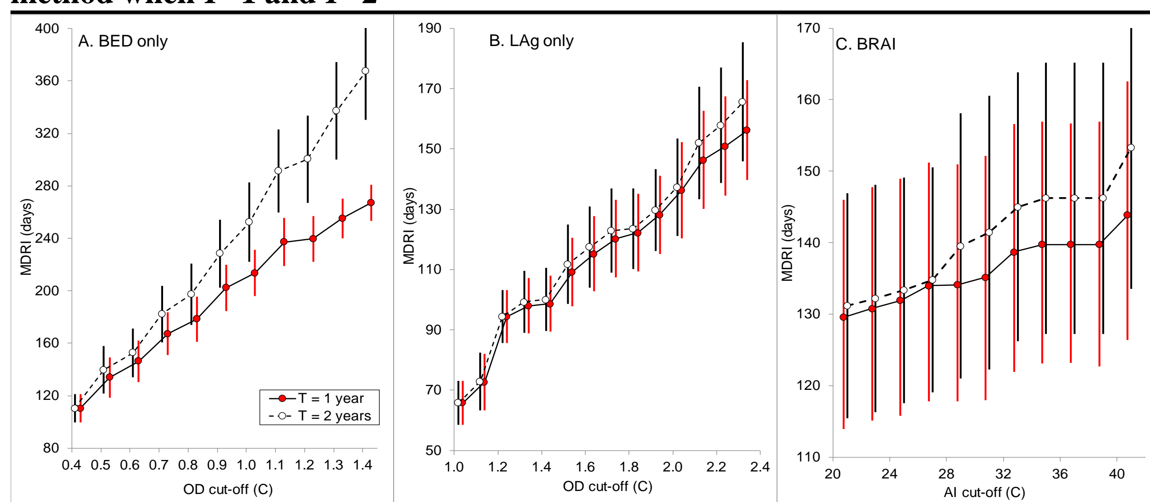
Figure 8-9a: Comparison of MDRI estimates for LAg when *T*=1 and *T*=2



Stellenbosch University https://scholar.sun.ac.za

We explored MDRI estimates for BED, LAg and BRAI obtained using survival analysis, with *T* taken either 1 or 2 years. For all analyses it was required that each case had a minimum of two HIV positive samples post-seroconversion and that the time between last negative and first positive HIV tests was at most 120 days. (Figure 8-9b).

Figure 8-9b: Comparison of MDRI estimates for BED, Lag and BRAI using SA method when *T*=1 and *T*=2



For BED and BRAI, MDRI point estimates increased more than for LAg when *T* was increased from 1 to 2 years, but the differences only attained statistical significant for BED for larger values of *C* than are likely to be used in practice. Accordingly, we took *T*=1 year for all further calculations of MDRI for the different assays with (Section 8.2.5).

Stellenbosch University https://scholar.sun.ac.za

8.2.5 Comparison of MDRI based on NLMM, SA, Binomial Regression and r/s for the three assays

We compared the MDRI obtained for all three serological biomarkers using binomial regression, survival analysis (SA) and the ratio *r*/*s*. For BED and LAg, but not for BRAI, we also used NLMM (Figure 8- 10).

Using NLMM, at cut-off values of 0.8 for BED and 1.5 for LAg gives MDRI estimates of 188 (95% CI 180 -196) days and 104 (95% CI 98-110) days, respectively (Figure 8- 10). The MDRI obtained using NLMM were lower than those obtained using binomial regression analysis by 6.3% for BED and 0.8 % for LAg at similar cut-offs. However, for both LAg and BED, the narrowest CI was always achieved when MDRI was calculated using the NLMM.

Based on Turnbull SA methods at cut-off 30% and 40% for BRAI, the estimated MDRIs were 135 (95% CI 120 -151) and 144 (95% CI 128 -160) days, respectively. (Figure 8- 10). In contrast, the estimates obtained using binomial regression analysis were lower by 6% and 7% at cut-off 30% and 40% respectively with wide confidence intervals for both estimates when T= 1 year.

It will be noted that, when either BED or LAg is used alone (*i.e*., without using the viral load) to identify recent infections, all of the different methods provide very similar MDRI estimates (Figure 8- 10). The low values of BED and LAg MDRI, habitually obtained using the ZVITAMBO data, do not appear to be due to some simple error in our statistical methods or their application.

For BRAI, however, whereas the estimates derived using either survival analysis or binomial regression are closely similar, the *r*/*s* estimates are consistently and markedly lower.

This comparative analysis of the different statistical methods using the original HIV serology data (Figure 8- 10). It is useful in highlighting for each assay, the method that produces the most precise MDRI estimate (based on low CV) which is NLMM for BED and LAg while for BRAI it was the SA method.

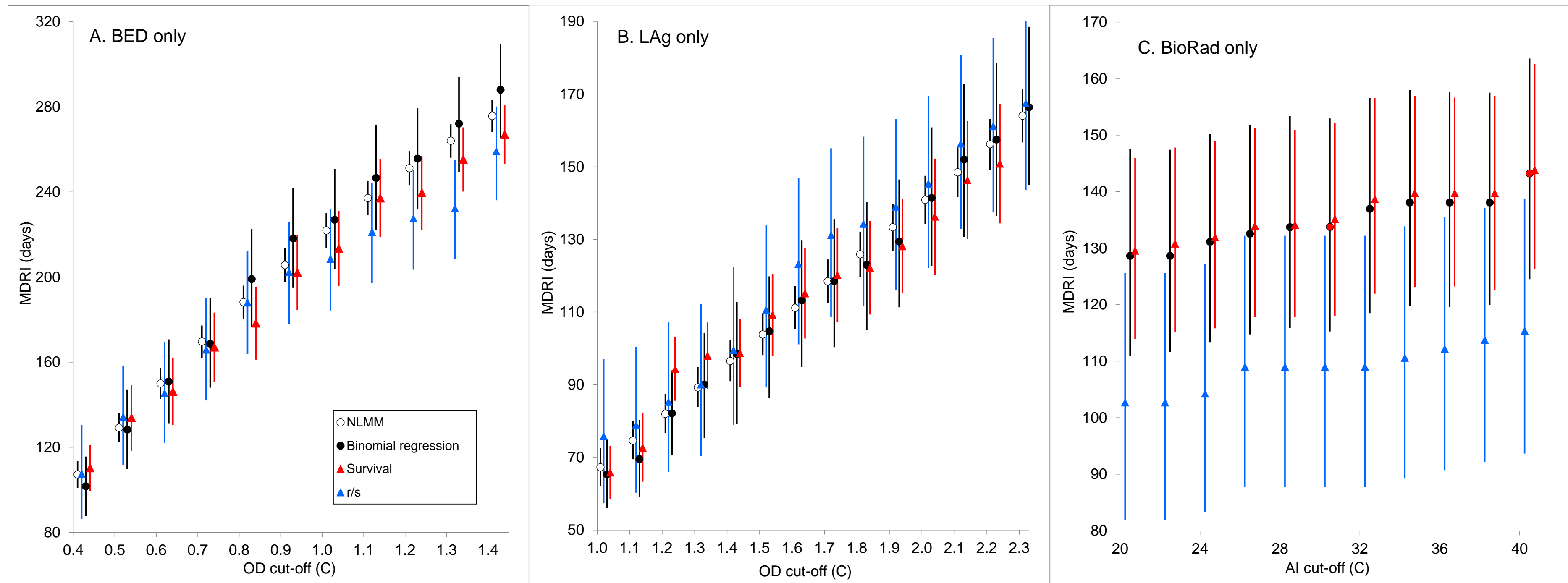
Stellenbosch University https://scholar.sun.ac.za

From this evaluation, we note that there are merits and demerits for different statistical methods in calculating MDRI. Duong et al. (2015) highlighted that MDRI calculations can be affected by the duration between last negative and first positive results, the collection interval and frequency of specimen collection as well as avidity kinetics.[69] Firstly, the NLMM is based on the Bayesian approach and so allows for use of all posterior measurement of individual readings after seroconversion. However, we could not apply the NLMM to BRAI, since the avidity index (AI) does not increase according to a parametric form and BRAI has a fixed upper-limit of 100%, unlike the BED and LAg. In contrast, the SA method is able to accommodate AI trend reversals observed with BRAI, thus making it a suitable method for MDRI calculation. Binomial regression models utilise all data points following seroconversion and can be applied to both LAg and BRAI. (Figure 8- 10). Hanson et al. (2016), reported that the major difference in MDRI calculated using SA (parametric or non-parametric) or binomial regression (parametric logit and parametric with random intercept) or non-parametric Generalized Additive Model (GAM) methods was in accuracy and precision.[65] However, they recommended that for BRAI and BED, the revised SA method that take into account the probability of remaining in the recent state as a function of time since seroconversion should be applied in calculation of annual incidence in United States of America.

As highlighted in methods section 6.6.7, there is generally no difference in the MDRI point estimate derived when we use the NLMM, Binomial and R/S methods (Figure 8- 10). In this case we rely on the CV to show differences in precision of methods used. The complete overlap of the 95% confidence intervals shows that there is no significant difference between the point estimates and so based on the CV, the only difference is their precision.

Stellenbosch University https://scholar.sun.ac.za

Figure 8-10: Comparison of MDRI based on NLMM, SA, Binomial Regression and the Ratio (R/S) for BED, LAg and BRAI [ns>=2; t_0 <=120 days; T = 1 year]



Method of calculating MDRI (95% CI; CV)	BED C=0.8	LAg C=1.5	LAg C=2.0	BRAI C=30%	BRAI C=40%
SA	178 (161 -196; 4.9%)	109 (98-121; 5.3%)	136 (120-152; 6.0%)	135 (120-151; 5.9%)	144 (128-160; 5.6%)
NLMM	188 (180-196; 2.1%)	104 (98 -110; 2.9%)	141 (134-148; 2.4%)		
Binomial	199 (176 -223; 6.1%)	105 (86-120; 8.2%)	141(123-161; 6.8%)	134 (115 -153; 7.2%)	143 (125 -164; 7.0%)
r/s	188 (164 – 212; 6.4%)	111 (89 – 134; 10.0%)	145 (122 – 170; 8.1%)	109 (88 -132; 10.1%)	115 (94 – 139; 9.7%)

Stellenbosch University https://scholar.sun.ac.za

Based on the above analysis, we do not know what the biases are: but what the results imply is that either: (i) If at least one of the methods provides approximately unbiased estimates then all of the methods provide approximately unbiased estimates. (ii) If all of the estimate are biased then they are biased to approximately the same extent.

Stellenbosch University https://scholar.sun.ac.za

8.2.6a Impact of adding VL to MDRI using Survival Analysis

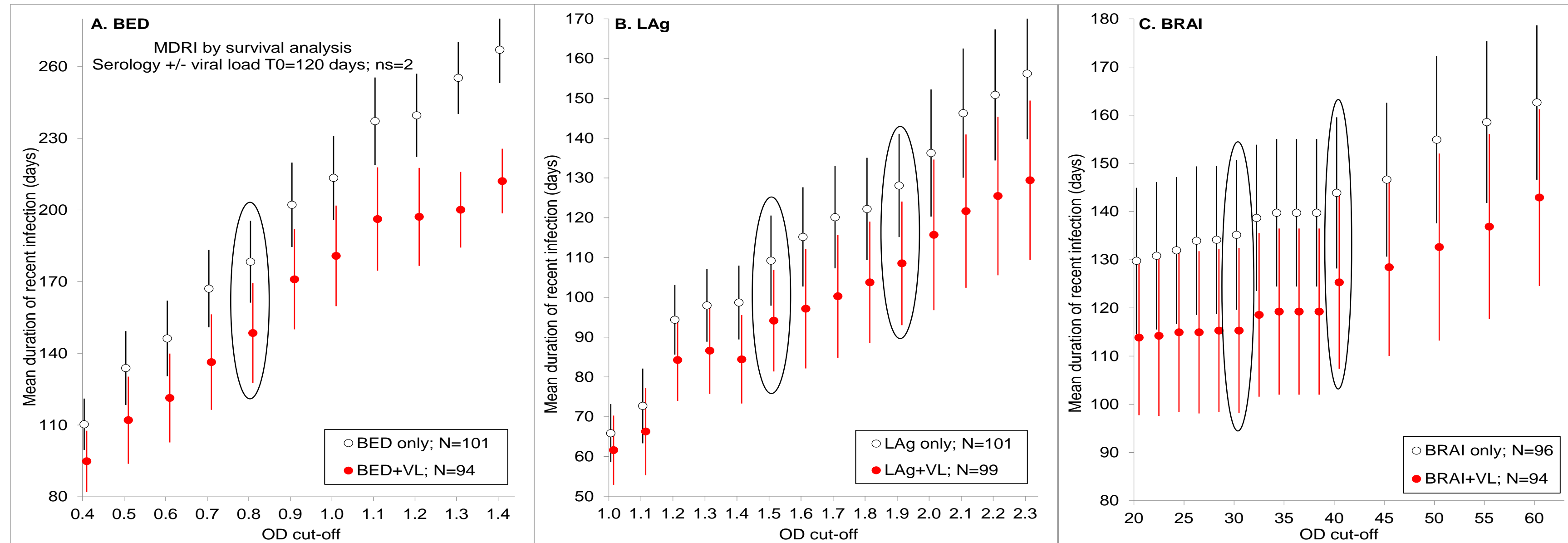
Recent studies have shown that viral load suppression in long-term infections is a strong confounder when serological assays are applied to long-term infections.[87,90,91] Results show that long-term infections return ODn/AI readings that mimic recent infection. Using VL as part of serological algorithms therefore helps in reducing misclassifications, and thus the FRR. This should thus result in improved HIV incidence measurements.

We included results of the VL assay in the algorithm, for determining MDRI using the SA methods. We performed this analysis using the ZVITAMBO original HIV serology classification. We initially define cases with VL < 1000 copies/ml as long-term infections and so they are not included in the calculation of MDRI. We show the absolute number of samples remaining in the calculation of MDRI using Turnbull SA and ns=2 and t_0=120 days for all three assays (Figure 8- 11).

The addition of VL resulted in the reduction of the absolute number of cases that are included in the analysis, because cases were screened, based on the level of VL. We show the number of samples (N) included in the analysis with or without VL (Figure 8- 11). The resulting MDRI for all three assays is lowered by addition of VL for all three assays.

Stellenbosch University https://scholar.sun.ac.za

Figure 8-11: MDRI for BED, LAg and BRAI and VL MAA using Turnbull's survival analysis (ns=2 and t_0=120)



Assay Recommended Cut-off	BED 0.8	Lag 1.5	Lag 2.0	BRAI 30%	BRAI* 40%
Assay only (95% CI) (CV; N)	178 (161-196) (4.9%; 96)	109 (98-121) (5.3%; 101)	136 (120-152) (6.0%; 101)	135 (120-151) (5.9%; 96)	144 (128-160) (5.6%; 96)
Assay and VL (95% CI) (CV; N)	149 (132-165) (5.7%; 94)	94 (82-106) (6.4%; 99)	116 (101-130) (6.5%; 99)	115 (102-129) (6.0%; 94)	125 (110-141) (6.3%; 94)

*Included based on Kassanjee et al. 2014

Stellenbosch University https://scholar.sun.ac.za

8.2.6b Comparison of MDRI derived for assay and assay with VL based on Binomial regression, Survival Analysis and R/S Method

We compared the MDRI obtained using a serological assay only and calculated using the Binomial Regression model, SA and R/S methods (Figure 8-12). We added the comparison of estimates obtained using an MAA in which we used VL in the initial screening to remove cases that were considered as long-term infections, if VL ≥ 1000 regardless of the biomarker level. We used three analytical methods to estimate the MDRI for BED and LAg, and three for BRAI (Figure 8 - 12).

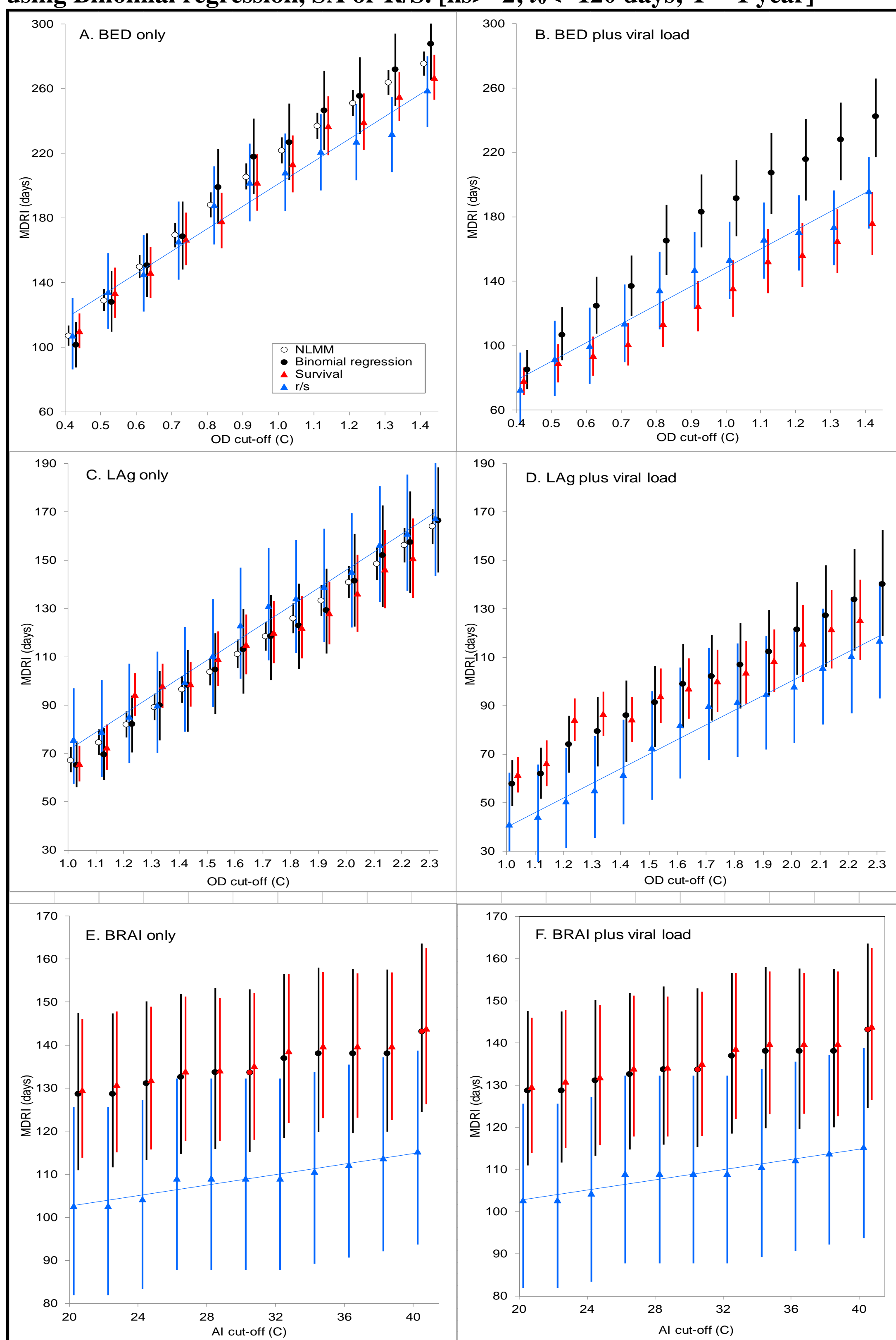
In our comparison of MDRI obtained using the serological biomarker assay only, we note that the three analytical methods provide estimates that are numerically similar for BED and LAg. The complete overlap of the 95% confidence intervals in each case shows that the estimates do not differ significantly at the 0.05 level of probability (Figure 8 - 12 A, C).

The same is true for BRAI, when we compare the estimates derived using survival analysis and those emerging from binomial regression. The *r*/*s* estimates, however, are consistently lower than those calculate using the other two methods (Figure 8 - 12E).

When both serology and VL are used to identify recent infections, the NLMM can no longer be used for any of the serological biomarkers and only three methods are compared. For LAg and BRAI the binomial regression and survival analysis estimates are closely similar, whereas the *r*/*s* estimates are consistently lower (Figure 8 - 12 B, D and F). In contrast, the survival analysis estimates for BED are lower than those from binomial regression with generally no overlap in the 95% confidence intervals. In this case the *r*/*s* estimates are intermediate between the estimates from the other two methods.

Stellenbosch University https://scholar.sun.ac.za

Figure 8 - 12: Comparison of MDRI for BED, LAg, BRAI with VL estimated using Binomial regression, SA or R/S. [ns>=2; t_0<=120 days; T = 1 year]



Stellenbosch University https://scholar.sun.ac.za

8.2.7 Exploration of low MDRI for BED, LAg and BRAI arising from ZVITAMBO postpartum cohort

The MDRI of an assay is a key input in the calculation of HIV incidence. Several evaluations have been conducted to determine the MDRI of laboratory assays including BED, LAg and BRAI as highlighted in Section 3.3 and Section 8.0. Researchers have estimated MDRIs using different statistical methods. [65, 69, 73] In most cases the analysis have provide slight differences in the MDRI obtained for each laboratory assay. [65, 69, 73]

In general, differences in MDRI have been attributed to HIV-subtypes, geographic location, host factors such as age, duration of HIV infection and to kinetics of the humoral response. The number of samples and interval of specimen collection also affect this parameter.

The ZVITAMBO cohort study samples (1997-200) used in this current evaluation were first used in Zimbabwe to characterise MDRI for the BED-capture EIA.[2] At OD cut-off = 0.8, the estimated BED MDRI was 196 (95% CI 188 - 204) days. The CEPHIA trial (2014) evaluated the BED, LAg and BRAI assays among others and obtained MDRI values that were much higher than those observed in current evaluation (Figure 8- 13).[73] Based on these observations, in Sections 8.1-8.2.7 we have provided an analysis of MDRI using several statistical methods with variations in T0 and ns. This analysis led to a further exploration of the data and resulted in the publication of a Short Communication by Hargrove et al. (2017).[92]

The Hargrove et al. (2017) paper tested the hypothesis that the low MDRI values obtained for LAg and BRAI in a population of postpartum women, compared to the general population, were due in part to differences in physiological state. Postpartum women exhibit a heightened antibody response compared to women in the general population. The analysis tested this by comparing MDRI values estimated using data for women who seroconverted at different times after giving birth. In making these calculations we used at T=2 years in order to match the published data.[92] We present the detailed sub-analysis in Sections 8.2.7a-b and Section 8.2.8.

Stellenbosch University https://scholar.sun.ac.za

8.2.7 a MDRI for BED, LAg and BRAI by stage of seroconversion

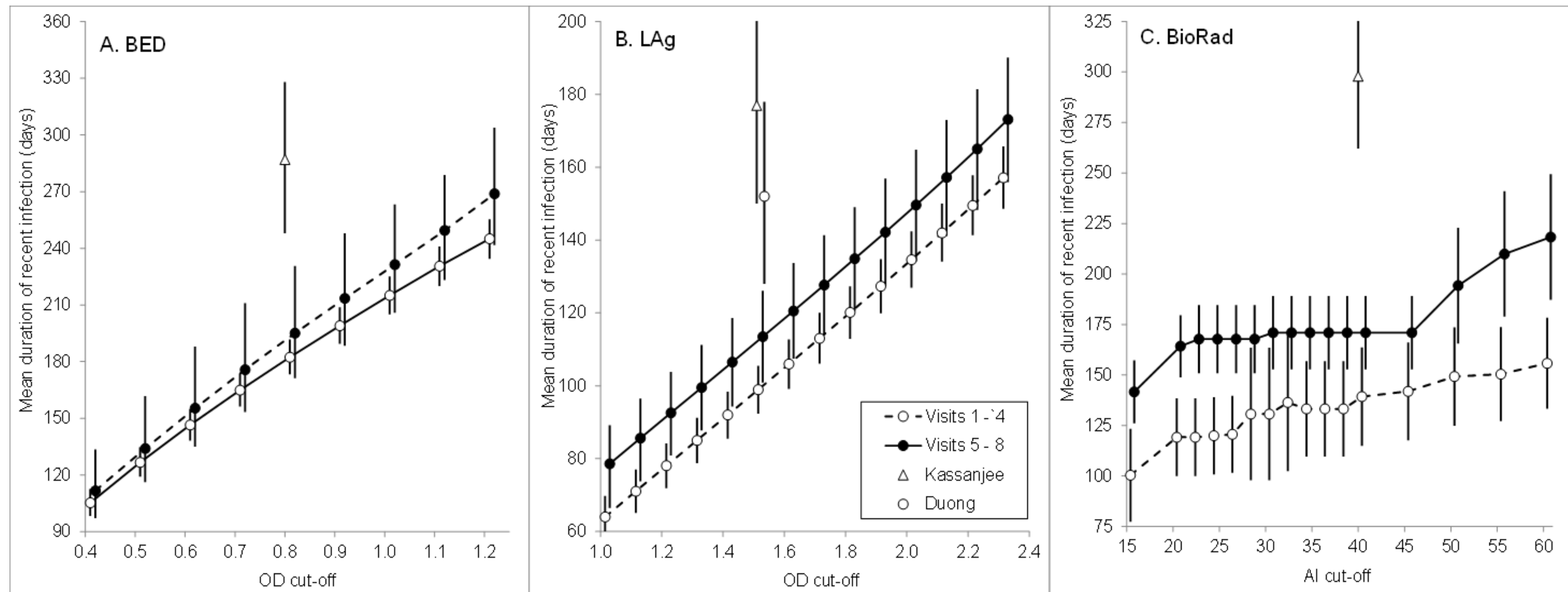
For the ZVITAMBO postpartum women, the analysis of MDRI was conducted for women who seroconverted in the first 9 months (Visits 1-4) compared to those seroconverting at 12 months and later. We used NLMM to estimate MDRI values for LAg and BED and Turnbull SA for BRAI: we used the recommended pre-set cut-off values for *C* (LAg=1.5, BED=0.8 and BRAI=40%).

In the initial analysis, when all seroconverting samples (*ns*=101) are included in the analysis of MDRI were 192 (180-201; *ns*=96) for BED, 104 (98-110; *ns*=101) for LAg and 144 (128-160; *ns*=96) for BRAI. These MDRI were 33%.32% and 52% lower than the 287(248-328), 152 (128-178) and 298 (262-338) days on respective assays of Clade C samples. [69, 73, 92]

When the data were analysed by stage of seroconversion, 0-9months and ≥12months, the MDRI obtained in the first group were 185 (176-195; ns=65) for BED, 99 (92-106; ns=70) for LAg and 139 (115-164; ns=65) days for BRAI. MDRI in >12months group were 195 (171-231; ns=31) for BED, 113 (101-126; ns=31) for LAg and 171 (153-189; ns=31) for BRAI (Figure 8 - 13).[92] We note that cases classified as later seroconverters does not mean that they were not seen for at least a year after infection. It means simply that the cases first tested HIV positive at, or after, Visit 5 (12 months postpartum). Criteria for inclusion of these cases in the analysis were still that the time between last HIV negative and first HIV positive tests could not exceed 120 days since they were supposed to be followed up every 3 months. Regardless of stage of seroconversion, all point estimates of MDRI were consistently lower than the published results for all three assays.

Stellenbosch University https://scholar.sun.ac.za

Figure 8-13: MDRI for BED, LAg, and BRAI by stage of seroconversion and compared to published results

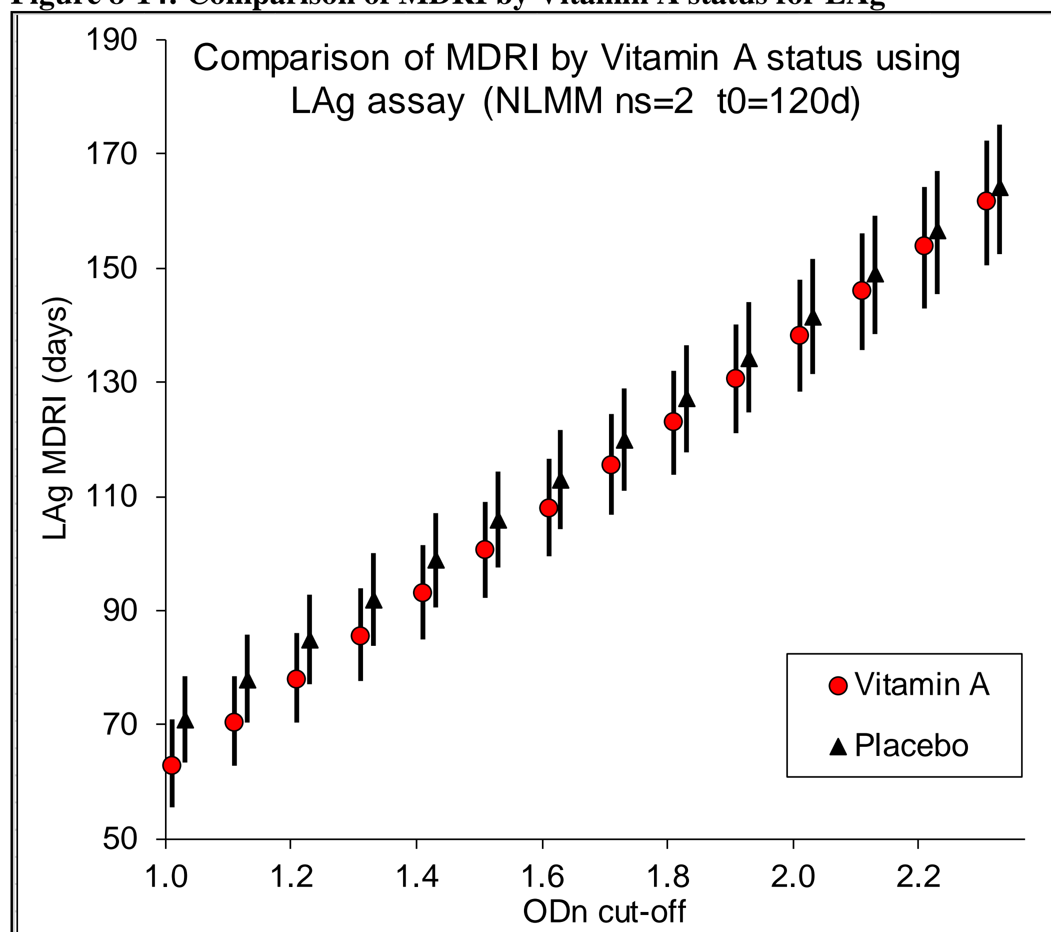


Stellenbosch University https://scholar.sun.ac.za

8.2.7b Comparison of MDRI by Vitamin A arm in original ZVITAMBO trial

Humphrey et al. (2006) reported no difference in rates of acquisition of HIV infection postpartum among women and their infants who received Vitamin A and those who received placebo.[25] We explored the impact of Vitamin A and placebo treatment on MDRI. In particular we needed to exclude the possibility that the very low MDRI values we estimate are an artefact of the vitamin A treatment. We explored this using analysis of LAg MDRI (Figure 8- 14).[92]

Figure 8-14: Comparison of MDRI by Vitamin A status for LAg



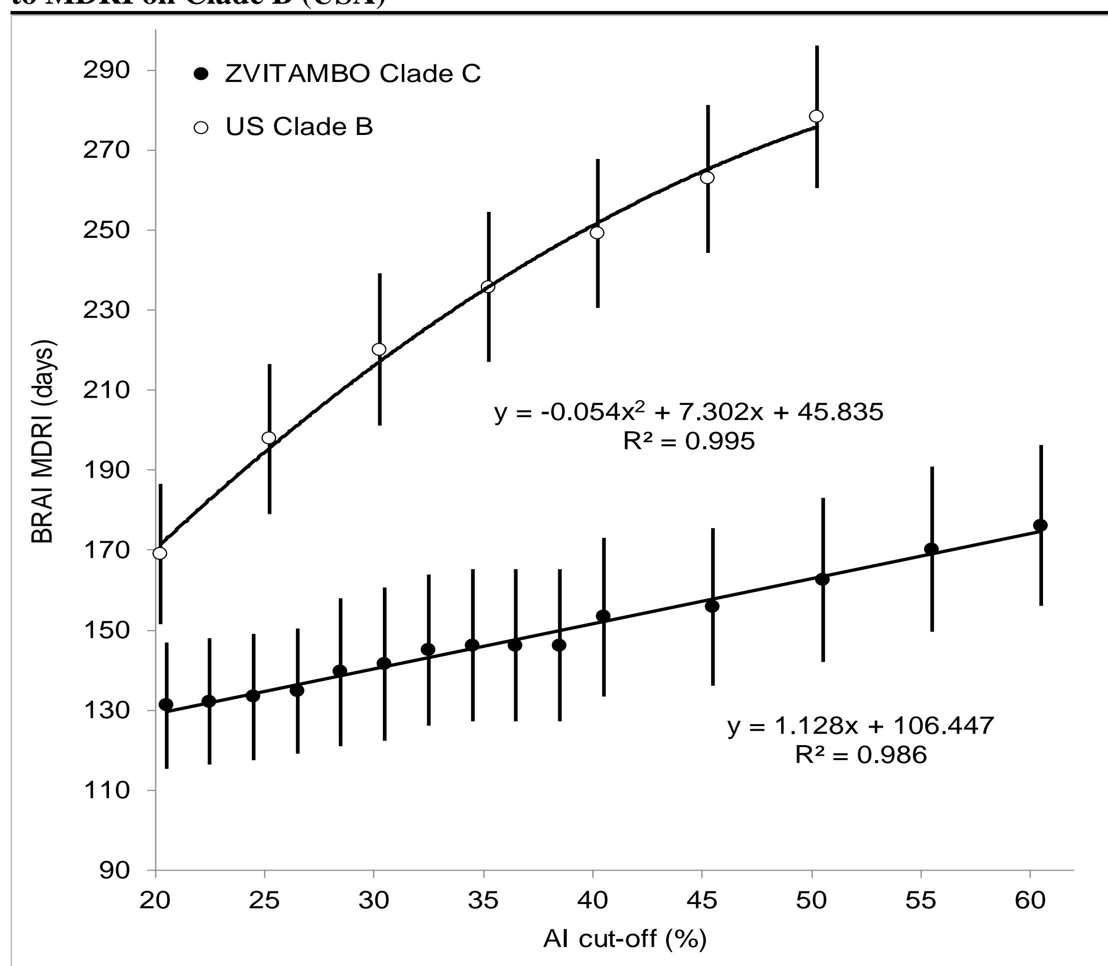
At each cut-off, there was no difference in MDRI between women who received Vitamin A and those who received placebo. Therefore it was not likely that the low MDRI obtained using LAg assay was in anyway associated with Vitamin A.

Stellenbosch University https://scholar.sun.ac.za

8.2.8 Comparison of MDRI for Clade C and Clade B using BRAI Assay

The BRAI assays has been evaluated using Clade B samples in America.[65] We compared our results for BRAI MDRI on Clade C to the analysis of BRAI MDRI on clade B.[65] Data and analyses were kindly provided by Debra Hanson (Figure 8- 15).[65]

Figure 8- 15: MDRI obtained using BRAI on Clade C (ZVITAMBO) compared to MDRI on Clade B (USA)



***Used with permission from Debra Hanson**[65]

The MDRI is markedly lower for the postpartum women with clade C, ZVITAMBO samples than for clade B samples from a general population in the USA. We point out that the differences in the Clade differences shown here could be attributable to the bias in postpartum specimens noted previously. At the recommended cut-off of 30%, the MDRI for clade B is 239 days compared to141days for ZVITAMBO, Clade C samples.

Stellenbosch University https://scholar.sun.ac.za

8.3 Discussion of results of MDRI for BED, LAg and BRAI in this evaluation

We found variations in MDRI estimates obtained using different statistical methods and when ODn and AI cut-off were varied for BED, LAg and BRAI.[21, 65 ,69, 73] These differences can be explained partly by the variations in assay kinetics in response to antibody maturation and growth. For BED and LAg, we found that there was no statistically significant difference between estimates derived using the NLMM and binomial regression methods, but the former estimates had lower CVs. Survival analysis estimates for BED and LAg had larger CVs. For BRAI, where we could not use NLMM, there was no statistically significant difference between estimates derived using survival analysis and binomial regression methods

The recalibration of MDRI by Duong et al. (2015) showed that, for the same cut-off, there were no significant differences in MDRI obtained for the LAg assay by using different analytical methods.[69] Our evaluation showed, likewise, that four different analytical techniques produced closely similar estimates over a large range of C (Figure 8-10).

For the NLMM method, the MDRI at cut-off 2.0 was 141 days (134-148), significantly higher than the MDRI at C = 1.5 of 104 (98-110). This suggests the possibly that a cut-off of C = 2.0 could be more appealing, provided the FRR remain low as was found to be the case in this evaluation (Section 9.0).

Although, the *r*/*s* method makes use of samples that are HIV negative at baseline thus precluding any long-term false serology positive samples, the high CV across MDRI for all assays estimates warrants a cautious approach in their use. In practical terms using the *r*/*s* method requires that all cases be HIV negative at the beginning and seroconvert during the follow-up period. In a cross-sectional survey set-up, this is not feasible, therefore this calculation is only useful in showing an alternative and rudimentary calculation of MDRI.

Other noteworthy findings include the finding that, for the LAg assay, the MDRI did not differ between the women who received Vitamin A and those on given a placebo. This supports the finding that Vitamin A did not alter the immune composition of the

Stellenbosch University https://scholar.sun.ac.za

postpartum cohort.[25,26]

A major finding in this evaluation is the fact that the MDRI for LAg and BRAI were much lower than those published of 287 days (95% CI 248 -328) for BED), 177 days (95% CI 150 -206) for LAg and 298 days (95% CI 262 -338) for BRAI applied to clade C samples.[73] A recalibration of incidence data by Duong et al. (2015) suggested that at the optimum cut-off for LAg of 1.5, the associated MDRI was 152 (128-178) days.[69] A recent evaluation of GS HIV Combo Ag/Ab (BRAI, Redmond, WA) using Clade B samples found that MDRI ranged from 50.6 days at AI 20% to 275.6 days for AI of 90%. [91] The developers recommended that BRAI assay cut-off be raised to 50% to provide an MDRI of 135 days.[91] If we considered an AI of 50% in this evaluation, the BRAI MDRI would be 163 (95% CI 142 - 183), and would still fall below the 298 days established in the CEPHIA study.[73] Similarly, for LAg, using a cut-off of 2.0 would result in an MDRI of 141 (134-148) days for NLMM methods and much lower, 137.0 (95% CI 121.1 -153.1) when SA methods are applied. This analysis clearly shows that our low values of MDRI are not attributable to the statistical methods that were used and therefore we need to find a plausible explanation as to why these discrepancies exist.

We suggest that the low MDRI estimates obtained using the ZVITAMBO samples are result from the fact that these samples are from an immunologically unique group of people. Studies have highlighted that there are immune-suppressive processes that occur during pregnancy that are required to maintain the pregnancy: there is then immune reconstitution in the postpartum period.[93,94] We were able to show that when the analysis was stratified by stage of follow-up, the MDRI values were lower for women seroconverting during the first nine months of follow-up compared to those who seroconvert at later times postpartum. These findings are consistent with work showing that the postpartum period is characterised by heightened and activated innate and specific immune defences. Furthermore, when the postpartum period is compared with non-pregnant women, these responses are elevated in women who are breastfeeding, rather than for those women with formula-fed infants [93] In the ZVITAMBO situation, where 99.1%, 94.0%, and 59.1% of the women breastfed their babies for at least 6, 12, and 18 months postpartum, respectively, it is therefore

Stellenbosch University https://scholar.sun.ac.za

expected that we find decreased MDRI values for women seroconverting during the first year, relative to the second, year postpartum.[95,95]

Given that pregnant women exhibit a state of relative immune-suppression, relative to non-pregnant women, thereby increasing the tolerance to foetal antigens, it is plausible that MDRI values appropriate for non-pregnant women might be even higher than the values we estimate for the second year postpartum period.[94, 96] While in this current evaluation, we acknowledge that the differences in MDRI were not significant at the 5% level of probability, these results may be limited by the limited numbers of samples seroconverting in the period >9 months postpartum, thus limiting the power of statistical analysis. Further evaluations of impact of gender and immune response on MDRI are critical for ongoing developments of laboratory assays in HIV incidence surveillance.

MDRI estimates will vary according to the estimated timing of HIV infection, which depends on the HIV test system employed. Kassanjee et al. (2014) used a Western blot to define infection, while the ZVITAMBO study employed two independent Elisa tests, with a Western blot tiebreaker in cases where Elisa results were discordant. Studies have shown, however, that the difference between the time of infection estimated using the two diagnostic systems differ by only about 5 days.[96] It is therefore unlikely that this small difference accounts for the large differences in the published MDRI and those obtained in this evaluation.

The low MDRI estimates in this cohort of postpartum women reinforce the need to establish values appropriate for any given population. Since incidence estimates change linearly with the inverse of the MDRI, application of inappropriate MDRI estimates will inevitably lead to errors in incidence estimates. Theoretically, because of the immune suppressed status of pregnant women, MDRI values might higher in pregnant than in postpartum women – but there is no information on the MDRI values appropriate for pregnant women.[95] In general, it is clearly important to obtain MDRI values appropriate for any given study population. Given the importance of antenatal clinic samples in monitoring changes in HIV prevalence and incidence, it is particularly important that we accurately estimate MDRI values appropriate for the analysis of samples emanating from pregnant women. Most importantly, this

Stellenbosch University https://scholar.sun.ac.za

evaluation supports the fact that MDRI not only varies by subtype but also varies by physiological status of participants even when infected with the same clade virus.

Researchers have advocated that an MDRI should be long enough to allow the counting of all recent cases and this was proposed to be at least six months to one year.[20] The values obtained for LAg or BRAI fall below this desired optimum. This evaluation and other studies either points to the need to re-evaluate the utility of this arbitrary value or that the assays need further refinement to meet this criterion.[69,73]

Finally in our current study the higher values of MDRI obtained using SA for BRAI (135 days) should favour this assay over LAg (104 days), but of concern is the higher coefficient of variation round MDRI for BRAI compared to that of LAg. While it is desirable to have a high MDRI value, the best value is one where there is a balance between a sufficiently large MDRI and an accompanying low FRR.

Stellenbosch University https://scholar.sun.ac.za

9.0 False Recency Rate (FRR) of HIV Infection

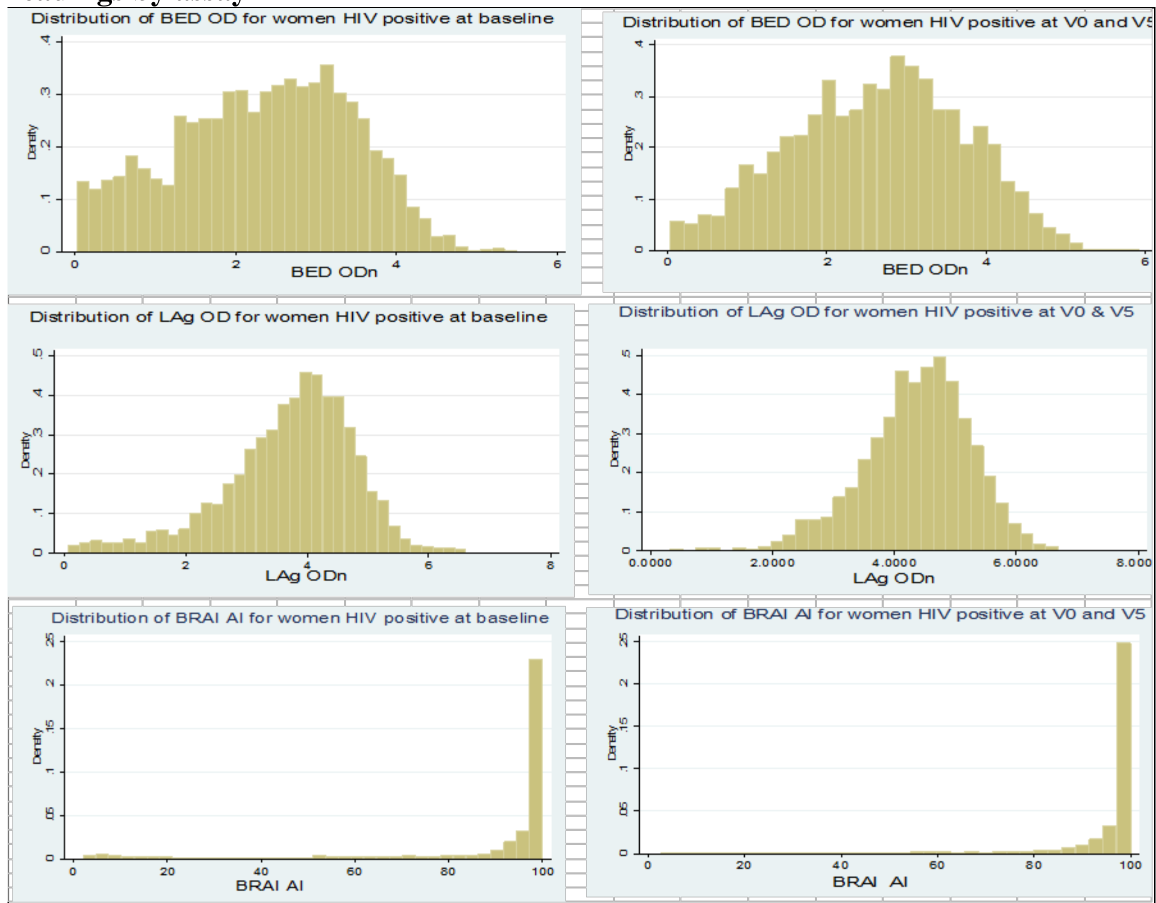
We retrieved and tested 2825 samples using LAg and 2824 using BRAI assay in the evaluation of FRR. Those women who tested HIV positive at enrolment and at 12 months in the ZVITAMBO trial study provided a good opportunity to measure FRR. However, because of the 18 samples that appeared to show a different HIV serology, we explored analysis using both the original data and the reclassified "New" data.

9.1 Distribution of ODn readings for BED and Lag and AI for BRAI

For each of the three assays, we plotted the distribution of ODn/AI values for women who were HIV positive at baseline. We also plotted these values for the subset of these women who were then also tested a period of 1 year later (Figure 9 1). This analysis of distribution patterns allowed us to see the distribution patterns of OD/AI readings for the same sample of women tested by the three different assays. Although this is not of great analytical value, the analysis enables us to see a crude picture of frequency and distribution of the readings that can be related to assay performance. Both for the women who were HIV positive at baseline and for those who seroconverted during follow-up, BED had the widest spread (distribution) of ODn readings. In comparison, the LAg had a symmetrical (normal) distribution, while the AI readings for BRAI clustered at the tail end of the graph. The BRAI pattern characterises the assay's kinetics whereby avidity quickly reaches the maximum 100% for with increase in time after infection.

Stellenbosch University https://scholar.sun.ac.za

Figure 9-1: Distribution of optical density (BED, LAg) or avidity index (BRAI) readings by assay



	BED	LAg	BRAI
Mean OD/AI for Women HIV positive at V0 (95% CI)	2.353 (n=2,735) (2.321 - 2.368)	3.724 (n=4,468) (3.692 - 3.755)	88.12% (n=4,463) (87.39 -88.84)
Mean OD/AI for Women HIV positive at V0 &V5 (95% CI)	2.635 (n=2,735) (2.594 -2.676)	4.371 (n=2,823) (4.338 - 4.404)	94.98% (n=2,814) (94.51 - 95.45)

Stellenbosch University https://scholar.sun.ac.za

9.2 FRR for BED, LAg and BRAI when Original and New Data are used with or without viral load

We define the False Recent Rate (FRR) for any test for recency as the proportion of cases that test as recent when we know that they have actually been HIV positive for some pre-defined (extended) time *T*: for analysis of the ZVITAMBO Trial data we defined *T* = 1 year. To estimate the FRR we used samples from women who tested HIV positive at baseline and then at 12-months postpartum. We carried out these analyses for all three assays using both the original HIV serology classifications (the "Original data-set") and the "New" dataset based on the reclassifications detailed in Section 7.2. We calculated FRRs for recency tests based only on the serological biomarkers, and then on tests where each of three biomarkers was used in conjunction with viral load.

9.2.1 Viral loads for cases testing HIV positive at Baseline and Visit 5

We also used viral load (VL) results in an MAA to screen samples before applying the laboratory assay and calculating FRR. When all samples with VL <1000 were classified as long-term, we calculated the proportion that remained recent for each assay.

Figure 9 - 2A and B show the distributions of log(10) viral loads for samples testing HIV positive at Baseline, and for those testing positive at Visit 5, that had tested HIV positive at Baseline 12-months previously. The peaks at the origin indicate cases with undetectable viral load. For all cases where virus was detectable, the means for the log(10) viral load were: Baseline, 4.11 (95%CI; 2.79 – 5.44, N = 4100); Visit 5, 4.23 (95% CI; 2.93 – 5.43; N = 2659). Given the complete overlap of the 95% confidence intervals for the means of the two distributions, there is clearly no statistically significant difference between the means.

Stellenbosch University https://scholar.sun.ac.za

Figure 9-2A: Distributions of viral loads for samples taken at A. Baseline, for cases testing HIV testing positive at that time

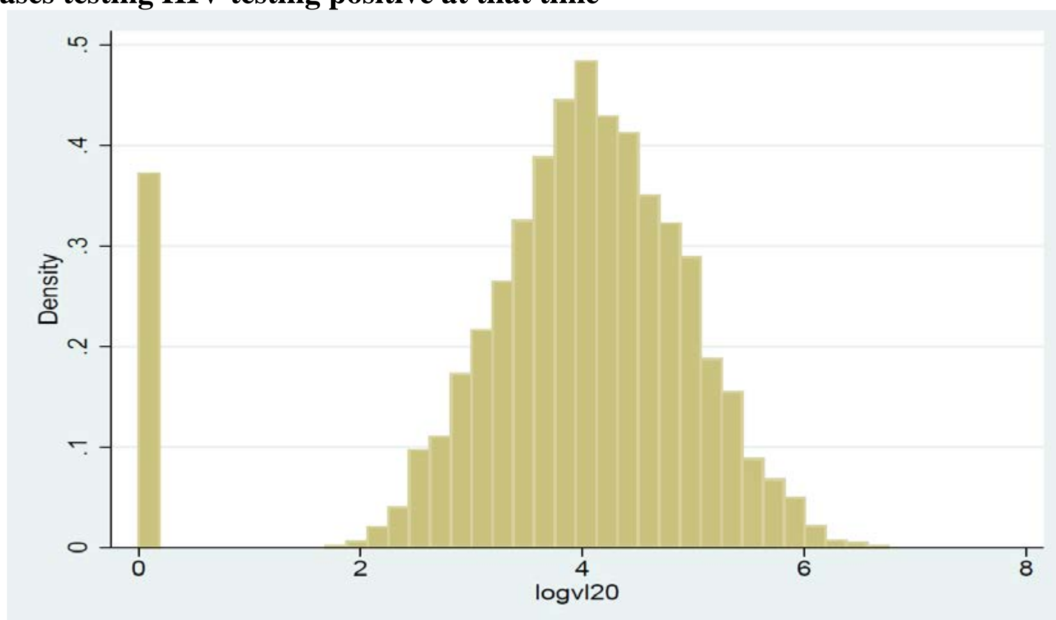
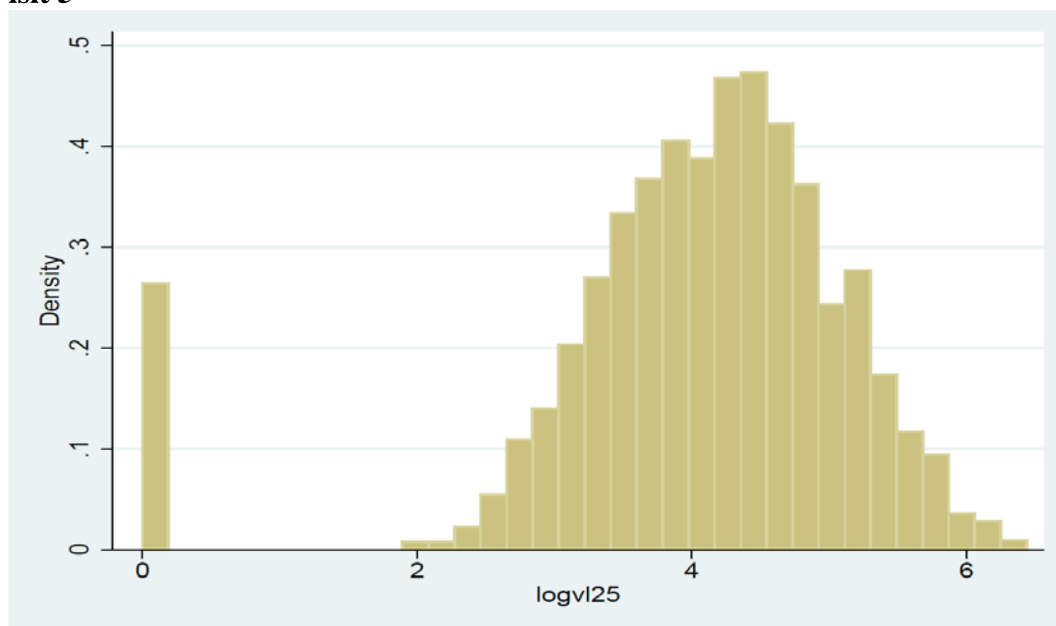


Figure 9-2B: Visit 5 for cases testing HIV testing positive both at Baseline and at Visit 5



Note: The difference in scales on the two X axes. The peaks at the origin indicate cases with undetectable viral load.

Stellenbosch University https://scholar.sun.ac.za

Table 9 -1 provides the numbers of samples testing “recent” or “long-term” as a function of log(10) viral load for cases who (A) tested HIV at Baseline or (B) tested HIV positive at both Baseline and Visit 5.

Table 9-1: Distribution of cases testing as “recent” or “long term” infections (using LAg with a cut-off of 1.5) as function of log(10) viral load

		LAg infection status			
		Recent	Long Term	Missing	Total
Log(10) Viral load	2	**39**	271	2	312
	3	13	352	0	365
	4	51	1437	1	1489
	5	66	1603	3	1672
	6	30	509	1	540
	7	3	27	0	30
	Missing	**6**	76	5	87
	Total	208	4275	12	4495

Note: - 1. Values denote the upper limit of each interval
2. Baseline samples: restricted to cases testing HIV positive at baseline.

We used the latter samples in the calculation of the FRR. Notice that, in principle, the 44 missing viral load values could mean a much-reduced sample size when we use both viral load and LAg serology to diagnose false-recent cases at Visit 5. In fact, however, the problem is mainly a theoretical one and does not pose a real practical problem. Thus, we define as “long-term” any case that tests as such by LAg serology, regardless of its viral load or even if no viral load is available. Problems could thus only arise when a samples tests “recent” by LAg serology. For those samples, only, we need to know whether the viral load is greater, or less than, 1000.

Stellenbosch University https://scholar.sun.ac.za

From Table 9-2 we see, however, that the above problem does not arise in practice. In all cases where the viral load was missing, the case either tested "long-term" by LAg, or was anyway also missing the LAg assay. Thus, in this example, when LAg alone is used to identify recent infections, a total of 16 cases test recent: when viral load is used as well, the number is reduced only by 5 – those that test recent by LAg, but where log(10) viral load < 3, i.e., viral load < 1000.

Table 9-2: Distribution of cases testing as "recent" or "long term" infections (using LAg with a cut-off of 1.5) as function of log(10) viral load

		LAg infection status			
		Recent	Long Term	Missing	Total
Log(10) Viral load	2	**5**	136	1	142
	3	1	161	0	162
	4	4	872	5	881
	5	3	1140	1	1144
	6	3	441	1	445
	7	0	25	0	25
	Missing	**0**	34	10	44
	Total	16	2809	12	2843

Note: 1.Values denote the upper limit of each interval

2. Baseline samples: restricted to cases testing HIV positive at baseline and V5

Highlighting:

Blue: Cases where missing viral load means we cannot decide if a case is "recent" or "long term".

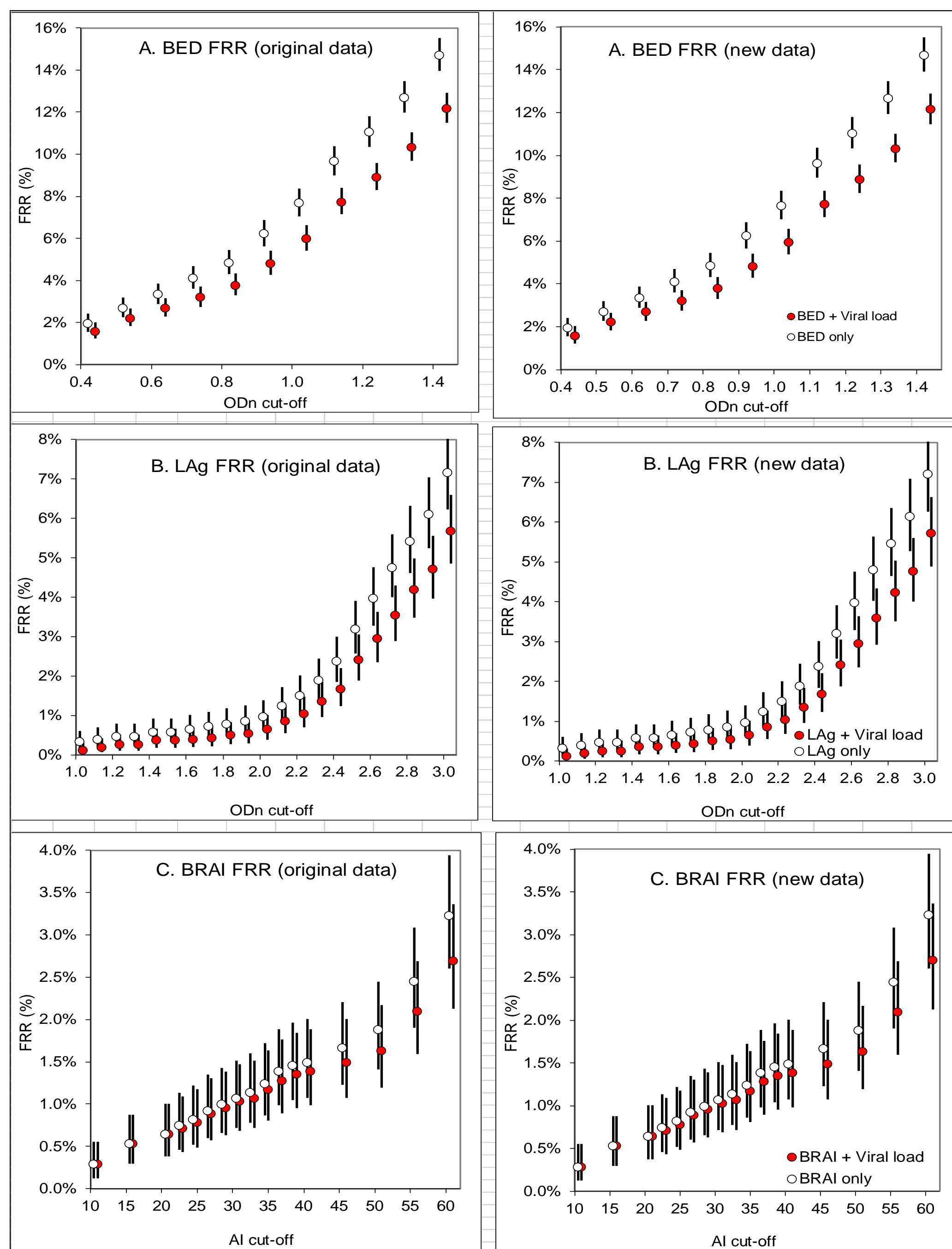
Green: Cases testing "recent" by LAg, but "long term" by viral load.

We plotted comparative analyses of the FRR results obtained using the "original' vs "new" HIV serology classification and serological assay only vs serological assay + VL at different cut-offs (Figure 9-2). Our results show minimal variation in the FRR when we compare results derived using Original or New data. This lack of any visible difference is likely due to the fact that we only changed the HIV serology status in 17 cases.

When we used VL in conjunction with a serology biomarker to assess recency of HIV infection, FRR values were numerically lower for each of the three biomarkers tested. The reduction was more marked for BED, but very much smaller for LAg and almost negligible for BRAI. This decrease in FRR is to be expected since a case is now only classified as "recent" if it has both ODn/AI < C and VL>1000.

Stellenbosch University https://scholar.sun.ac.za

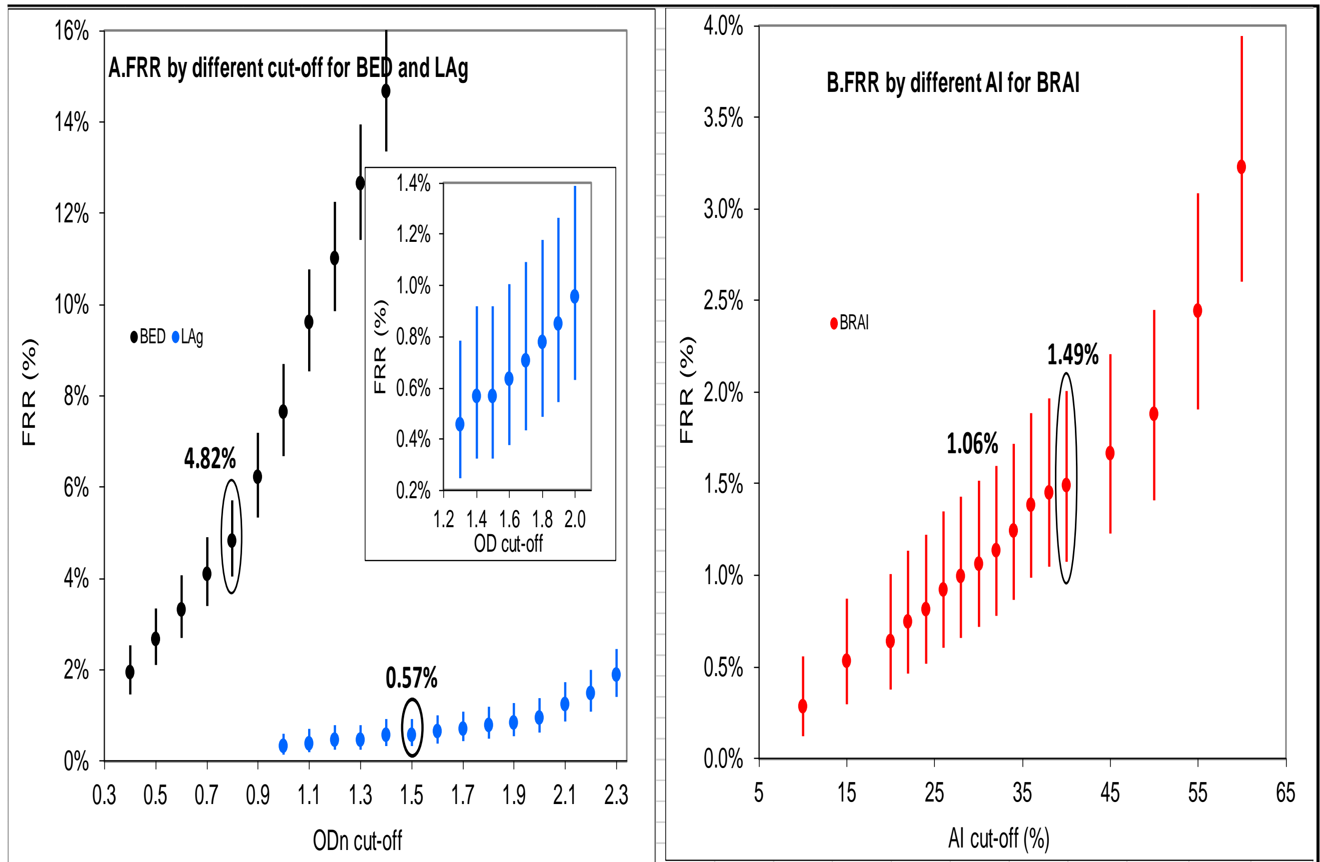
Figure 9-3: FRR by different cut-offs for BED, LAg and BRAI assay either used alone or in combination with VL



Stellenbosch University https://scholar.sun.ac.za

The FRR increases monotonically with cut-off, doing so much more rapidly for BED than LAg or BRAI. The FRR for LAg is much lower than for the BED for all $C < 2.0$. However, comparisons of FRR to cut-off are complicated by the fact that MDRI also increases with cut-off and comparisons between the FRR for LAg and BRAI and are also complicated by the differences in the ranges on the C axis (Figure 9- 4).

Figure 9-4: Summary of FRR by cut-off for BED, LAg and BRAI

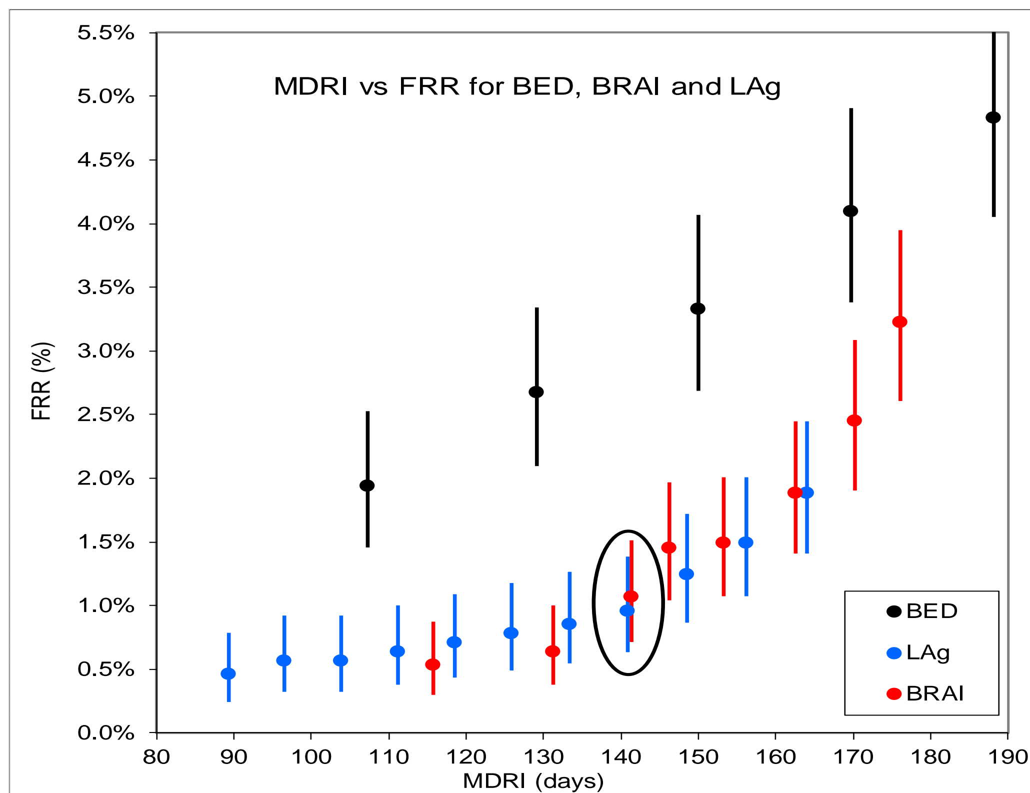


By varying the assay cut-off, we increase the threshold for classifying cases as recent and, as expected, the FRR increases with increasing cut-off (Figure 9 - 4).

Stellenbosch University https://scholar.sun.ac.za

The MDRI and FRR are inter-related indicators that characterise the performance of an assay. In characterising the laboratory assay's ability to measure HIV incidence with increased accuracy, researchers have concluded that the MDRI must be sufficiently large to capture all truly recent cases, while the FRR must remain small.[20] To overcome the problem of the differences in axis for BED, LAg and BRAI, we plotted the FRR against MDRI for the three assays (Figure 9- 5).

Figure 9-5: Summary FRR vs MDRI for BED, LAg and BRAI



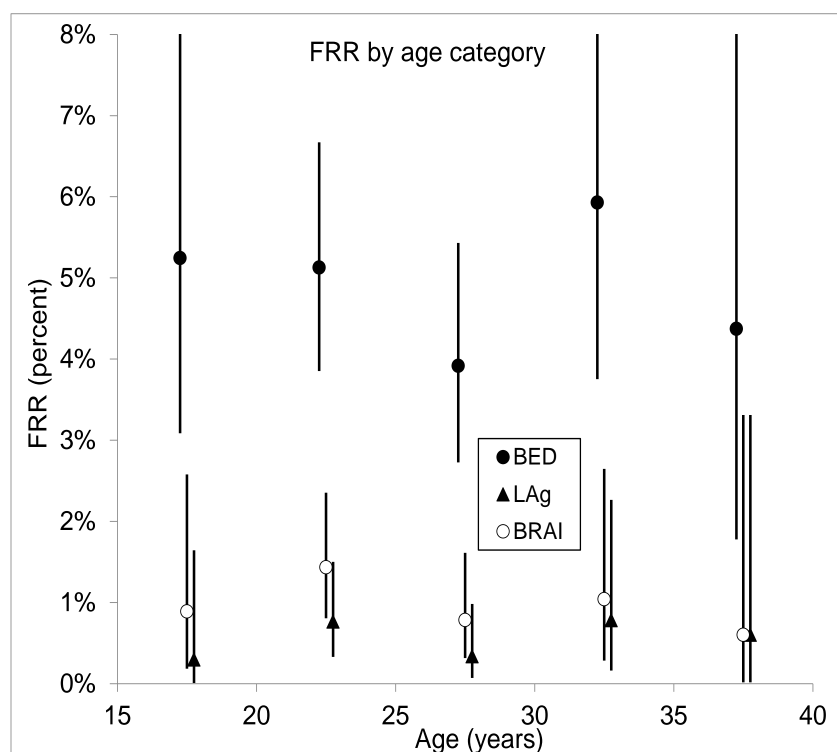
Since FRR and MDRI both increase with the pre-set cut-off C, FRRs also increase with increasing MDRI. For BED, we observed a steep increase in FRR with increasing MDRI: for LAg or BRAI, the rate of increase was much lower for all MDRIs less than about 140 days. At their respective recommended cut-offs, LAg has a smaller FRR (0.6% at $C = 1.5$) than for BRAI (1.1% at $C = 30\%$). However, when FRRs are compared at similar values of MDRI we find that, for MDRI values between 115 and 165 days there was no significant difference (CI overlap) between the LAg and BRAI FRRs. For example, at an MDRI of 141, the LAg and BRAI FRRs are closely similar; 0.96% and 1.1%, respectively (Figure 9-5).

Stellenbosch University https://scholar.sun.ac.za

9.3 FRR by Age Group of Mother

We investigated, for each of the three serological biomarkers, the effect of maternal age on FRR. For this analysis, we assigned each mothers to one of five age categories; < 20, 20-24, 25-29, 30-34 and ≥35 years old. We cross-tabulated the numbers of samples testing recent, or long-term, among HIV positive samples taken from mothers at Visit 5, among those who had previously tested HIV positive at Baseline. For all three serological biomarkers there was no significant difference between the proportions of "recent" diagnoses found in the different age categories (P>0.5, χ^2, 4 df). We illustrate the results graphically in Figure 9-6, from which it is clear that there is no consistent trend in FRR with age.

Figure 9-6: FRR by maternal age



Stellenbosch University https://scholar.sun.ac.za

9.4 Comparison of MDRI and FRR evaluation results to other studies

Evaluation studies have shown that the MDRI and FRR depend on HIV subtype. We compared our results with published results based on evaluations done on Clade C subtype. In order to be able to compare our estimates to those published by Kassanjee et al. (2014) or Duong et al. (2015), we did not include viral load as an initial screening tool (Table 9-3)

Table 9-3: Estimates of MDRI and FRR values for clade C HIV made in this study compared to studies of Kassanjee et al. (2014) and Duong et al. (2015).

	Cut-off	Published	ZVITAMBO	
		MDRI (days)	**MDRI (days)**	**[CV %]**
BED	**0.8**	287 (248 - 328)	188 (180 - 196)	[2.1%][1]
			197 (174 - 221)	[6.0%][2]
LAg	**1.5**	177 (150 - 206)[3]	104 (98 - 110)	[2.9%][1]
		152 (128 - 178)[4]	109 (98 -121)	[5.3%][2]
	2.0		141(134-148)	[2.4%][1]
			136 (120 -152)	[6.0%][2]
BRAI	**30%**		135 (120 - 151)	[5.9%][2]
	40%	298 (262 - 338)[3]	144 (128 - 160)	[5.6%][2]
		FRR	**FRR**	
BED	**0.8**	7.3% (2.6 - 15.7)[3]	4.8% (4.1 - 5.7)	
LAg	**1.5**	1.3% (0.0 - 7.2)[3]	0.6% (0.3 - 0.9)	
	2.0		0.96% (0.6-1.4)	
BRAI	**30%**		1.1% (0.7 - 1.5)	
	40%	6.7% (2.2 - 14.9)[3]	1.5% (1.1 - 2.0)	

Notes: 1= NLMM 2=SA 3=Kassanjee *et al.* (2014) 4= Duong *et al.* (2015)

As we have seen above, the MDRIs for BED, LAg and BRAI were significantly shorter, at the same cut-off, than those observed in Clade C samples from the general population. The FRRs are also all smaller: thus, we report FRRs of 4.8%, 0.6%, 1.06% and 1.5% for BED, LAg and BRAI at the standard cut-offs of 0.8, 1.5, 30% and 40% respectively. These are significantly lower than the 7.3%, 1.3% and 6.7% established in Subtype C cohorts within the region for BED, LAg and BRAI.[68, 69, 73] For the LAg assay, using to a cut-off of 2.0 yields an FRR of 0.96 at an MDRI of 137 days.

Stellenbosch University https://scholar.sun.ac.za

9.5 Discussion of the evaluation FRR for BED, LAg and BRAI assays

We carried out our evaluation on a cohort of postpartum women during a period when ART was not generally available in Zimbabwe. Because HIV affects fertility, it is biologically plausible that our sample did not include women who had been infected for an extended period. The availability of ART was limited to private sector clients in Zimbabwe during the period 1997-2002, and it is thus highly unlikely that our sample included women who had initiated and were already on ART at enrolment in the ZVITAMBO study. The limited availability of VL data for all women in the ZVITAMBO Trial makes it difficult to rule out the presence of elite controllers. We believe that our inability to rule out the presence of elite controllers or to rule out that any women were on ART does not in any way affect the accuracy of the results obtained in this evaluation. We believe that we can use the FRR and MDRI values obtained in this evaluation to estimate HIV incidence among postpartum women infected with sub-type C virus.

An evaluation by Hauser et al. (2014) suggested that BRAI has results that are more accurate for people recently infected with Subtype B virus.[64] In Hauser's evaluation among recently infected individuals, 60% (15/25) were correctly classified by BED, 88% (22/25) by BRAI and 48% (12/25) by LAg.[64] The FRR for BRAI and LAg were both 2% in other subtype B samples and 6% for non-type B and for BED, 7% subtype B (7/101) and 25% non-subtype B (4/16). Our evaluation did not have the same capacity to evaluate explicitly FRRs by stage of infection: this aspect requires further exploration.

In a detailed review of serological assays compared to newer molecular diagnostics technologies, used for determining recent HIV infections, Moyo et al. (2015) highlighted the significant progress made in improving accuracy and precision, but found that there were still challenges.[67, 97,] Serological tools that use anti p24 IgG3 and avidity assays were more sensitive than the detuned assays and more stable to variations in immune response, HIV subtype and the use of ART. Better still, molecular diagnostic tools displayed greater improvements by returning even lower misclassification rates.[18, 19,98] Combination tools that detect both antigens and antibodies provide simultaneous detection of HIV infection and prevalent

Stellenbosch University https://scholar.sun.ac.za

infection.[16,83, 91] These, and many other, tools provide innovative and cost effective means of improving sensitivity and specificity of assays for determining population level incidence.

The relationship between FRR and MDRI is such that an increase in MDRI is also associated with an increase in FRR. [20,73] The right combination of MDRI and FRR characteristics are critical in the performance of an assay. The utility of these parameters is defined by the need to have a sufficiently low FRR in order to screen out false assay positives (false recent) cases while ensuring that the MDRI maximally captures incident cases using minimal sample size. When applied in the calculation of HIV incidence, they should provide HIV incidence estimates that closely match follow-up estimates. The sample size required to achieve this accuracy is important, and workers at CDC and SACEMA have accordingly provided calculators for these considerations. Given the inter-variability between MDRI and FRR, Kassanjee et al. (2014) suggested that developers should focus on ensuring that, regardless of the optimum values, they judge assays on their ability to provide precise estimates of HIV incidence.[20]

Stellenbosch University https://scholar.sun.ac.za

10.0 Incidence Estimates derived using the three biomarkers

10.1 Preamble

Although the MDRI and FRR of a laboratory assay are key indicators of laboratory assay's performance, assay, the ultimate goal is its ability to provide an accurate measure of HIV incidence.[20, 62] The idea is to use some physiological measure (biomarker) that increases with time after HIV infection, set a convenient cut-off value for the biomarker, and define cases, which return a measure below predefined cut-off as recent infections. A true validation of such an assay would require that the assay be applied to samples collected from a different cohort of samples (than in which the FRR and MDRI where derived). In this evaluation, our analysis was limited to the same cohort.

Following significant investments in the development of laboratory assays, the USA Centers for Disease Control and Prevention (CDC) and UNAIDS, FHI 360 and the Bill and Melinda Gates Foundation have supported evaluations of candidate assays. The CEPHIA evaluation (2013), is probably one of the most significant evaluations that characterised five candidate assays; BED-CEIA, LAg avidity EIA, Vitros-Less Sensitive, Vitros Avidity and Bio-Rad Avidity Index EIA.[73] This evaluation concluded that, although LAg had a low FRR (1.3%) and a relatively high corresponding MDRI of 188 days for all specimens, none of the candidate assays met the criteria deemed critical for the accurate measurements of HIV incidence on their own[2, 73,97,98]

A test for recent infection need to have an MDRI that was long enough to include all incident cases (> 6 months) and a low FRR (<2%). Kassanjee et al. (2014) made a strong point that the primary focus of an assay's suitability should not be anchored on the assay's ability to correctly classify individuals as incident or long term (i.e., sensitivity or specificity) but rather on its ability to measure HIV incidence. However, this could only be achieved by ensuring an optimal trade-off between the FRR and MDRI.[73] The most precise estimates of HIV incidence are produced when the MDRI is large enough (to include all recent infections), while maintaining a small FRR (to

Stellenbosch University https://scholar.sun.ac.za

exclude long-term infections). Given the wide variations in immune response, i.e., antibody titre or maturation kinetics in response to HIV infection, HIV subtype and other demographic variations there is clearly a need to continue optimization of assays in order to improve on their performance.

The CEPHIA evaluation had a large diverse sample of ~ 2500 samples. While this large sample allowed for the determination of FRR and MDRI in sub-groups and a large diverse sample, our current evaluation provides a critical next step in that we provide FRR and MDRI for BED-CEIA, LAg and BRAI in a different sample and our evaluation is coming after two years of continuous improvements on the laboratory assays. Additionally, these samples are coming from a cohort where it was possible to document follow-up incidence estimates.[25]

Using the MDRI and FRR derived in this evaluation through the testing of ZVITAMBO trial samples collected at Visit 5, we were able to calculate HIV incidence estimates for the 12 months postpartum period and compare them to the longitudinal follow-up incidence. We note immediately that we are aware of the fact that both sets of estimates are derived from the same set of data, and that, in particular, the data have been used to estimate both the MDRI and FRR used for incidence estimation. If, therefore, all calculations have been correctly carried out, and all underlying assumptions are valid, there should be a good correspondence between the follow-up and biomarker estimates of incidence. Such correspondence does not, therefore, constitute a validation of the biomarker estimation procedure. Such a validation can only be achieved through the application of the methodology to data that have been independently derived. The importance, and interest, of the comparisons arises when there are differences between the follow-up and biomarker incidence estimates and/or differences between the various biomarker incidence estimates themselves.

We also used these MDRI and FRR parameter estimates to estimate HIV incidence using the baseline (V0) samples. These incidence estimates pertain to the period when the women were pregnant: we have independent estimate of HIV incidence for this period but it is of interest to compare these incidence estimates with those calculated from the Visit 5 samples.

Stellenbosch University https://scholar.sun.ac.za

We used the HIV serology results derived in the ZVITAMBO project trial to create a file of "Original Data". We also explored analysis with the "New Data" in which the HIV serology was changed for 18 out of the total 269 cases that were originally classified as seroconverting at V5 to reflect the revised status based on HIV serology retest and on HIV VL results. In addition, we assumed an FRR of zero for unadjusted analysis and used FRR in adjusted analysis at different cut-off values. We used the *SACEMA ABIE Version May 2015*-incidence calculator to calculate HIV incidence from our data.[99]

10.1.1 Samples used for HIV incidence estimation

Baseline testing of 14,110 mothers recruited to the ZVITAMBO study produced 4495 HIV positive and 9562 HIV negative results; 53 women could not be unequivocally classified as either HIV positive or negative. Among the HIV positive cases, the number testing as recent infections depends, of course, on the biomarker being used and the pre-set cut-off that has been chosen. When LAg was used with an OD cut-off of 1.5, 208 cases tested as recent infections.

At 12-months postpartum similar testing showed that 3081 women now tested HIV positive and 6537 HIV negative. The HIV positive cases included 2843 women who had tested HIV positive at baseline, and 231 who had seroconverted between baseline and 12-months postpartum. There were also 7 cases where the women whose HIV status could not be determined unequivocally at baseline, but who tested HIV positive at 12-months postpartum. When LAg was used with an OD cut-off of 1.5, 86 cases tested as recent infections.

Stellenbosch University https://scholar.sun.ac.za

10.2 HIV incidence rates over the first 12-months postpartum as estimated by follow-up and calculated using BED, LAg or BRAI.

The ZVITAMBO Project was able to measure follow-up incidence among postpartum women enrolled for follow-up for periods ranging from 12 - 24 months. Of the 9562 HIV negative women enrolled at baseline, 351 HIV women seroconverted in the first 12 months postpartum follow-up period. ZVITAMBO estimated HIV incidence using Turnbull's analysis with 95% confidence intervals calculated by the 'bootstrap' method.[89,100] The HIV incidence estimate for the first 12 months postpartum was 3.4% (95% CI 3.0-3.8) per annum.[25]

Data from the prospective follow-up provided us an opportunity to compare this 'gold standard' with HIV estimates calculated using all three of the serological biomarkers studied here, with the appropriate MDRI and FRR derived in this evaluation. We carried out these analyses for the categorisations made using the serological biomarkers by themselves, and for categorisations made using viral load plus a serological biomarker. Moreover, we carried out all of these analyses using either the "original" or the "new" data.

We draw attention to the fact that we have used ZVITAMBO data from seroconverting mothers to estimate MDRI values, and have used long-term infections to estimate the FRR values, and that these same ZVITAMBO data have been used to estimate the follow-up HIV incidence estimates. Where we have appropriate MDRI and FRR estimates, and have satisfied assumptions about the distribution of seroconversion events, there must then be a close correspondence between the follow-up and biomarker incidence estimates. The more interesting situations, which require explanation, arise when there is *not* a close correspondence between the estimates arising from the two approaches.

We also calculated incidence where we used viral load as an additional screening tool for removing cases that were long-term HIV infections. In this analysis, cases with VL <1000 were considered as long-term infections and were accordingly reclassified as long term, regardless of the level of the serological biomarker.

Stellenbosch University https://scholar.sun.ac.za

In all analyses, we show the confidence bounds of these estimates and the follow-up rate (black horizontal lines) in order to show these estimates in relation to the follow-up estimates.

In our initial analysis, we calculated HIV incidence rates using MDRI values estimated using either SA or NLMM for BED and LAg, but just SA for BRAI (Figure 10-1). We used our estimates of FRR to calculate the adjusted HIV incidence, but also set the FRR to zero, to calculate the (biased) *unadjusted* estimates. We acknowledge that unadjusted estimates are biased (not accurate) since in reality laboratory assay will misclassify some patients as recent, therefore when FRR is set to zero, the incidence estimates are indeed erroneous. The value of calculating the unadjusted estimates is to provide a measure of the error involved when we do *not* account for the FRR, and to see how this error varies when we estimated incidence using different biomarker systems.

10.2.1 Estimates using only serological biomarkers: original data.

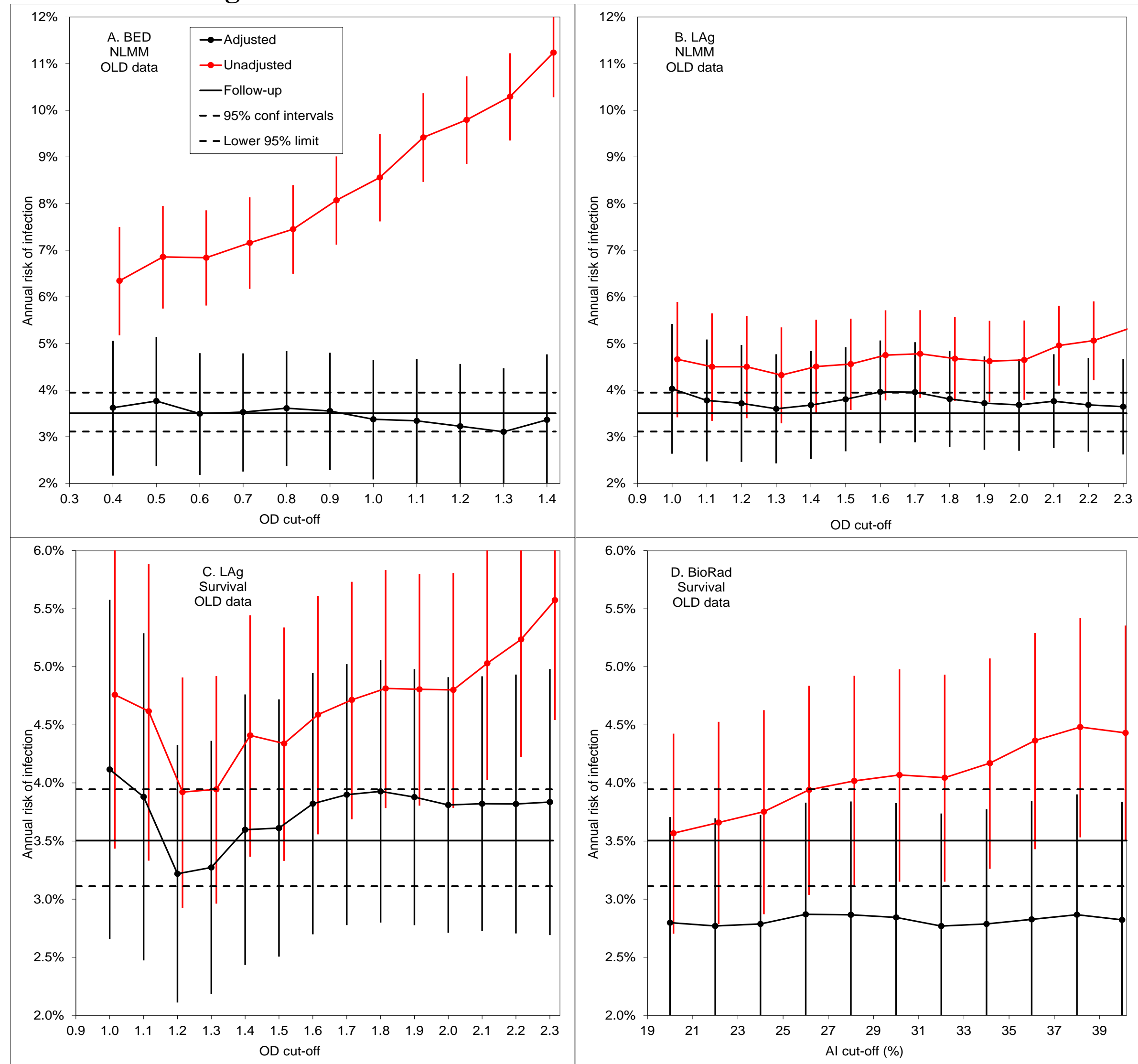
The unadjusted incidence estimates for BED were always higher than the follow-up and adjusted estimates, and the discrepancy increased with increasing cut-off C (Figure 10-1A). For LAg, the difference between the adjusted and unadjusted incidence estimate was always much smaller – in keeping with the much reduced FRR for LAg (Figure 10-1B). Nonetheless, the unadjusted incidence estimates were still biased and over-estimated the follow-up incidence. By contrast, the adjusted incidence estimates for both BED and LAg, over almost the range of C, lay within the 95% confidence intervals of the follow-up incidence (Figure 10-1A, B).

For BED and LAg, NLMM and survival analysis give very similar MDRI estimates and, accordingly, the resulting HIV incidence estimates are similar whichever of the sets of MDRI estimates we use. For comparisons between the LAg and BRAI incidence estimates, we use the survival analysis MDRI estimates, since it was not possible to use NLMM to estimate the BRAI MDRI.

Stellenbosch University https://scholar.sun.ac.za

Figure 10-1 C and D show that, whereas for LAg the adjusted incidence estimates show a close correspondence with the follow-up estimate, the BRAI estimates lie consistently below the lower 95% confidence interval band for follow-up estimate.

Figure 10-1: HIV Incidence over the first 12 months postpartum derived using ZVITAMBO original data.

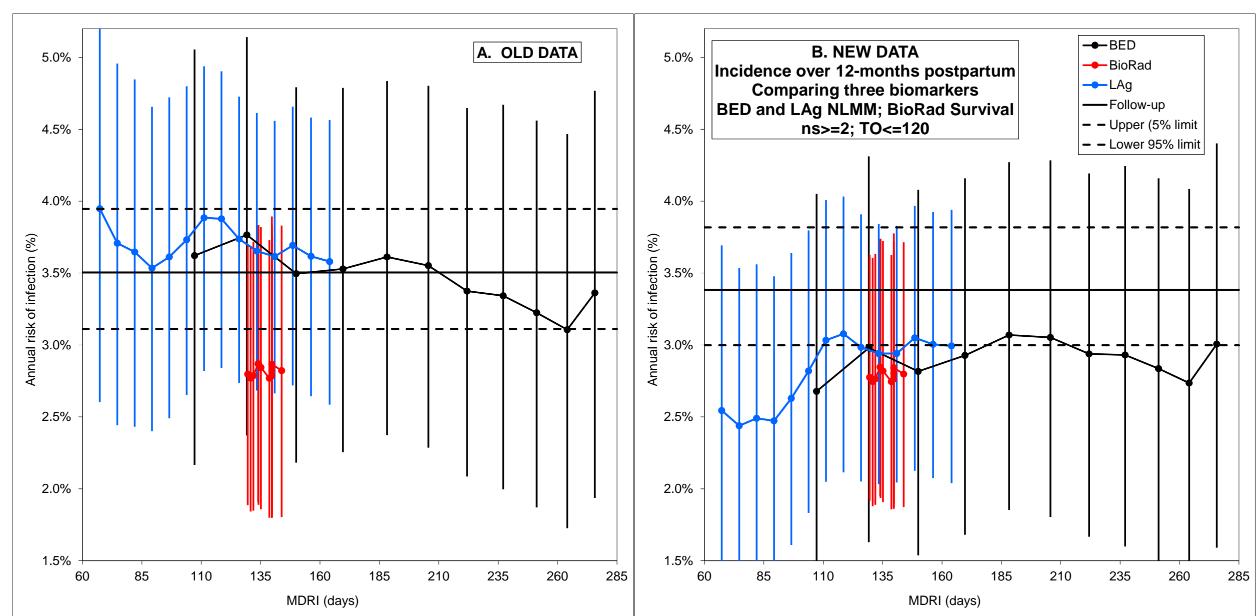


Notes:
A and B: BED and LAg (MDRI calculated using NLMM).
C and D: LAg and BRAI (MDRI calculated using survival analysis).
Notice the difference in scales on the Y-axis for A and B *vs* C and D.

Stellenbosch University https://scholar.sun.ac.za

We compared HIV incidence estimates obtained when we used the original data compared to when we changed the HIV serology of 17 samples from HIV positive, and recent, infections to HIV negative. This analysis is key in showing the potential impact of the misclassification of HIV serology (Figure 10-2). This analysis is based on *ns*=2, t_0> 120 days and MDRI calculated using NLMM for BED and LAg and SA for BRAI.

Figure 10-2: Adjusted HIV incidence estimates for BED, LAg and BRAI comparing Original and New Data



The adjusted estimates using the original data for BED and LAg had values within the 95% CI, while the adjusted values for BRAI assay were below the lower bound of the 95% CI, whether we used the original or the new classifications. Using both recommended cut-off values considered in other evaluations, we found that, for the LAg assay, there was no difference between incidence established at cut-off 1.5 or 2.0. However, the “new” estimates were lower than when we used the original classifications. These differences, between incidence estimates arising from the use of original vs new classifications, were not obvious when we applied BRAI analyses to the samples. (Figure 10-2)

Stellenbosch University https://scholar.sun.ac.za

10.2.2 Estimates using serological biomarkers plus viral load.

As seen from the results presented so far, the HIV incidence estimates can vary according to the biomarker we use to make the estimates – and whether we use the original or the new diagnoses. The situation is further complicated if we explore the possibility of using viral load as an additional tool for identifying recent infections. We now illustrate this situation with reference to estimates made using the LAg biomarker.

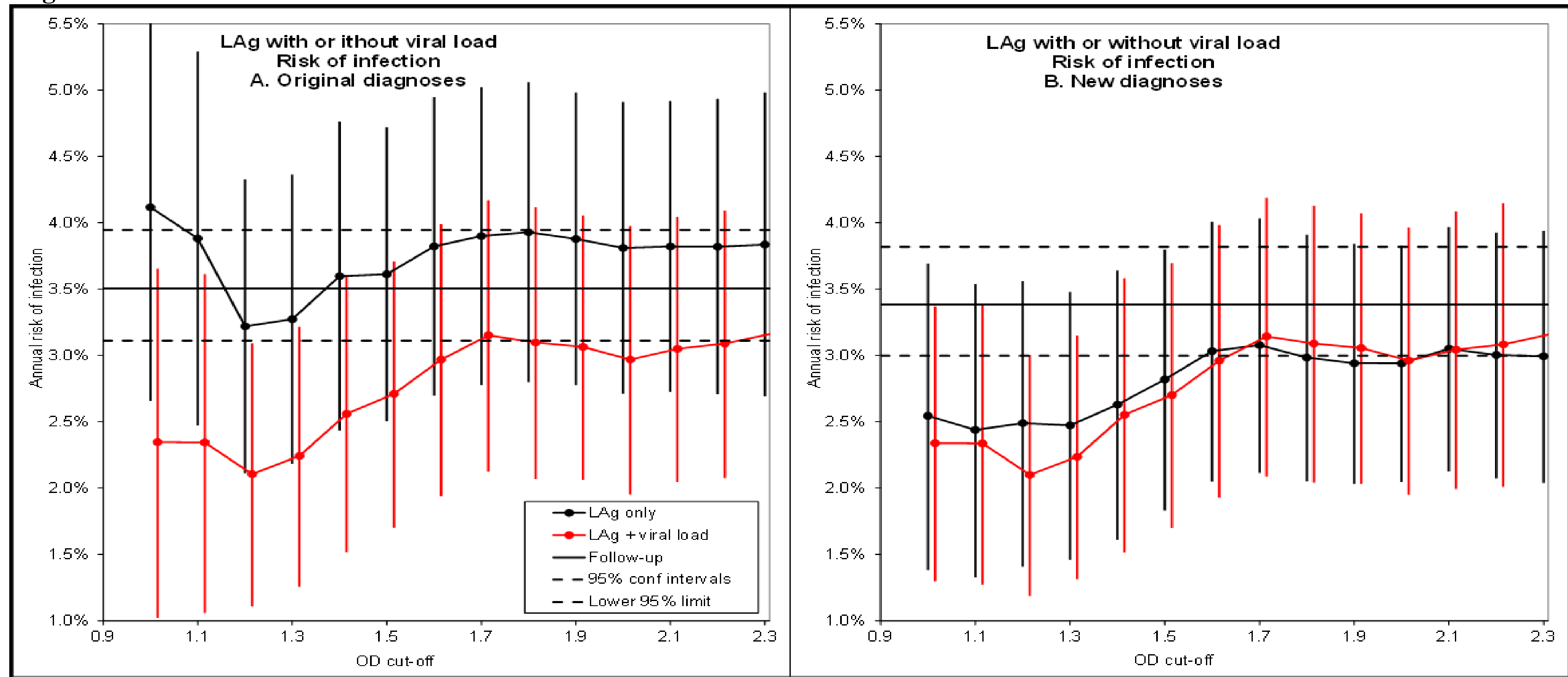
When we use LAg only to identify recent infections, using the original HIV diagnoses, then we see, as above, that there is a close correspondence between the adjusted HIV incidence estimates at Visit 5 and the ZVITAMBO follow-up estimates (Figure 10- 3A). When we also identify any case with a viral load <1000 as a long-term infection, regardless of the LAg reading, the estimated incidence drops by order 25% and, for almost all values of *C* used, the point estimates lie below the lower bound of the 95% confidence interval of the follow-up incidence. (Figure 10- 3A).

When, however, we used the new HIV diagnoses there was virtually no difference between the estimates made using LAg with or without the added use of viral load levels. (Figure 10- 3B). Moreover, these estimates were closely similar to the estimates obtained when we used both LAg and viral load in the analysis of the original data. (Figure 10- 3B).

These results can be understood when we realise that all of the disputed cases that constitute the difference between the old and the new data are samples which had viral load<1000. Accordingly, we always classify these cases as long-term infections, whether we are using the old or the new diagnoses. Moreover, when we are using the new diagnoses, all of these cases are classified as HIV negative and, accordingly do not appear as recent infections.

Stellenbosch University https://scholar.sun.ac.za

Figure 10-3: Annual risk of infection as calculated using LAg with or without VL and using either the original or the new HIV diagnoses.



Note: MDRI estimated using survival analysis: $ns = 2$; $t_0 = 120$ days.

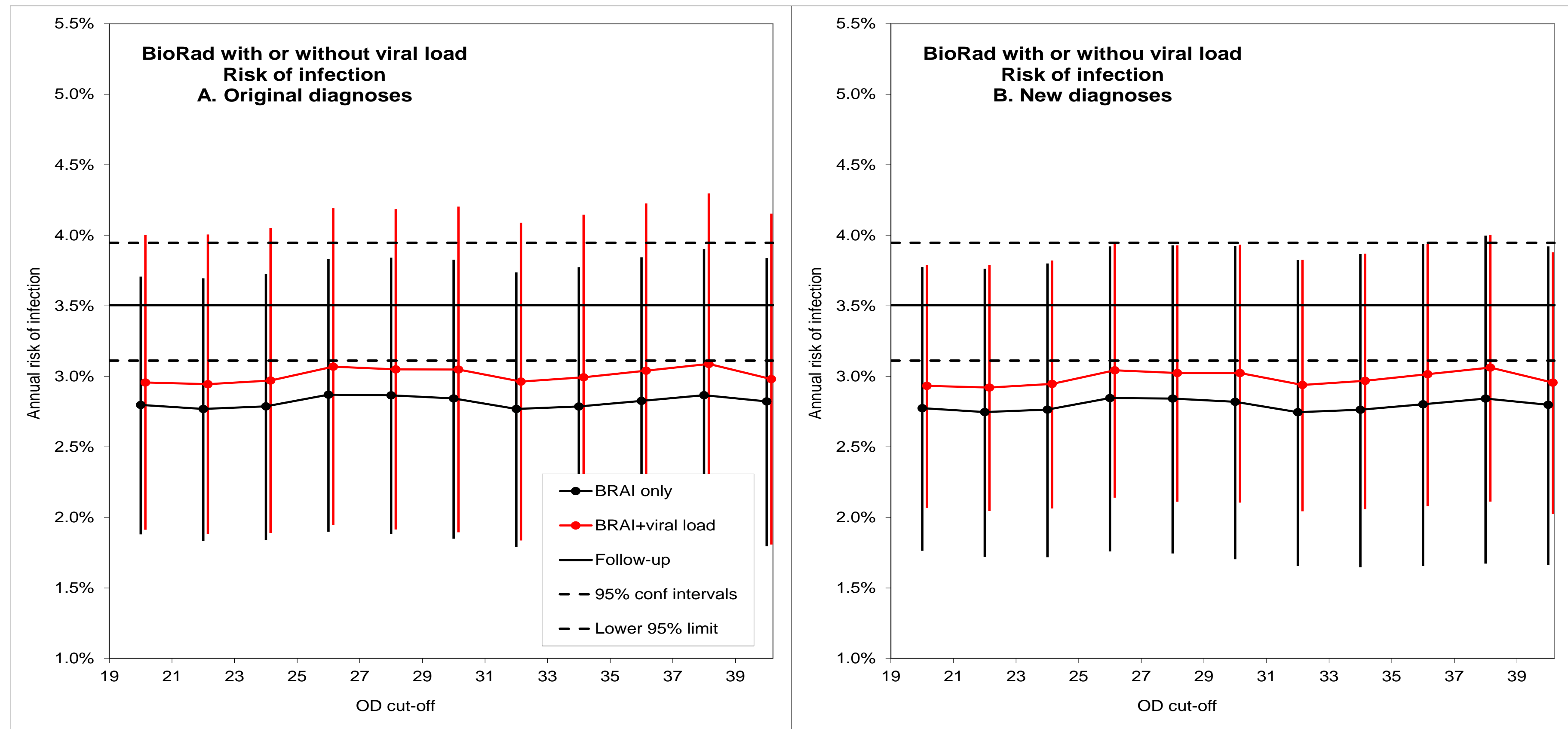
Stellenbosch University https://scholar.sun.ac.za

For BRAI, it made no difference whether we used the original or the new diagnoses, or whether we use BRAI alone, or in combination with viral load, to identify recent infections (Figure 10- 4). This differed from the results for LAg and BED using similar situation in that the estimates obtained using BRAI applied to the original data no longer differed from the other estimates – and were consistently lower than the follow-up estimates. We can understand this result on the basis that the 18 samples are always classified as "invalid" when analysed using BRAI. They can therefore never be classified as recent infections, regardless of whether we use the new or old data set, nor whether we use, or do not use, viral load to identify recent infections.

We compared the estimates of HIV incidence when MDRI was calculated using SA methods for all three assays using original (Figure 10- 5) and new data (Figure 10- 6).

Stellenbosch University https://scholar.sun.ac.za

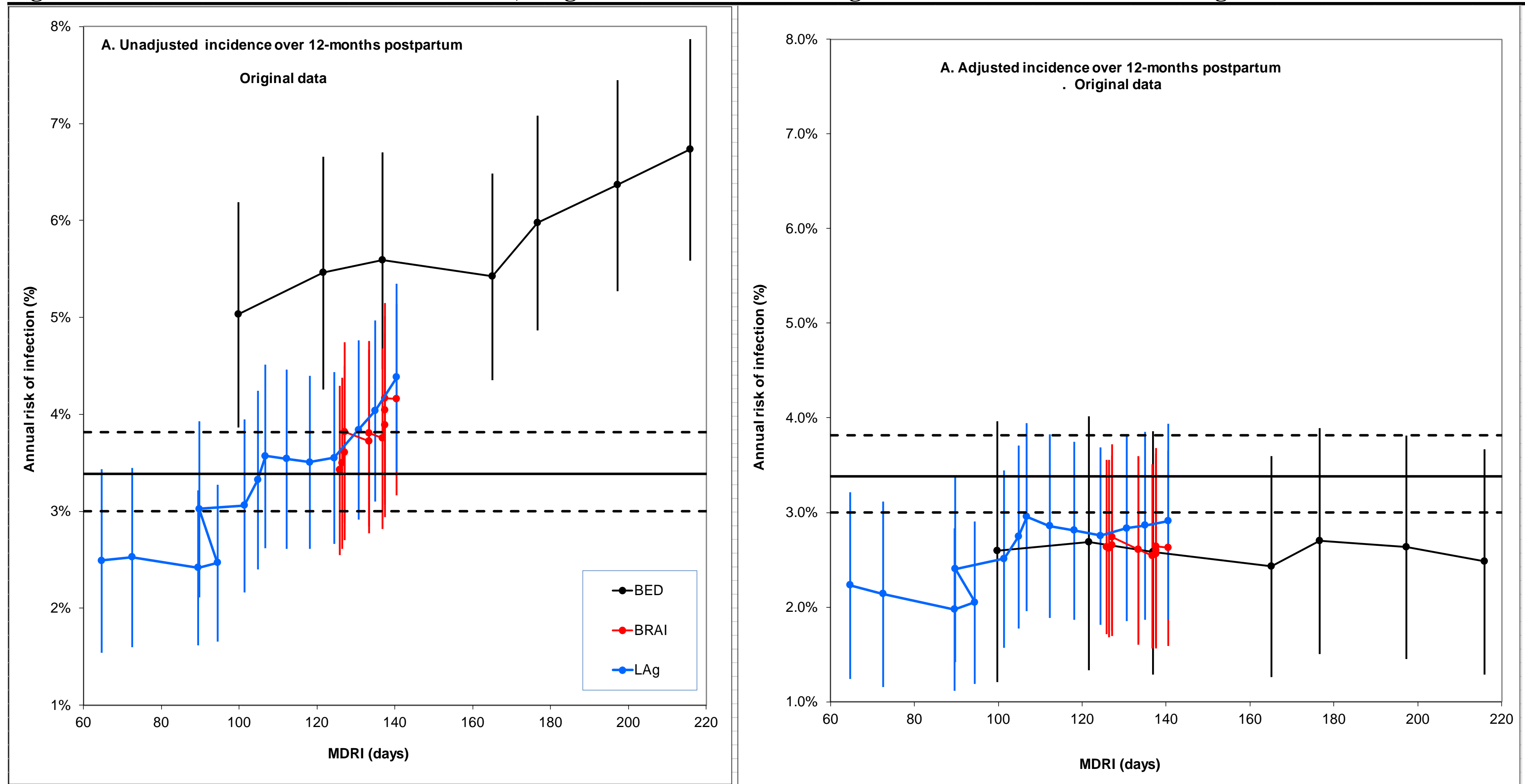
Figure 10-4: Annual risk of infection as calculated using BRAI with or without VL and using the original and new HIV diagnoses.



Notes: MDRI estimated using survival analysis: $ns = 2$; $t_0 = 120$ days.

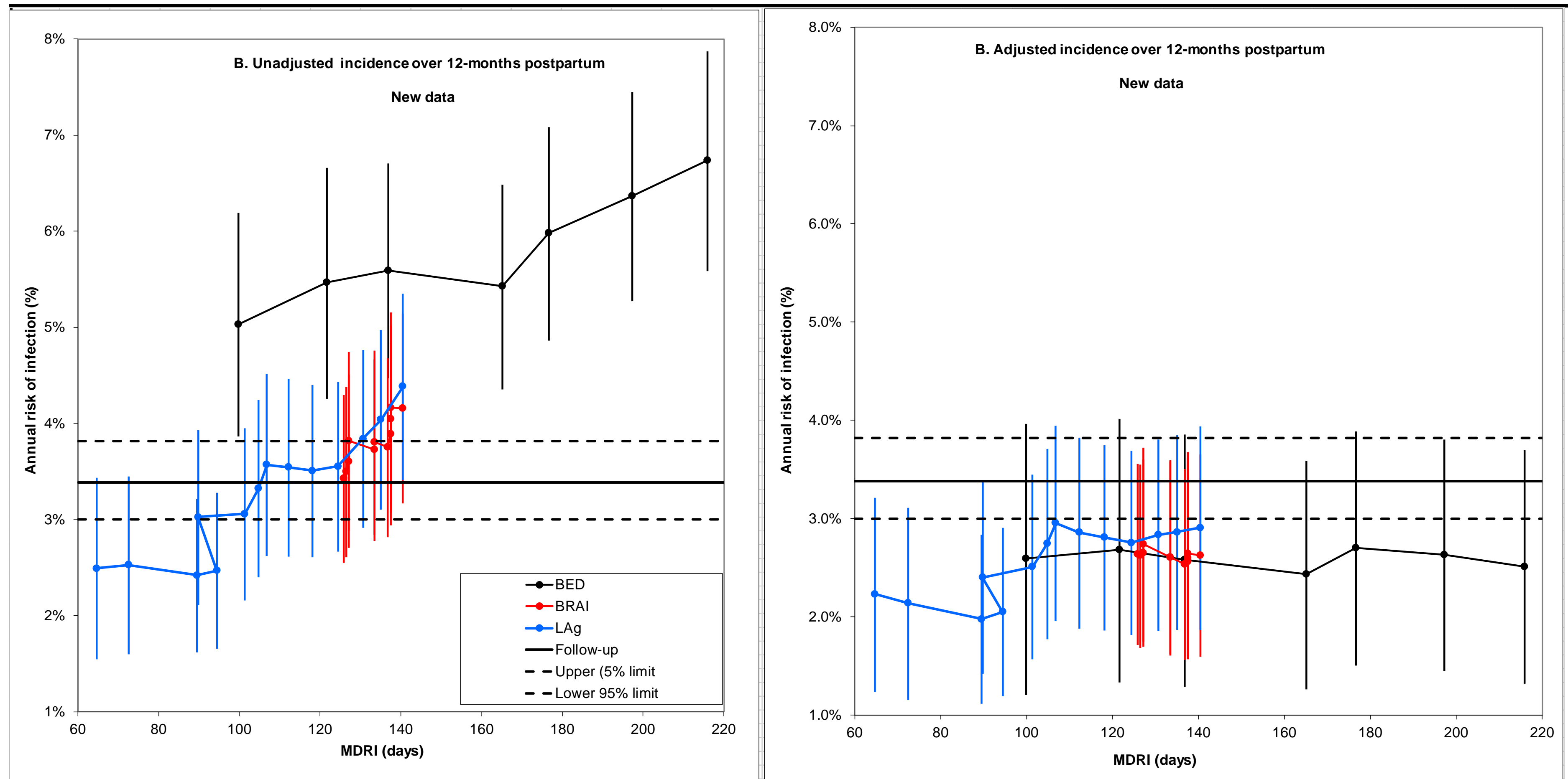
Stellenbosch University https://scholar.sun.ac.za

Figure 10-5: Annual risk of infection for BED, LAg and BRAI with VL using SA methods for MDRI and Original Data



Stellenbosch University https://scholar.sun.ac.za

Figure 10-6: Annual risk of infection for BED, LAg and BRAI with VL using SA methods for MDRI and New Data



Stellenbosch University https://scholar.sun.ac.za

10.3 Baseline HIV Incidence using BED, LAg and BRAI with or without viral load

10.3.1 Using serological biomarkers only

Women who were enrolled in the study, within 96 hours postpartum, provided a baseline blood sample. Using samples that were HIV positive, we calculated the baseline HIV incidence. We assume that these HIV incidence estimates are useful in inferring HIV incidence associated with the period when the women were pregnant.

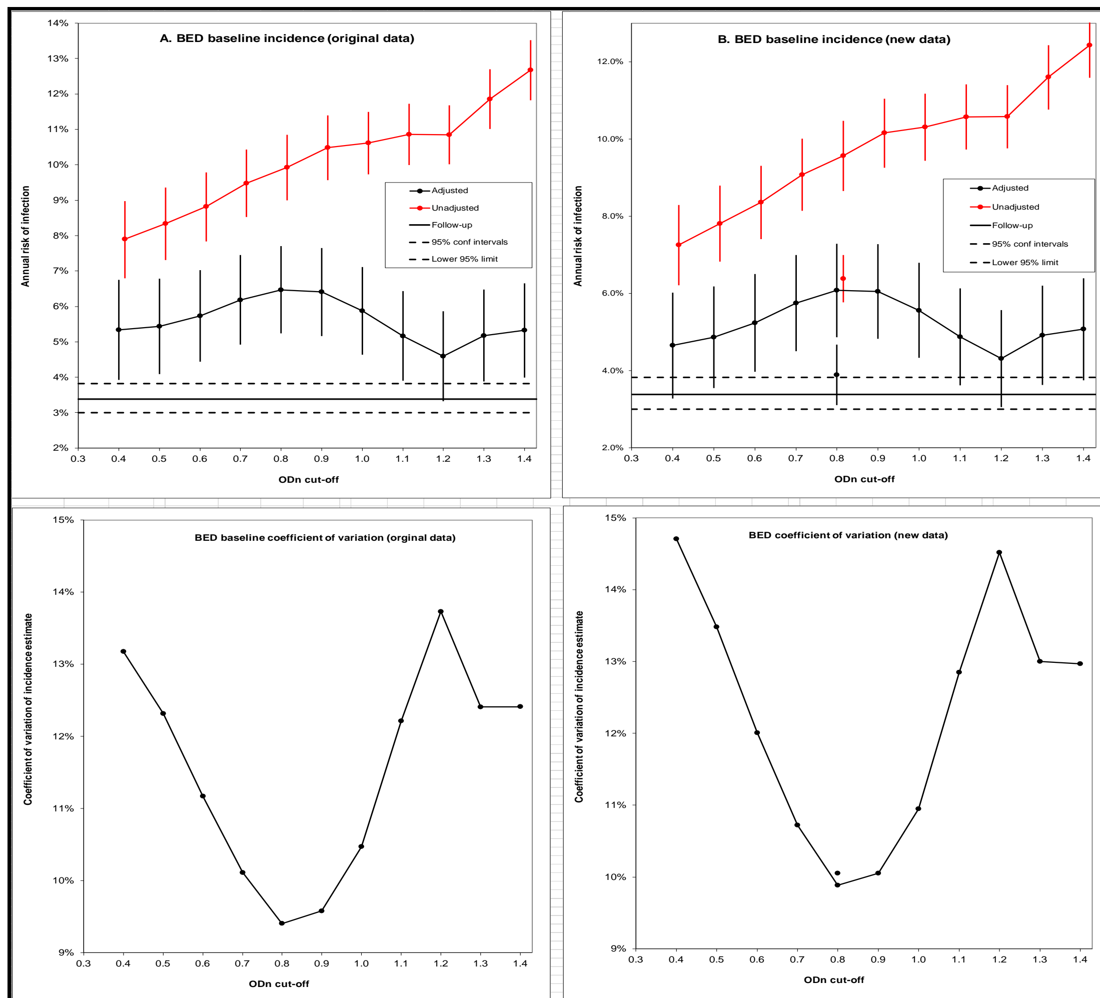
There is published evidence that the risk of infection is higher during pregnancy than it is during the postpartum period.[101] In presenting the baseline incidence estimates we have also shown the postpartum follow-up incidence estimate, simply to show the differences between our HIV incidence estimates during two very different physiological states. There is no implied suggestion about the validity of either set of estimates.

Using the most precise MDRI estimates (NLMM estimates for BED and LAg, and SA for BRAI) and associated FRRs, we calculated the HIV adjusted and unadjusted incidence estimates at baseline. We also calculated a multi-assay incidence where we used the viral load results to identify as long term infections any samples with VL<1000 copies per ml.

Stellenbosch University https://scholar.sun.ac.za

For the BED assay, at the recommended cut-off of 0.8, baseline unadjusted HIV incidence was 10.5% (95% CI 9.4- 11.5) and 10.1 % (9.1 -11.1) for old and new data respectively (Figure 10-7). In contrast, the adjusted incidence were, as expected, significantly lower and did not show the strong monotonic increase with increasing C that was seen in the unadjusted estimates. Indeed, the incidence estimates declined slightly between cut-off 0.8 to 1.2. The adjusted estimates at $C = 0.8$ were 6.7% (95% CI 5.5 - 7.9) and 6.1% (95% CI 4.9 - 7.2) for original and new data, respectively. As expected, these adjusted estimates are higher than the follow-up rate of 3.4% obtained in the ZVITAMBO trial. Of note was the fact that the coefficient of variation in the estimate of adjusted incidence took a unique minimum at the preferred C = 0.8%.

Figure 10-7: Baseline (V0) HIV incidence for BED by cut-off using original and new data

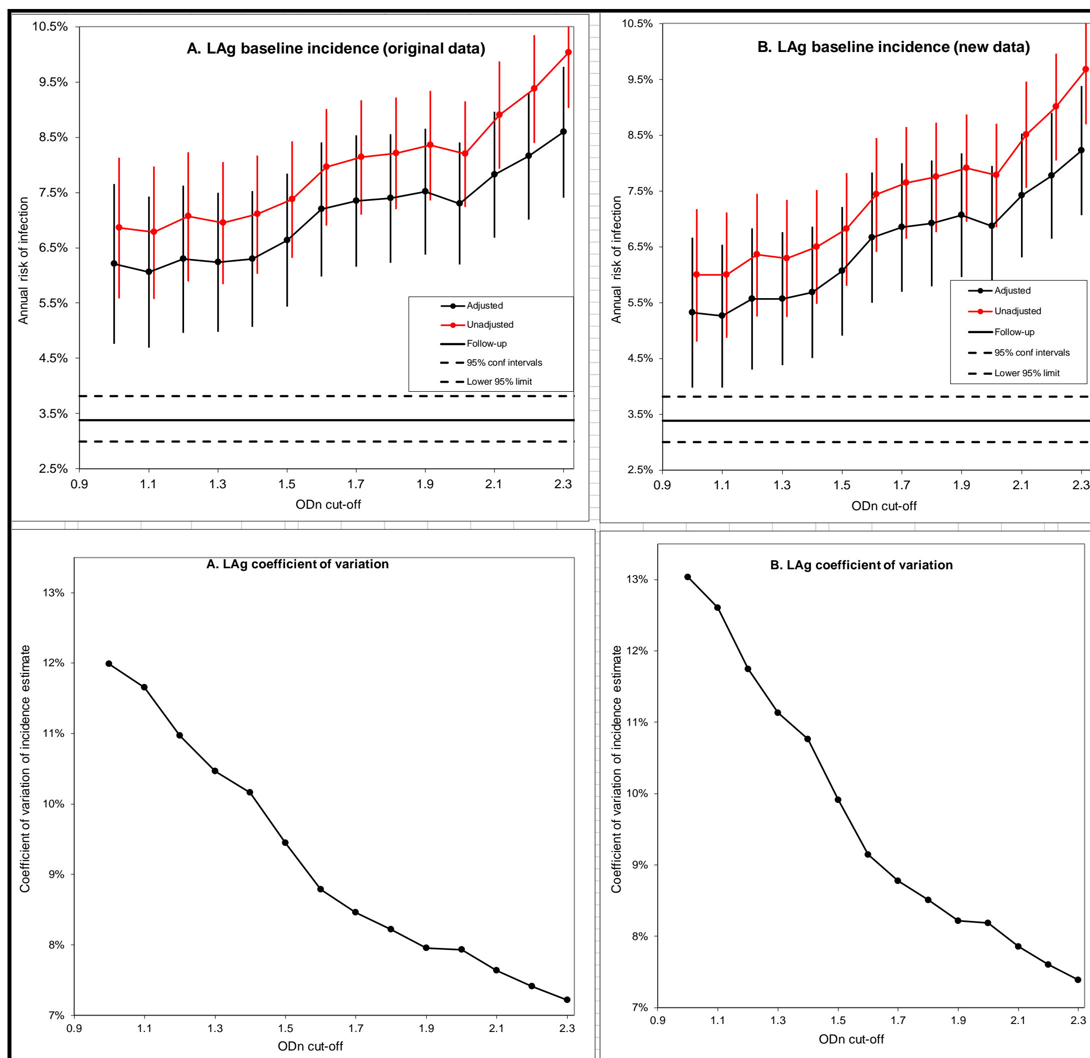


Stellenbosch University https://scholar.sun.ac.za

For the LAg assay, both the adjusted and unadjusted estimates increase with *C*, in contrast to the BED results, where adjusted estimates were independent of C (Figure 10- 8) For the LAg assay, the coefficient of variation decreases with increase in cut-off over the entire range of cut-off that we tested. There were no significant differences between 'Original' and 'New' data. At the standard cut-off =1.5 the adjusted incidences were 6.9 % (95% CI 5.7- 8.1) and 6.1% (95% 5.0- 7.1) for old and new data respectively.

Stellenbosch University https://scholar.sun.ac.za

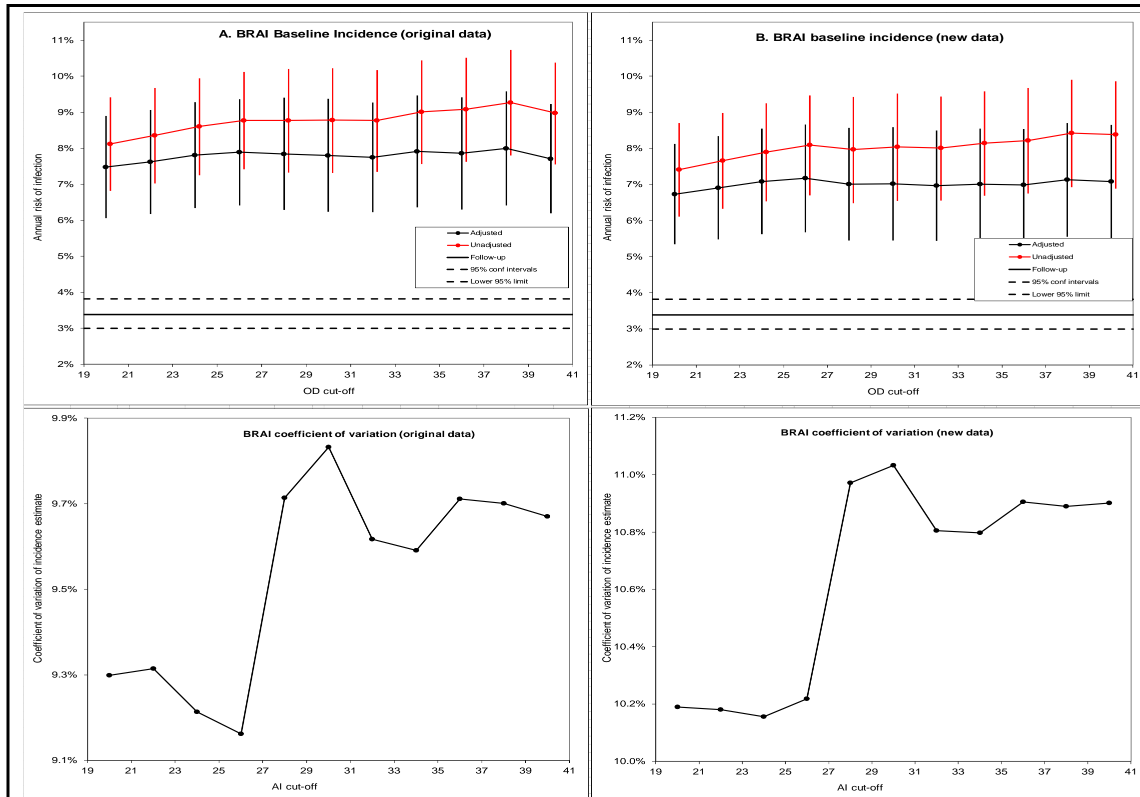
Figure 10-8: Baseline (V0) HIV incidence for LAg by cut-off using original and new data



Stellenbosch University https://scholar.sun.ac.za

For the BRAI assay both the unadjusted and adjusted estimates are largely independent of cut-off (Figure 10- 9). Moreover, the overlap of the 95% confidence intervals shows that there was no statistically significant difference between the adjusted and unadjusted incidence estimates – though the former were consistently lower at every cut-off point.

Figure 10-9: Baseline (V0) incidence for BRAI by cut-off using original and new data



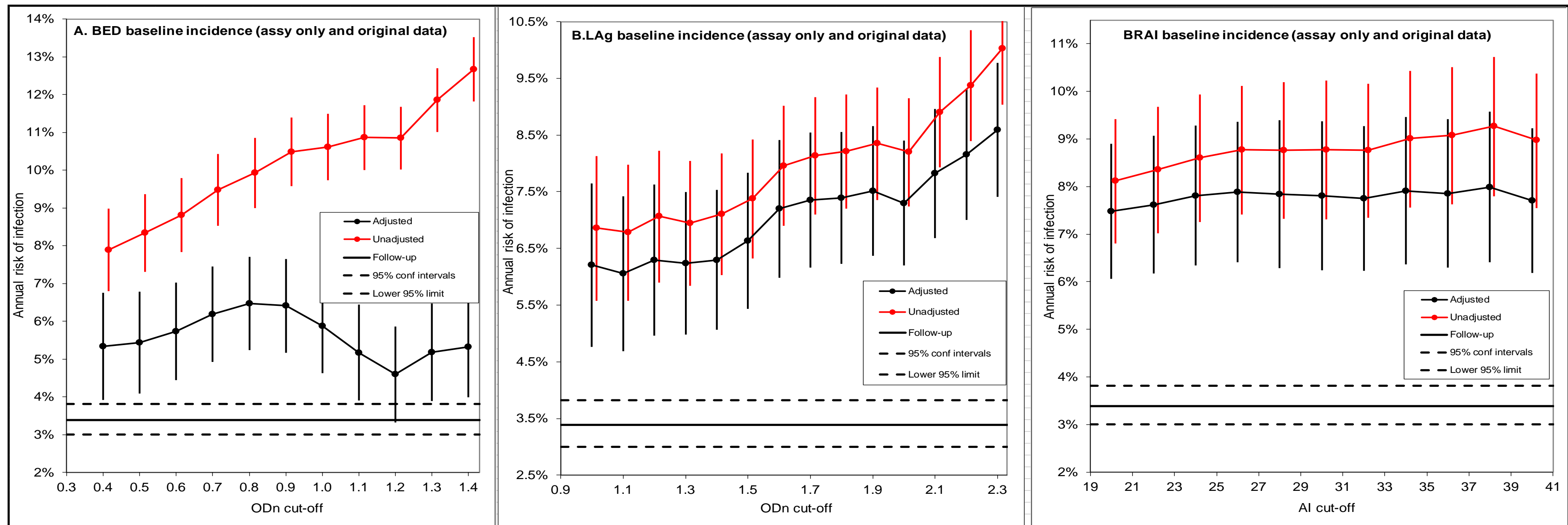
The overlap of 95% confidence intervals shows that there were no statistically significant differences between the incidence estimates calculated using the 'Original' or the 'New" data. At an AI cut-off of 40% the adjusted estimates were 7.0% (95% CI 5.6 - 8.5) and 7.8% (95% CI 6.3 - 9.2%) for old and new data respectively. There is significant variation in coefficient of variation with increase in cut-off and we note that the CV actually takes a maximum value at the preferred $C = 30\%$. In fairness, however, the absolute size of the variation in the CV is small, varying only between about 9.2 and 9.8. The lower end of this range is still higher, however, than the CV for the LAg estimate of incidence made at the preferred value of $C = 1.5$

Stellenbosch University https://scholar.sun.ac.za

For an easy visual comparison of these variations for the three different serological biomarkers we plot the baseline incidence against cut-off for each assay (Figure 10- 10).

Stellenbosch University https://scholar.sun.ac.za

Figure 10-10: Baseline incidence estimates for BED, LAg (NLMM) and BRAI (Survival) unadjusted and adjusted estimates (original data)



Stellenbosch University https://scholar.sun.ac.za

The difficulty with comparing the incidence estimates obtained, for different biomarkers, over a range of values of the pre-set cut-off values (*C*) is that *C* has analysis entirely different range for each of the biomarkers and each range is associated with a quite different range of MDRI values.

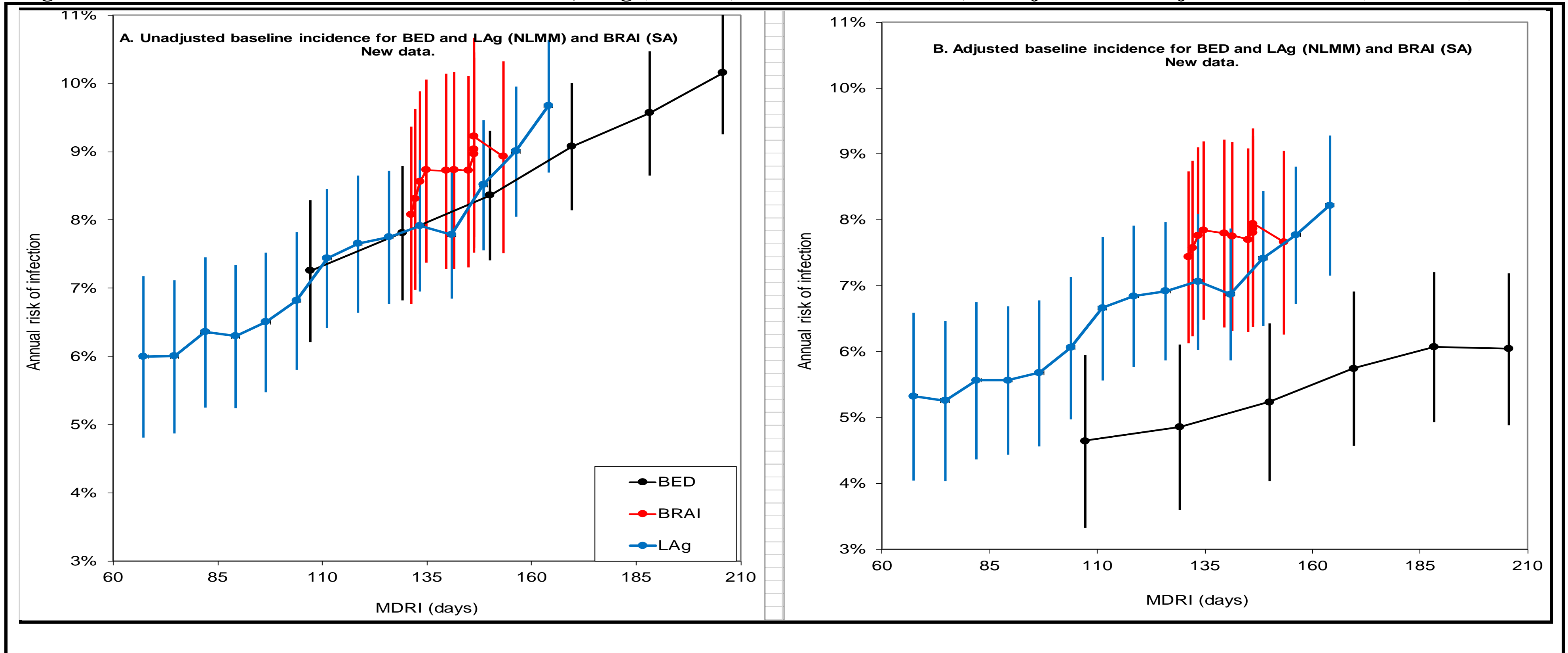
Accordingly, in an attempt to get a more meaningful comparison of biomarker performance, we plotted the estimated baseline HIV incidence against estimated MDRI for the three different assay systems (Figure 10- 11). We calculated HIV incidence estimates using the most precise estimate of MDRI i.e., NLMM estimates for BED and LAg, and SA estimates for BRAI.

Unadjusted HIV incidence estimates increased approximately linearly, from about 6% to 10%, across the whole range of observed MDRIs of 60-210 days, with the results for all biomarkers lying close to the same trend line. The trend for BRAI is not as clearly marked because the MDRI changes relatively little with changes in *C (*Figure 10- 11A).

Adjusted HIV incidence estimates were, of course, lower than the unadjusted estimates for all three assays (Figure 10- 11B). In keeping with the markedly higher FRR for BED, the adjusted incidence estimates for this biomarker were significantly lower than for the other two biomarkers (Figure 10- 11B). Adjusted estimates increased noticeably with MDRI for LAg (5.3%-8.2% for MDRI 67-164 days) and BED (4.7%-6.1% for MDRI 107-206 days). For BRAI the small range of MDRI values again made it difficult to discern a trend.

Stellenbosch University https://scholar.sun.ac.za

Figure 10-11: Baseline incidence estimates for BED, LAg (NLMM) and BRAI (Survival) unadjusted and adjusted estimates (New data)



Stellenbosch University https://scholar.sun.ac.za

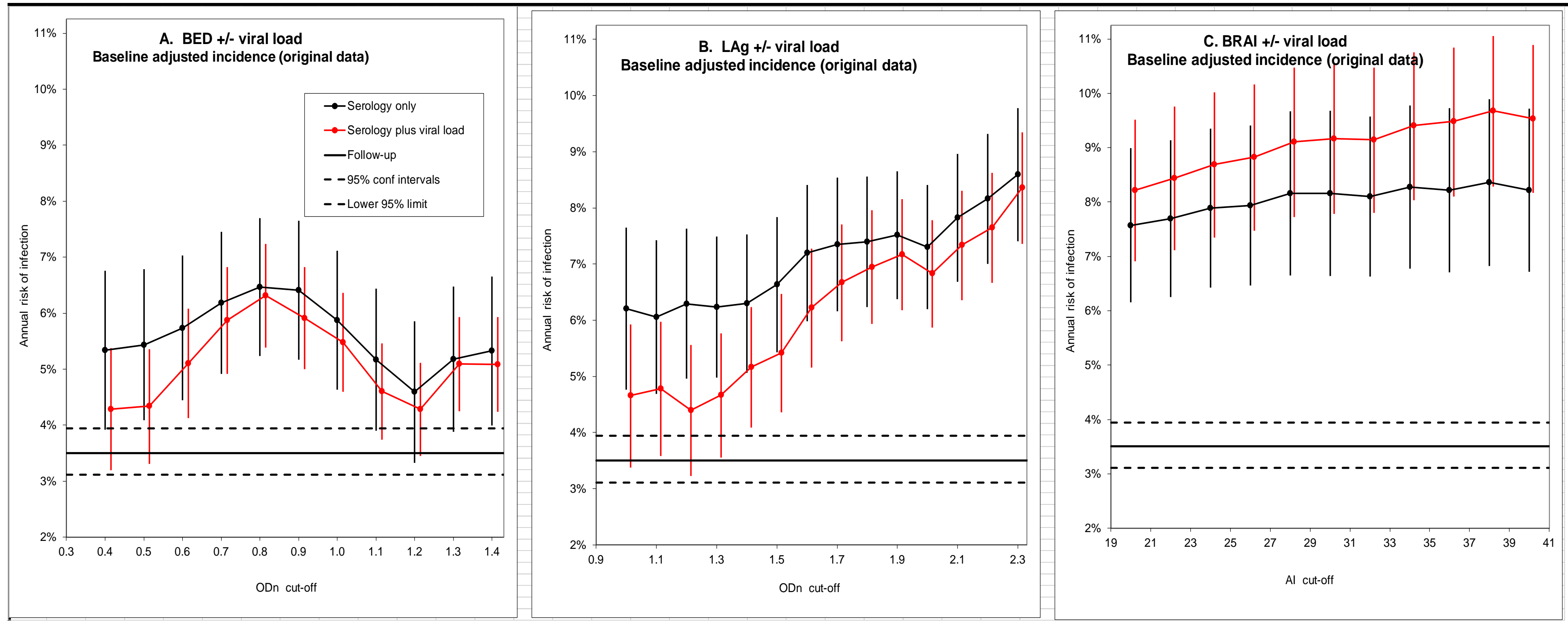
10.3.2 Using viral load in conjunction with a serological biomarker for baseline estimates of incidence

In this evaluation, we considered all samples with viral load (VL) <1000 copies/ml as long-term infections. The baseline adjusted HIV incidence estimates were lower when we used VL together with BED or LAg although, at most values of *C* overlap of the 95% confidence intervals showed that the means did not differ significantly (Figure 10-12). For BRAI, the situation was reverse, with the estimates derived using the BRAI by itself producing lower estimates than when we used BRAI in conjunction with VL. Again, however, there were no significant differences between the means

When we use viral load as well as serology to identify recent infections, there will be fewer cases testing recent in any cross sectional survey. This tends to decrease the HIV incidence estimate. However, the FRR and the MDRI are also lower when viral load is used and those changes tend to *increase* the incidence. It is thus not easy to predict how the two sets of estimates will differ. The point to bear in mind is that there exists a *true* incidence level – and if our estimation procedures are good, and the MDRI and FRR estimates appropriate, then we should get estimates that approximate the true incidence level, whether we use just the serology or the serology *plus* the viral load.

Stellenbosch University https://scholar.sun.ac.za

Figure 10-12: Comparison of baseline HIV incidence for BED, LAg and BRAI assays used either alone or in combination with VL

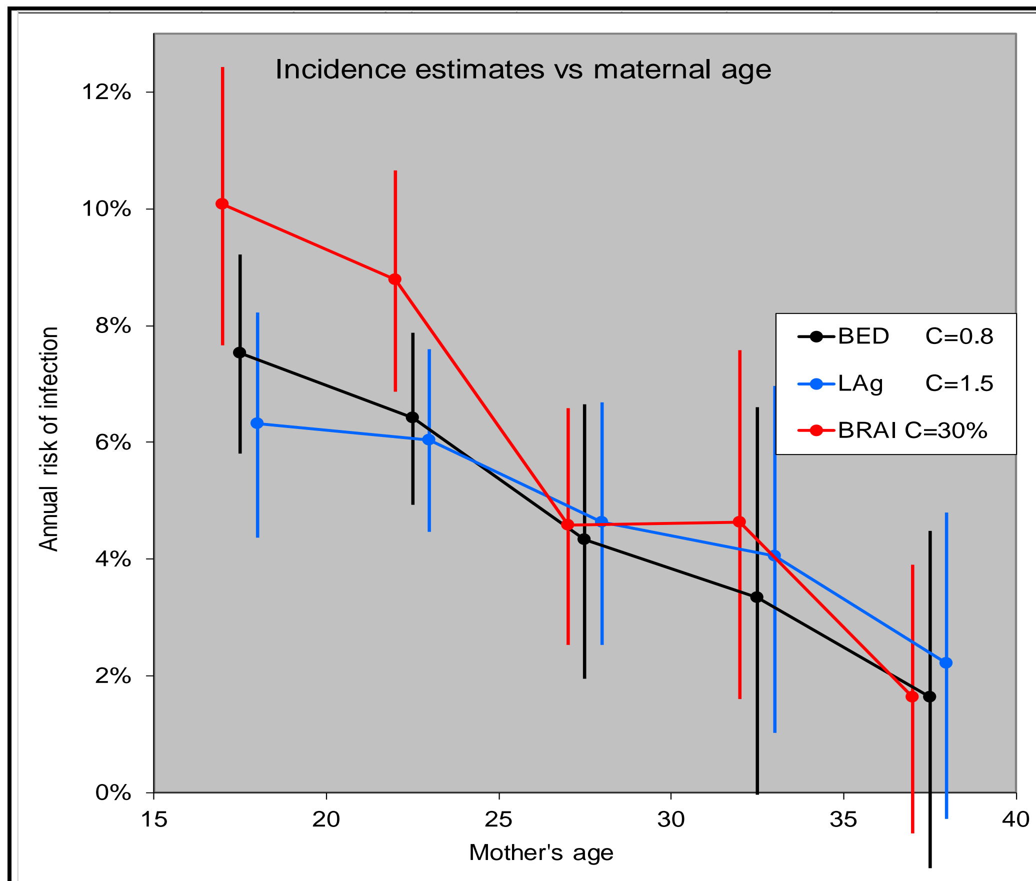


Stellenbosch University https://scholar.sun.ac.za

10.4: HIV incidence as a function of age or parity

In the ZVITAMBO cohort of 14,011 women, about 58% were below 24years old, 37% were in the age-group 25-34 and only 5% were 35-42 years. We estimated HIV incidence as a function of baseline maternal age, since age can be an important predictor of disease acquisition and progression.[102] For all three assays, adjusted baseline HIV incidence decreased with age, and was particularly high in young women 15-24 years (Figure 10-13). These results mirrored the follow-up estimates observed in the ZVITAMBO cohort where incidence declined from 5.4% (age <20) to 0.6% (age 38-41 years).[25]

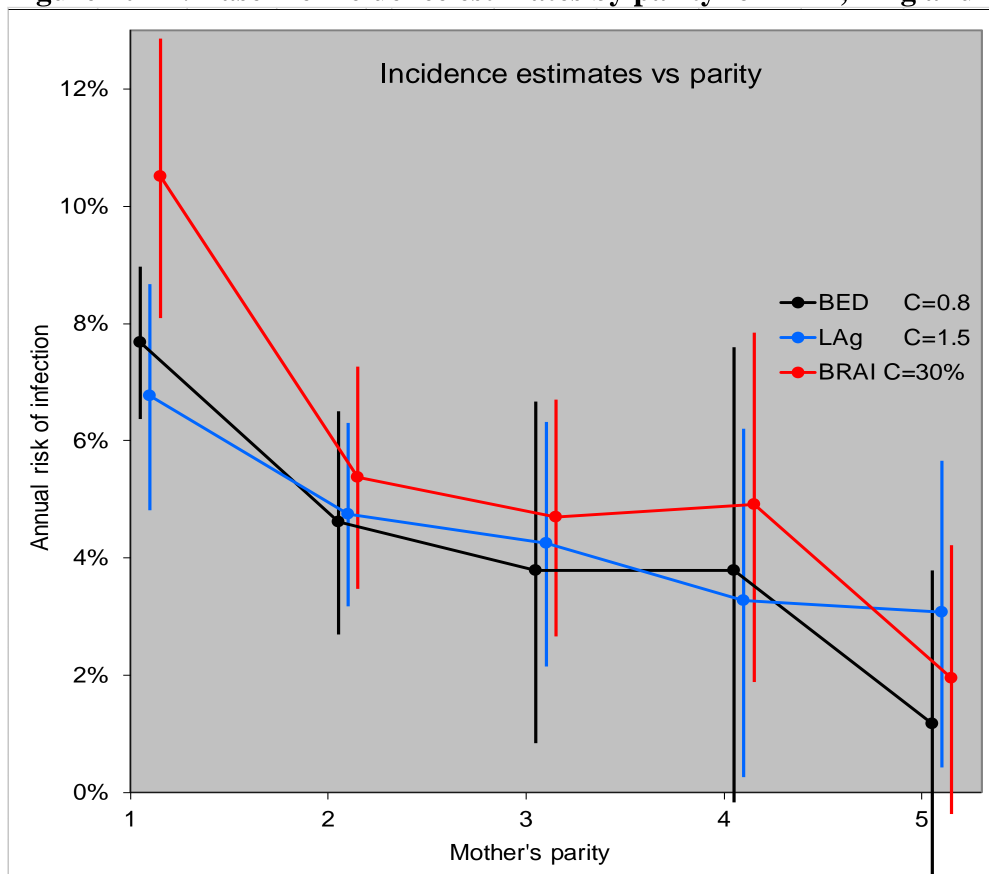
Figure 10-13: Baseline incidence estimates and maternal age for BED, LAg, and BRAI



Stellenbosch University https://scholar.sun.ac.za

We explored the rates of infection by parity of women in the ZVITAMBO cohort at baseline. Naturally older women tend to have more children than the younger women. Not surprisingly, therefore, we observed a similar trend in the relationship of parity and HIV incidence as that observed with age and HIV incidence (Figure 10-14).

Figure 10-14: Baseline incidence estimates by parity for BED, LAg and BRAI



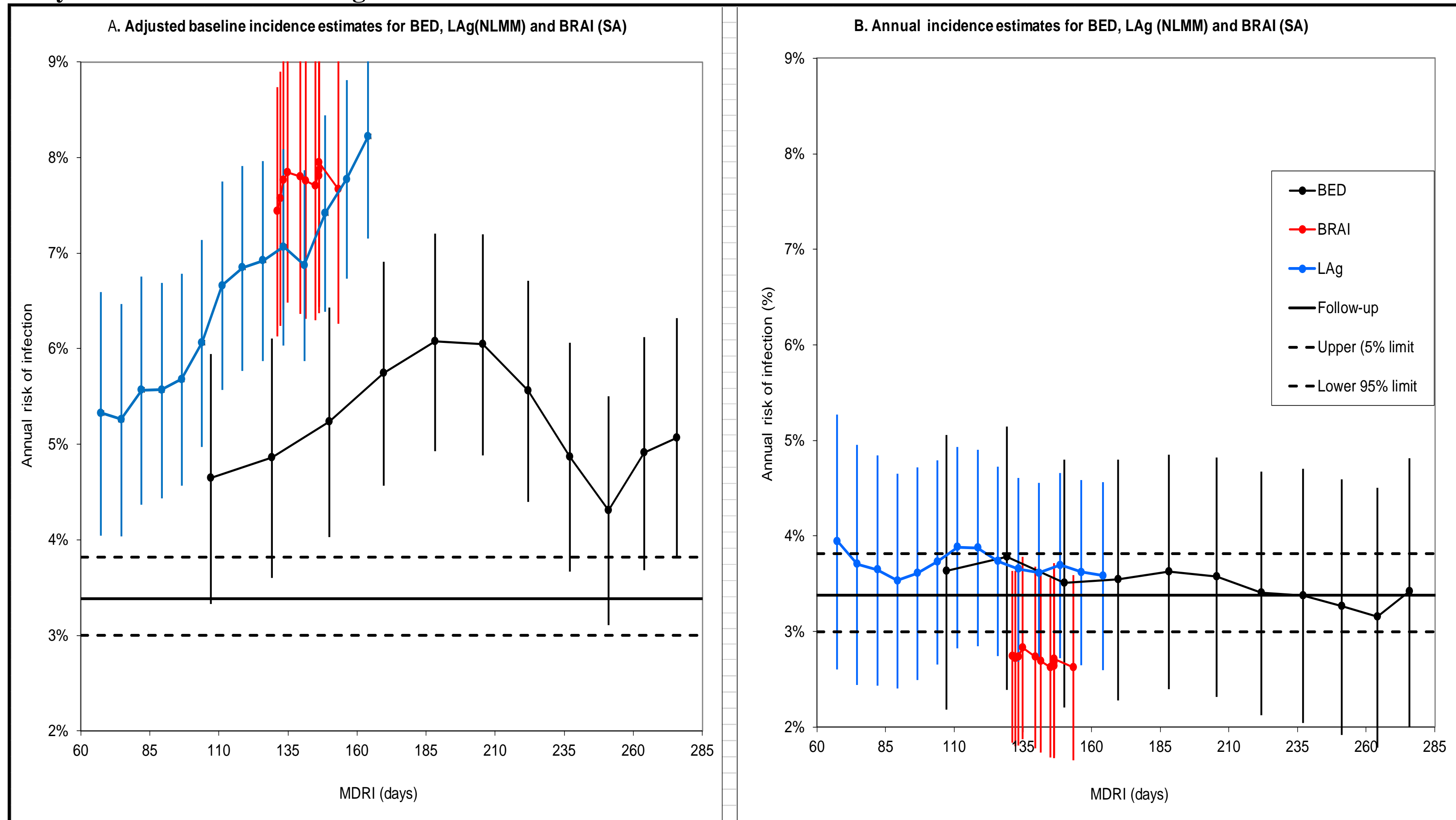
Stellenbosch University https://scholar.sun.ac.za

10.5 Comparison of HIV incidence estimated from samples at Baseline and at 12-months Postpartum

We plotted the annual risk of HIV infection against varying MDRI for the three assays for HIV estimates made using samples collected either at baseline or at 12 months postpartum (Figure 10-15). HIV incidence is markedly higher, at all levels of MDRI, when estimated at Baseline than at 12-months postpartum. Moreover, particularly for BED and LAg estimates, incidence measured at baseline changes markedly as a function of MDRI. By contrast, incidence estimated from samples collected at 12-months postpartum shows much less variation with MDRI.

Stellenbosch University https://scholar.sun.ac.za

Figure 10-15: Comparison of adjusted baseline and postpartum incidence estimates obtained by application of the BED, LAg and BRAI assays to the ZVITAMBO original data.



10.6 Summary HIV Incidence at Baseline and 12 Months Postpartum

Table 10-1: Summary of estimates for BED, LAg and BRAI serology (Original data)

		Serological biomarkers only				Incidence estimates			
						Postpartum		Baseline	
	C	MDRI (ci)	CoV (%)	FRR (ci)	CoV (%)	adjusted (ci)	unadjusted (ci)	adjusted (ci)	unadjusted
BRAI(SA)	30%	135	5.9	1.1%	18.2	2.6	3.9	8.1	9.1
		(120-151)		(0.7-1.5)		(1.6-3.7)	(2.9 -4.9)	(6.6-9.7)	(7.6-10.7)
BED (NLMM)	0.8	188	2.1	4.8%	8.5	2.7	6.2	6.7	10.1
		(180-196)		(4.1-5.7)		(1.5 - 4.0)	(5.0 - 7.3)	(5.5 -7.9)	(9.1-11.1)
LAg NLMM	1.5	104	2.9	0.6%	24.9	2.5	3.1	6.9	7.1
		(98-110)		(0.3-0.9)		(1.6 - 3.5)	(2.1-4.0)	(5.7 - 8.1)	(6.0-8.1)
						Follow-up			
						3.4			
						(3.0 -3.8)			

Table 10-2: Summary of estimates for BED, LAg and BRAI when VL is included (Original data)

		Serological biomarkers plus viral load				Incidence estimates			
						Postpartum		Baseline	
	C	MDRI (ci)	CoV (%)	FRR (ci)	CoV (%)	adjusted (ci)	unadjusted (ci)	adjusted (ci)	unadjusted
BRAI(SA)	30%	133	7.8	1.0%	18.5	2.7	3.9	7.3	8.4
		(113-154)		(0.7-1.5)		(1.5-4.0)	(2.9-4.9)	(5.7-8.8)	(6.8-10.0)
BED (NLMM)	0.8	177	6.4	3.8%	9.7	2.5	6.2	5.4	8.6
		(155-199)		(3.1-4.6)		(1.6-3.5)	(5.0-7.3)	(4.0-6.7)	(7.2-10.0)
LAg NLMM	1.5	101	7	0.4%	31.6	2.6	3.1	5.1	5.7
		(87-115)		(0.2-0.7)		(1.6-3.7)	(2.1-4.0)	(3.9-6.3)	(4.5-6.9)
						Follow-up			
						3.4			
						(3.0 -3.8)			

In summary, when only serological biomarkers were used to identify recent infections, the unadjusted HIV incidence for postpartum period derived for BRAI and LAg were numerically higher than the adjusted estimates (Table 10-1) although not significantly different (CI overlap). The BED unadjusted estimates were significantly higher than the adjusted estimates. BRAI unadjusted incidence, estimated from Visit 5 samples, were slightly higher (3.9%) than the ZVITAMBO follow-up rate, whereas the adjusted incidence was slightly lower (2.7%). For the Lag, unadjusted incidence (3.1%) was higher and the adjusted incidence (2.6%) was lower than the follow-up value.

Stellenbosch University https://scholar.sun.ac.za

At baseline, where there is no follow-up rate for comparison, the adjusted incidence estimate for BRAI (9.1%) was numerically, but not significantly, higher than for BED (10.1%) and LAg (7.1%). The incidence estimates derived from samples collected at baseline provide a proxy measure for HIV acquisition during pregnancy and so we expect the incidence rates to be different.

The addition of viral load in multi-assay analysis affects the data by removing some of the false assay positives (recent) from the numerator. The incidence estimates are generally lower than when only the serological biomarker is used to identify recent infections (cf Tables 10-1 and Table 10-2). This analysis is based on the original data.

Stellenbosch University https://scholar.sun.ac.za

10.7 Discussion of HIV incidence estimates

In order to evaluate the performance of the new avidity assays, LAg and BRAI, we measured the MDRI and FRR of each assay, and used these parameters to calculate HIV incidence using blood samples collected during the ZVITAMBO cohort. We compared the HIV incidence, derived using these parameters, to the prospective cohort follow-up rate, measured over the first 12-months postpartum and reported by the ZVITAMBO project in 2006.[32] Given that there was large overlap in the data used to make the two sets of estimates, we anticipate that the two approaches should give very similar estimates. Agreement between the estimates does not, therefore, constitute validation of the biomarker estimates. On the other hand, differences between the sets of estimates may be informative about possible problems with the biomarker approach.

10.7.1 Using serological biomarkers only to identify recent infections

When we used the original ZVITAMBO diagnoses for HIV infection status, we found that, at the recommended cut-offs of 30% and 40% AI, BRAI has MDRIs of 135 days (95% CI 120 - 151) and 144 days (128 - 160), respectively. The LAg has MDRI 104 days (95% CI 98 - 110) at cut-off 1.5 and 141 days (134-148) at cut-off 2.0. The coefficient of variation (CV) of these MDRI estimates for BRAI were 5.9% and 5.6% and those for LAg were 2.9% and 3.4%, respectively. The FRR associated with these cut-offs were 1.1% and 1.4 for BRAI and 0.6% and 1.0% for LAg.

When we used these parameters at cut-off 30% AI for BRAI and 1.5% for LAg, the estimated, unadjusted, annual rates of postpartum HIV incidences were 3.9 (95% CI 3.0 -4.9) and 4.7% (95% CI 3.7-5.7), respectively. The adjusted rates were 2.9% (95% CI 1.9-3.9) and 3.8% (95% CI 2.7-4.9), respectively. The 95% confidence intervals for these adjusted estimates overlapped the follow-up incidence estimate of 3.4% (95% CI 3.0 -3.8%) and did not thus differ significantly from that value. The BRAI estimate did however, lie below the lower 95% confidence band.

In earlier analyses of the ZVITAMBO data using the BED biomarker there were large, statistically significant, differences between the unadjusted and adjusted estimates. With the much-reduced FRR values typical of the BRAI and LAg

Stellenbosch University https://scholar.sun.ac.za

biomarkers the differences between unadjusted and adjusted postpartum HIV incidences for BRAI and LAg were also much reduced. Nonetheless, the adjusted estimates were closer to the follow-up estimates.

Given the uncertainty about the HIV status of a small number of samples, we repeated the analyses using the suggested new diagnoses. This involved reclassifying 18 cases at Baseline, and 17 cases at Visit 5, where the initial diagnosis was HIV positive and we considered that there was sufficient evidence to change the diagnosis to HIV negative. When we used this new/revised data set, we found that the incidence measured at Visit 5 using LAg declined substantially from the 3.80% quoted above, to 2.85% (95% CI 1.85-3.87). Again, the 95% confidence interval overlaps the mean value of 3.80%. Nonetheless the point estimate obtained using the new diagnoses is 1-2.85/3.80 = 25% lower than the estimate obtained using the original data.

At first sight, this appears a large change given the small proportion of reclassified. The reason for the large effect is that the HIV incidence estimator is very sensitive to small changes in the number of recent cases. The effect is accordingly large, because all of the case reclassified were not only HIV positive cases, but were all also classified as “recent” infections.

The BRAI postpartum adjusted estimate remained approximately unchanged. Using the old, and new, diagnoses the estimates were 2.88% (95% CI 1.87-3.90) and 2.86% (95% CI 1.85-3.87), respectively. This may appear contradictory, given the relatively larger change seen when using LAg. The reason here is that, in using the BRAI method for estimating the incidence, we were never able to make sensible use of cases that tested “invalid”. When we assumed that the “original” diagnoses were correct, we simply had to ignore any cases that tested “invalid”: we were assuming that the case was HIV positive, but we had no information to decide whether the case was a “recent” or a “long-term” infection. When we assumed that the “new” diagnoses were correct, there was no need to test the samples with BRAI.

It will be clear from the previous paragraph that the number classified as “recent”, using BRAI, is the same whichever data set we choose to accept as correct. The only difference between the two analyses lies in the total numbers of cases classified as

Stellenbosch University https://scholar.sun.ac.za

HIV positive and negative. Since these numbers are much larger than the numbers classified as “recent”, the resulting HIV incidence estimates differ little, whether we use the “original” or the “new” data.

10.7.2 Using serological biomarkers together with viral load to identify recent infections

We found that addition of viral load resulted in slight declines in incidence as measured using BRAI (-0.1%) and LAg (-1.2%). When we use both serology and VL <1000 copies to identify recent infections, the issue of discordant HIV serology becomes irrelevant, regardless of whether we use BED, LAg or BRAI. This is because all of the discrepant cases had very low or undetectable viral load. Thus, in the event that we used the “original” diagnoses, all of these cases were categorised as “long-term” infections, regardless of the level measured using the serological biomarker. Alternatively, if we used the new diagnoses, they were classified as HIV negative. In neither case, therefore, were any of them classified as “recent” (HIV positive) cases. The difference in the outcome, between the use of original or new data sets, thus rests on relatively small changes in the numbers of either HIV negative, or HIV positive long-term, infections.

Using viral load as an initial screening tool ensures that samples that are HIV serology negative, suppressed on antiretroviral therapy or are elite controllers are not considered in the overall assay data (since laboratory assay are only applied to HIV positive samples) and subsequently they are removed from the numerator as false positives. These findings provide evidence in favour of the inclusion of VL as part of an MAA in the estimation of HIV incidence using cross-sectional survey samples. Inclusion of VL would guard against situations where significant proportions of samples are misclassified as HIV positive. Inclusion of VL would also be useful in epidemiological studies of risk factors for seroconversion and especially where individualised treatment or partner notification based on HIV incidence results are implemented. The Swaziland HIV Incidence Measurement Survey (SHIMS) study showed that using VL results in HIV incidence results that are similar to follow-up estimates. This finding is further supported by recent publications by Kassanjee et al. (2016) [103]

Stellenbosch University https://scholar.sun.ac.za

10.7.3 High HIV incidence measured at Baseline

While there are no independent estimates of HIV incidence over the pregnancy period with which to compare the baseline biomarker-based HIV incidence estimates, these results are useful in showing the risk of infection by physiological status of the individual. The baseline incidence pertains to the risk of HIV acquisition during pregnancy. The adjusted HIV incidence at baseline of 8.1% per annum (95% CI 6.6 - 9.7) for BRAI and 6.9% per annum (95% CI 5.7 - 8.1) for LAg are significantly higher than the follow-up level of 3.4% per annum, and the adjusted biomarker rates measured 12 months postpartum.

The high baseline adjusted incidence estimates suggest a high rate of seroconversion during pregnancy The condition of pregnancy is characterised by immune suppression that facilitates embryonic implantation, its maintenance and growth of the foetus as a foreign body.[94] The higher baseline HIV incidence supports this hypothesis. An alternative explanation for the high baseline HIV incidence is that the true levels of MDRI and FRR during pregnancy are markedly higher than the estimates we estimated from samples collected in the postpartum period.

The addition of viral load resulted in a decrease in unadjusted and adjusted baseline estimates for all three assays, but most markedly for the BED assay, consistent with the fact that BED has a higher assay FRR.

While evaluations have focused on the performance of a laboratory assay in characterizing an infection as either recent or long term in the estimation of HIV incidence, Kassanjee et al. (2014) made a strong point that the focus should be on the accuracy of the final incidence estimator. The focus should not be on the laboratory assay's performance in terms of sensitivity/specificity.[20] The similarity of calculated unadjusted and adjusted annual incidence at 12 months postpartum measured by BRAI and LAg to the ZVITAMBO prospective cohort estimate support the application of these assays for measuring incidence among postpartum women.

Stellenbosch University https://scholar.sun.ac.za

11.0 Limitations of the study

The most serious concern we have about our study relates to the uncertainty of the HIV diagnoses in a number of samples collected during the ZVITAMBO study. One should, however, put these concerns in perspective. At baseline, there is a suggestion that we should reclassify 18 cases as HIV originally diagnosed as HIV negative: this is only 0.4% of the 4495 originally classified as HIV positive at baseline. The problem is that we must classify all of the disputed cases are as "recent" infections if we decide that they are indeed classified as HIV positive. We could not resolve this dilemma with the data available – particularly as we had no way of doing any further testing of viral loads.

We note, further, that the samples used in this study were collected in the ZVITAMBO study 1997-2000. Although we conducted viability tests in which we compared the BED optical density readings on specimens tested in 2013-2014 against readings obtained in the 2006 BED evaluation and found that the correlation was good, we cannot rule out the possibility that some of the HIV serology classifications were affected by the long-term storage.

Stellenbosch University https://scholar.sun.ac.za

12.0 Conclusion

We observed significantly lower MDRI and FRR values for LAg and BRAI when we applied these avidity assays to ZVITAMBO samples as compared to evaluations conducted on Clade C samples from the general population. At the recommended cut-offs, BRAI has a longer MDRI (141days) than LAg (104 days) but, conversely, the FRR was 1.1% for BRAI and only 0.6% for LAg.

The primary goal of a biomarker assay is to classify an infection accurately either as a recent, or as a long-term, infection – as defined relative to a predefined ODn/AI cut-off. To this end, the kinetics of an assay are an important predictor of its performance when applied to population level samples. The major difference in the two assays is the variations in plots of avidity index versus the cut-off in BRAI when compared to plots of optical density versus cut-off for LAg. For the BRAI assay, while the expected progression of the avidity is a plateauing at 100%, we observed that in some instances, AI, increased to a maximum then returned to below the cut-off AI. This reversal in trend is potentially problematic in that we will misclassify as recent a sample taken at the point where there is a reversal of AI to below the cut-off. The plots of variation in AI/OD versus cut-off are important in showing the capacity of assay in tracking biological changes as indicated by increase in ODn/AI as measures of increase in antigen/antibody with infection

Another important finding in this evaluation was that when we used different statistical methods to calculate MDRI, the NLMM method provided more precise estimates for the LAg assay than SA or binomial regression analyses. When SA was applied to both new biomarkers, the MDRI for LAg at cut-off 1.5 (112 days 95% CI 99-125 CV 6.0%) and cut-off 2.0 (137days 95% CI 122 -154, CV 6.0%) had narrower CI and slightly lower CV than BRAI at 30% (141days 95% CI 122-166 CV 6.9%) or at 40% (153 days 95% CI 134 -173 CV 6.6%).

In our analysis of MDRI for LAg and BRAI, the results suggested that beyond HIV subtype, the host's humoral response might be a key determinant in the estimation of MDRI. When we compared the MDRI in this evaluation to those published for Clade C samples from the general population, our point estimates of MDRI at similar cut-

Stellenbosch University https://scholar.sun.ac.za

offs were lower for both BRAI and LAg. Further work is critical in establishing more conclusively whether MDRIs among postpartum women are indeed consistently shorter than for women in the general populations. Alternatively is our result peculiar to our study population, or perhaps only to women in Zimbabwean? Or do other factors such as genetic make-up, or even ethnicity, have a role in the performance of laboratory assays? Until we have answers to these questions we cannot know to what extent our MDRI and FRR estimates can be used in using BRAI and LAg to estimate HIV incidence from samples collected in Zimbabwe, or indeed further afield. More generally, we need to know whether our findings are in any way generalizable to people other than postpartum women.

In countries experiencing a generalised epidemic and simultaneously scaling up antiretroviral treatment programs, an assay with a low FRR is highly desirable. The lower FRR at cut-off 1.5 for LAg compared to BRAI at cut-off 30% AI favour the use of LAg in population surveys such as HIV Impact Assessments that are currently being rolled out in Sub-Saharan Africa. We note, however, that at similar MDRI of 141 days, the FRR for both LAg and BRAI were 1.1%.

While the adjusted and unadjusted rates of annual risk of infection for BRAI and adjusted estimate for LAg closely approximate the follow-up ZVITAMBO rate, the correct classification of HIV status remains a critical piece in incidence estimation. At a population level, the HIV serology misclassification may result in either over- or under-estimation of HIV incidence. The development of fourth generation laboratory assays that have dual capabilities to both screen for HIV serology and classify as recent or long term is critical, but rigorous evaluations are required before we can use these in avidity assays. A benefit of a fourth generation assay is that it could detect cases that are antibody negative but have p24 antigen, which may bind avidly to the patient's p24 antigen. However, this may provide confounding results in that there is detection of both antibody and antigen in one compartment. By adding p24 antigen, one increases the window of detection at the left side (very early infections) by adding perhaps 1-2 weeks. One may, however, lose in terms of the correct identification of recent infections, and this problem is compounded if MDRI is shorter and/or if the FRR is bigger, as observed in the evaluation of fourth generation assay for incidence testing [98]

Stellenbosch University https://scholar.sun.ac.za

The ZVITAMBO results show that LAg + VL algorithm gives incidence estimates at 2-month postpartum that are similar to the observed incidence, without any further adjustment (FRR=0.0). This is an interesting finding and similar to what was observed in the Swaziland SHIMS study where the combined use of LAg + VL resulted in unadjusted HIV incidence estimates that were close to the observed follow-up estimates.[104] At a population level, the combined use of serology and VL biomarkers should likely result in improvements in the accuracy of HIV incidence estimates. One view is that the combined use of two biomarkers lowers the FRR to the point that further adjustment leads to under-estimation of the incidence and that therefore it is inappropriate to adjust for FRR. On this argument, we could regard the unadjusted incidence estimates as adequate. This is, however, a dangerous argument, based just on a correspondence between biomarker and follow-up incidence estimates that might be entirely coincidental. In our own study, we note that – when we used VL testing in conjunction with serological biomarkers – we see a reasonable correspondence, for both BRAI and LAg biomarkers at their recommended cut-offs, between unadjusted biomarker estimates and follow-up estimates. This matter requires further investigation.

We do note that adjusted incidence estimate at 12 months postpartum, using biomarkers alone, are largely independent of the cut-off used. This is not the case with unadjusted estimates, whether we use serological biomarkers by themselves or in tandem with viral load: estimated incidence increases with increasing cut-off. This is consistent with the fact that, as we increase the cut-of, we incorrectly classify an increasing number of cases as recent infections. The adjusted incidence estimates are more in tune with follow-up estimates. When measured over different periods postpartum in the ZVITAMBO Trial, seroconversion rates, measured using follow-up testing, were relatively constant over the first 24 months postpartum[1].

Baseline incidence estimates were higher, at all values of C, than estimates of incidence over the first 12 months postpartum. We ascribe this, provisionally, to a generally heightened risk of HIV infection during pregnancy. The adjusted incidence estimates also increased with increasing values of the pre-set cut-off C and thus of the MDRI. This, in turn we could explain if risk of infection were lower towards the end

Stellenbosch University https://scholar.sun.ac.za

of pregnancy. If that were the case, using higher values of C would result in measurement of HIV incidence over a longer time-frame prior to parturition and would thus involve women at progressively higher risk

In conclusion, the much-reduced FRR associated with BRAI and LAg avidity assays mark a major improvement in performance relative to assays such as the BED. The major difference between the performance of the BRAI and LAg assays, as measured only on the basis of the ZVITAMBO evaluation exercise, stems from the greater variability in the pattern of increase in BRAI AI than in LAg normalised OD. This implies better precision of LAg estimates compared to BRAI. We also note that our assessment of the relative merits of the LAg and BRAI methods is deliberately limited here to the results arising from the current evaluation exercise. Issues such as cost, and cost-effectiveness, are outside scope of this evaluation. Regardless of close approximations in the final incidence rates, the kinetics displayed by BRAI suggest the need to invest in improvements of the assay's performance compared to the LAg assay. Whichever method is used in the field it will be necessary to ensure that appropriate values of MDRI and FRR are chosen for the population being studied.

Stellenbosch University https://scholar.sun.ac.za

13.0 Recommendations from evaluation

This evaluation has successfully characterised the performance of LAg and BRAI in terms of their MDRI and FRR using samples from postpartum women in Harare, Zimbabwe. Based on this evaluation we recommend that:

1. Where possible MDRIs and FRRs should be estimated in other populations of postpartum women, in Zimbabwe and elsewhere, to establish the generalisability of our findings.
2. Similarly, MDRIs and FRRs should be estimated for LAg and BRAI in Zimbabwe using cohorts in the general population in order to provide values appropriate for the local population.
3. The MDRI and FRR parameter estimates we derived should be used in the estimation of HIV incidence from other cohorts of postpartum pregnant women, in Zimbabwe and elsewhere, in order to provide further testing of the validity of our estimates.
4. The MDRI and FRR values can and should be used for estimating HIV incidence among postpartum populations in Zimbabwe
5. There is need to continue exploration of the combined use of LAg and BRAI laboratory assays and Viral Load in estimation of incidence within the context of Surveillance.

Stellenbosch University https://scholar.sun.ac.za

15.0 Glossary

Term	Definition
Avidity Index	The factor obtained by dividing the optical density of the well containing the dissociating agent by the optical density of the untreated well. This is multiplied by 100 and presented as a percentage
Test for recent infection (TRI)	A laboratory test or combination of tests and the supplementary clinical information used to classify an HIV infection as recent or not recent
Mean duration of recent infection (MDRI) (ΩT) **previously known as "window period"**	The mean duration of recent HIV infection (MDRI) is defined as the average time that a person spends and is classified as recent by a given assay, for less than a predefined time (T)
False recent rate (FRR)	The proportion of individuals infected for greater than time (T set at 365 days) who are misclassified as recently infected when in fact they have been HIV positive for a longer time.

Stellenbosch University https://scholar.sun.ac.za

References

1. UNAIDS. *Trends in HIV incidence and prevalence: natural course of the epidemic or results of behavioural change?*. (1999).
2. UNAIDS AND WHO. When and How to Use Assays for Recent Infection to Estimate HIV Incidence at a Population Level. UNAIDS and WHO. (2013). Available at: http://www.aidsdatahub.org/when-and-how-to-use-assays-for-recent-infection-to-estimate-hiv-incidence-at-a-population-level-unaids-and-who. (Accessed: 18th June 2015)
3. Brookmeyer, R. Measuring the HIV/AIDS epidemic: approaches and challenges. *Epidemiol. Rev.* **32,** 26–37 (2010).
4. Ghys, P. D., Kufa, E. & on Estimates, M. V. G. for the U. R. G. Measuring trends in prevalence and incidence of HIV infection in countries with generalised epidemics. *Sex. Transm. Infect.* **82,** i52–i56 (2006).
5. Eaton, J. W. *et al.* Recent HIV prevalence trends among pregnant women and all women in sub-Saharan Africa: implications for HIV estimates. *AIDS Lond. Engl.* **28,** S507–S514 (2014).
6. Ben Lopman, S. G. When Did HIV Incidence Peak in Harare, Zimbabwe? Back-Calculation from Mortality Statistics. *PloS One* **3,** e1711 (2008).
7. Hallett, T. B. *et al.* Estimates of HIV incidence from household-based prevalence surveys. *AIDS Lond. Engl.* **24,** 147–152 (2010).
8. Rosenberg PS and Biggar RJ. Trends in HIV Incidence among young adults in the United States. *JAMA* **279,** 1894–1899 (1998).
9. Rutherford, G. W. *et al.* Public health triangulation: approach and application to synthesizing data to understand national and local HIV epidemics. *BMC Public Health* **10,** 447 (2010).

Stellenbosch University https://scholar.sun.ac.za

10. Incidence Assay Critical Path Working Group. More and Better Information to Tackle HIV Epidemics: Towards Improved HIV Incidence Assays. *PLoS Med* **8,** e1001045 (2011).

11. Le Vu, S. *et al.* Principles and uses of HIV incidence estimation from recent infection testing--a review. *Euro Surveill. Bull. Eur. Sur Mal. Transm. Eur. Commun. Dis. Bull.* **13,** (2008).

12. Brookmeyer, R. & Quinn, T. C. Estimation of current human immunodeficiency virus incidence rates from a cross-sectional survey using early diagnostic tests. *Am. J. Epidemiol.* **141,** 166–172 (1995).

13. Busch, M. P. *et al.* Beyond detuning: 10 years of progress and new challenges in the development and application of assays for HIV incidence estimation: *AIDS* **24,** 2763–2771 (2010).

14. Parekh, B. S. *et al.* Quantitative Detection of Increasing HIV Type 1 Antibodies after Seroconversion: A Simple Assay for Detecting Recent HIV Infection and Estimating Incidence. *AIDS Res. Hum. Retroviruses* **18,** 295–307 (2002).

15. Wei, X. *et al.* Development of two avidity-based assays to detect recent HIV type 1 seroconversion using a multisubtype gp41 recombinant protein. *AIDS Res. Hum. Retroviruses* **26,** 61–71 (2010).

16. Granade, T. C., Nguyen, S., Kuehl, D. S. & Parekh, B. S. Development of a novel rapid HIV test for simultaneous detection of recent or long-term HIV type 1 infection using a single testing device. *AIDS Res. Hum. Retroviruses* **29,** 61–67 (2013).

Stellenbosch University https://scholar.sun.ac.za

17. Smoleń-Dzirba, J. & Wąsik, T. J. Current and future assays for identifying recent HIV infections at the population level. *Med. Sci. Monit. Int. Med. J. Exp. Clin. Res.* **17,** RA124–133 (2011).

18. Parekh, B., Hu, D. J., Vanichseni, S., Satten, G. & Candal, D. Evaluation of a sensitive/less-sensitive testing algorithm using the 3A11-LS assay for detecting recent HIV seroconversion among individuals with HIV-1 subtype B or E infection in Thailand. *AIDS Res Hum Retroviruses* **17,** 453–458 (2001).

19. Young, C. L. *et al.* Evaluation of a sensitive/less sensitive testing algorithm using the bioMérieux Vironostika-LS assay for detecting recent HIV-1 subtype B' or E infection in Thailand. *AIDS Res. Hum. Retroviruses* **19,** 481–486 (2003).

20. Kassanjee, R., McWalter, T. A. & Welte, A. Short Communication: Defining Optimality of a Test for Recent Infection for HIV Incidence Surveillance. *AIDS Res. Hum. Retroviruses* **30,** 45–49 (2013).

21. Hargrove, J., Eastwood, H., Mahiane, G. & van Schalkwyk, C. How Should We Best Estimate the Mean Recency Duration for the BED Method? *PLoS One* **7,** e49661 (2012).

22. SEDIA TM HIV-1 LAg-Avidity EIA,Sedia Biosciences Corporation, Portland, Oregon, USA. Sedia™ HIV-1 LAg-Avidity EIA. (2011). Available at: http://www.sediabio.com/products/lag-avidity-eia. (Accessed: 23rd April 2016)

23. SEDIA TM HIV-1 LAg-Avidity EIA,Sedia Biosciences Corporation, Portland, Oregon, USA. Sedia™ HIV-1 LAg-Avidity EIA for DBS. (2011). Available at: http://www.sediabio.com/products/lag-avidity-eia-for-dbs. (Accessed: 23rd April 2016)

Stellenbosch University https://scholar.sun.ac.za

24. Bio-Rad Genetic Systems HIV-1/HIV-2 plus O Enzyme-Linked Avidity Incidence Assay. Catalogue 32588 & Modified by Centers for Disease Control and Prevention. Atlanta GA.USA.

25. Humphrey, J. H. *et al.* HIV incidence among post-partum women in Zimbabwe: risk factors and the effect of vitamin A supplementation. *AIDS Lond. Engl.* **20,** 1437–1446 (2006).

26. Humphrey, J. H. *et al.* Mother to child transmission of HIV among Zimbabwean women who seroconverted postnatally: prospective cohort study. *BMJ* **341,** c6580 (2010).

27. Hargrove, J. W. *et al.* Improved HIV-1 incidence estimates using the BED capture enzyme immunoassay. *AIDS Lond. Engl.* **22,** 511–518 (2008).

28. Mahy, M., Garcia-Calleja, J. M. & Marsh, K. A. Trends in HIV prevalence among young people in generalised epidemics: implications for monitoring the HIV epidemic. *Sex. Transm. Infect.* **88,** i65–i75 (2012).

29. Gouws, E. & The International Group on Analysis of Trends in HIV prevalence and Behaviours in Young People in Countries Most Affected by HIV. Trends in HIV prevalence and sexual behaviour among young people aged 15–24 years in countries most affected by HIV -- 86 (Suppl 2): ii72 -- Sexually Transmitted Infections. (2010). Available at: http://sti.bmj.com/content/86/Suppl_2/ii72.full. (Accessed: 23rd April 2016)

30. Avenir Health. Available at: http://www.avenirhealth.org/software-spectrum. (Accessed: 23rd April 2016)

31. Heaton, L. M., Komatsu, R., Low-Beer, D., Fowler, T. B. & Way, P. O. Estimating the number of HIV infections averted: an approach and its issues. *Sex. Transm. Infect.* **84,** i92–i96 (2008).

Stellenbosch University https://scholar.sun.ac.za

32. Garnett, G. P., Cousens, S., Hallett, T. B., Steketee, R. & Walker, N. Mathematical models in the evaluation of health programmes. *The Lancet* **378,** 515–525 (2011).

33. Hallett, T. B., Gregson, S., Mugurungi, O., Gonese, E. & Garnett, G. P. Assessing evidence for behaviour change affecting the course of HIV epidemics: a new mathematical modelling approach and application to data from Zimbabwe. *Epidemics* **1,** 108–117 (2009).

34. Mahiane, G., Ouifki, R., Brand, H., Delva, W. & Welte, A. A General HIV Incidence Inference Scheme Based on Likelihood of Individual Level Data and a Population Renewal Equation. *PLOS ONE* (2012).

35. Borquez, A. *et al.* The incidence Patterns Model to Estimate the Distribution of New HIV Infections in Sub-Saharan Africa:Development and Validation of a Mathematical Model. *PLoS Med.* **13,**

36. Rehle, T. M. *et al.* A decline in new HIV infections in South Africa: estimating HIV incidence from three national HIV surveys in 2002, 2005 and 2008. *PloS One* **5,** e11094 (2010).

37. Mehendale, S. M. *et al.* Low HIV-1 incidence among married serodiscordant couples in Pune, India. *J. Acquir. Immune Defic. Syndr. 1999* **41,** 371–373 (2006).

38. Cohen, M. S. *et al.* Prevention of HIV-1 Infection with Early Antiretroviral Therapy. *N. Engl. J. Med.* **365,** 493–505 (2011).

39. Chen, Y. Q. *et al.* Statistical considerations for the HPTN 052 Study to evaluate the effectiveness of early versus delayed antiretroviral strategies to prevent the sexual transmission of HIV-1 in serodiscordant couples. *Contemp. Clin. Trials* **33,** 1280–1286 (2012).

Stellenbosch University https://scholar.sun.ac.za

40. Ruzagira, E. *et al.* Prevalence and Incidence of HIV in a Rural Community-Based HIV Vaccine Preparedness Cohort in Masaka, Uganda. *PLoS ONE* **6,** (2011).

41. Beyrer, C. *et al.* Measuring HIV-1 incidence in northern Thailand: prospective cohort results and estimates based on early diagnostic tests. *J. Acquir. Immune Defic. Syndr. Hum. Retrovirology Off. Publ. Int. Retrovirology Assoc.* **12,** 495–499 (1996).

42. Kunawararak, P. *et al.* The epidemiology of HIV and syphilis among male commercial sex workers in northern Thailand. *AIDS Lond. Engl.* **9,** 517–521 (1995).

43. M T Mbizvo, A. S. L. HIV seroconversion among factory workers in Harare: who is getting newly infected? *Cent. Afr. J. Med.* **43,** 135–9 (1997).

44. Ray, S., Latif, A., Machekano, R. & Katzenstein, D. Sexual behaviour and risk assessment of HIV seroconvertors among urban male factory workers in Zimbabwe. *Soc. Sci. Med.* **47,** 1431–1443 (1998).

45. Gregson, S. *et al.* Estimating HIV incidence from age-specific prevalence data: comparison with concurrent cohort estimates in a study of male factory workers, Harare, Zimbabwe. *AIDS Lond. Engl.* **12,** 2049–2058 (1998).

46. Corbett, E. L. *et al.* HIV incidence during a cluster-randomized trial of two strategies providing voluntary counselling and testing at the workplace, Zimbabwe. *AIDS Lond. Engl.* **21,** 483–489 (2007).

47. Lopman, B. *et al.* HIV incidence in 3 years of follow-up of a Zimbabwe cohort—1998–2000 to 2001–03: contributions of proximate and underlying determinants to transmission. *Int. J. Epidemiol.* **37,** 88–105 (2008).

Stellenbosch University https://scholar.sun.ac.za

48. Janssen, R., Satten, G., Stramer, S. & Rawal, B. New testing strategy to detect early HIV-1 infection for use in incidence estimates and for clinical and prevention purposes. *JAMA* **280,** 42–48 (1998).

49. Duong, Y. T. *et al.* Detection of Recent HIV-1 Infection Using a New Limiting-Antigen Avidity Assay: Potential for HIV-1 Incidence Estimates and Avidity Maturation Studies. *PLoS ONE* **7,** (2012).

50. Konikoff, J. *et al.* Performance of a Limiting-Antigen Avidity Enzyme Immunoassay for Cross-Sectional Estimation of HIV Incidence in the United States. *PLoS ONE* **8,** (2013).

51. Suligoi, B. *et al.* Identifying recent HIV infections using the avidity index and an automated enzyme immunoassay. *J. Acquir. Immune Defic. Syndr.* **32,** 424–428 (2003).

52. Schupbach, J. & Gebhardt, M. . Assessment of recent HIV-1 infection by a line immunoassay for HIV-1/2 confirmation. PLoS Med. 2007 Dec;4(12):e343.

53. Parekh, B. S. *et al.* Determination of mean recency period for estimation of HIV type 1 Incidence with the BED-capture EIA in persons infected with diverse subtypes. *AIDS Res. Hum. Retroviruses* **27,** 265–273 (2011).

54. Hargrove, J. W. BED estimates of HIV incidence must be adjusted: *AIDS* **23,** 2061–2062 (2009).

55. Bärnighausen, T. *et al.* HIV Incidence in Rural South Africa: Comparison of Estimates from Longitudinal Surveillance and Cross-Sectional cBED Assay Testing. *PLoS ONE* **3,** (2008).

56. Edmore T Marinda, J. H. Significantly diminished long-term specificity of the BED capture enzyme immunoassay among patients with HIV-1 with very low CD4

Stellenbosch University https://scholar.sun.ac.za

counts and those on antiretroviral therapy. *J. Acquir. Immune Defic. Syndr. 1999* **53,** 496–9 (2010).

57. McNicholl, J. M. *et al.*, Thai-U.S. BED Assay Validation Working Group. Assessment of BED HIV-1 incidence assay in seroconverter cohorts: effect of individuals with long-term infection and importance of stable incidence. *PloS One* **6,** e14748 (2011).

58. Hallett, T. B., Ghys, P., Bärnighausen, T., Yan, P. & Garnett, G. P. Errors in 'BED'-Derived Estimates of HIV Incidence Will Vary by Place, Time and Age. *PLoS ONE* **4,** e5720 (2009).

59. Wendel, S. K. *et al.* Effect of Natural and ARV-Induced Viral Suppression and Viral Breakthrough on Anti-HIV Antibody Proportion and Avidity in Patients with HIV-1 Subtype B Infection. *PLoS ONE* **8,** (2013).

60. McDougal, J. S. BED estimates of HIV incidence must be adjusted. *AIDS Lond. Engl.* **23,** 2064–2065; author reply 2066–2068 (2009).

61. McWalter, T. A. & Welte, A. A comparison of biomarker based incidence estimators. *PloS One* **4,** e7368 (2009).

62. Kassanjee, R., McWalter, T. A., Bärnighausen, T. & Welte, A. A new general biomarker-based incidence estimator. *Epidemiol. Camb. Mass* **23,** 721–728 (2012).

63. Hargrove, J., van Schalkwyk, C. & Eastwood, H. BED Estimates of HIV Incidence: Resolving the Differences, Making Things Simpler. *PLoS ONE* **7,** e29736 (2012).

64. Hauser, A. *et al.* Improved Testing of Recent HIV-1 Infections with the BioRad Avidity Assay Compared to the Limiting Antigen Avidity Assay and BED Capture Enzyme Immunoassay: Evaluation Using Reference Sample Panels from the German Seroconverter Cohort. *PLOS ONE* **9,** e98038 (2014).

Stellenbosch University https://scholar.sun.ac.za

65. Hanson, D. L. *et al.* Mean Recency Period for Estimation of HIV-1 Incidence with the BED-Capture EIA and Bio-Rad Avidity in Persons Diagnosed in the United States with Subtype B. *PLOS ONE* **11,** e0152327 (2016).

66. Masciotra, S., Dobbs, T., Candal, D., Hanson, D. & Delaney, K. Antibody avidity-based assay for identifying recent HIV-1 infections Based on Genetic Systems [TM] 1/2 Plus O EIA [#937]. in (2010).

67. Moyo, S. *et al.* Evaluation of the False Recent Classification Rates of Multiassay Algorithms in Estimating HIV Type 1 Subtype C Incidence. *AIDS Res. Hum. Retroviruses* **30,** 29–36 (2013).

68. Andrew F Longosz, S. H. M. Incorrect identification of recent HIV infection in adults in the United States using a limiting-antigen avidity assay. *AIDS Lond. Engl.* **28,** (2014).

69. Yen T Duong, R. K. Recalibration of the Limiting Antigen Avidity EIA to Determine Mean Duration of Recent Infection in Divergent HIV-1 Subtypes. *PloS One* **10,** e0114947 (2015).

70. Brookmeyer, R., Konikoff, J., Laeyendecker, O. & Eshleman, S. H. Estimation of HIV Incidence Using Multiple Biomarkers. *Am. J. Epidemiol.* kws436 (2013). doi:10.1093/aje/kws436

71. Laeyendecker, O. *et al.* HIV incidence determination in the United States: a multiassay approach. *J. Infect. Dis.* **207,** 232–239 (2013).

72. Serhir, B. *et al.* Performance of Bio-Rad and Limiting Antigen Avidity Assays in Detecting Recent HIV Infections Using the Quebec Primary HIV-1 Infection Cohort. *PLOS ONE* **11,** (2016).

73. Kassanjee, R. *et al.* Independent assessment of candidate HIV incidence assays on specimens in the CEPHIA repository: *AIDS* **28,** 2439–2449 (2014).

Stellenbosch University https://scholar.sun.ac.za

74. Kim, A. A. *et al.* Estimating HIV incidence among adults in Kenya and Uganda: a systematic comparison of multiple methods. *PloS One* **6,** e17535 (2011).

75. SEDIA TM HIV-1 BED,INCIDENCE EIA,Sedia Biosciences Corporation, Portland, Oregon, USA. Sedia^{TM} BED HIV-1 Incidence EIA. (2009). Available at: http://www.sediabio.com/products/bed-eia. (Accessed: 23rd April 2016)

76. Erali, M. & Hillyard, D. . Evaluation of the Ultra Sensitive Roche Amplicor HIV-1 Monitor Assay for Quantitation of Human Immunodeficiency Virus Type 1 RNA. *J. Clin. Microbiol.* 792–795 (1999).

77. Sweeting, M., De Angelis, D., Parry, J. & Suligoi, B. Estimating the distribution of the window period for recent HIV infections: A comparison of statistical methods. *Stat. Med.* **29,** 3194–3202 (2010).

78. Iqbal, H. S., Solomon, S., Murugavel, K. G., Solomon, S. S. & Balakrishnan, P. Evaluation and Diagnostic Usefulness of Domestic and Imported Enzyme-Linked Immunosorbent Assays for Detection of Human Immunodeficiency Virus Type 1 Antibody in India. *Clin. Diagn. Lab. Immunol.* **12,** 1425–1428 (2005).

79. Bassey, O. *et al.* Evaluation of nine HIV rapid test kits to develop a national HIV testing algorithmn in Nigeria. *Afr. J. Lab. Med.* (2015).

80. Tchounga, B. K. *et al.* Re-testing and misclassification of HIV-2 and HIV-1&2 dually reactive patients among the HIV-2 cohort of The West African Database to evaluate AIDS collaboration. *J. Int. AIDS Soc.* **17,** (2014).

81. Delaney, K. P. *et al.* Evaluation of the Performance Characteristics of 6 Rapid HIV Antibody Tests. *Clin. Infect. Dis.* **52,** 257–263 (2011).

82. Everett, D. B. *et al.* Association of schistosomiasis with false-positive HIV test results in an African adolescent population. *J. Clin. Microbiol.* **48,** 1570–1577 (2010).

Stellenbosch University https://scholar.sun.ac.za

83. Bentsen, C. *et al.* Performance evaluation of the Bio-Rad Laboratories GS HIV Combo Ag/Ab EIA, a 4th generation HIV assay for the simultaneous detection of HIV p24 antigen and antibodies to HIV-1 (groups M and O) and HIV-2 in human serum or plasma. *J. Clin. Virol. Off. Publ. Pan Am. Soc. Clin. Virol.* **52 Suppl 1,** S57–61 (2011).

84. Pollyanna Chavez, L. W. Evaluation of the performance of the Abbott ARCHITECT HIV Ag/Ab combo assay. *J. Clin. Virol. Off. Publ. Pan Am. Soc. Clin. Virol.* **52 Suppl 1,** S51–5 (2011).

85. Constantine, N., Sill, A., Jack, N. & Kristen, K. Improved Classification of Recent HIV-1 Infection by Employing a Two-Stage Sensitive/Less-Sensitive Test Strategy. Available at: https://www.researchgate.net/publication/10963140_Improved_Classification_of_Recent_HIV-1_Infection_by_Employing_a_Two-Stage_SensitiveLess-Sensitive_Test_Strategy. (Accessed: 23rd April 2016)

86. Mitchell, E. O. *et al.* Performance comparison of the 4th generation Bio-Rad Laboratories GS HIV Combo Ag/Ab EIA on the EVOLIS™ automated system versus Abbott ARCHITECT HIV Ag/Ab Combo, Ortho Anti-HIV 1 + 2 EIA on Vitros ECi and Siemens HIV-1/O/2 enhanced on Advia Centaur. *J. Clin. Virol.* **58, Supplement 1,** e79–e84 (2013).

87. Warkentin, T. E., Sheppard, J. I., Moore, J. C., Sigouin, C. S. & Kelton, J. G. Quantitative interpretation of optical density measurements using PF4-dependent enzyme-immunoassays. *J. Thromb. Haemost. JTH* **6,** 1304–1312 (2008).

88. Fiebig, E. W. *et al.* Dynamics of HIV viremia and antibody seroconversion in plasma donors: implications for diagnosis and staging of primary HIV infection. *AIDS Lond. Engl.* **17,** 1871–1879 (2003).

Stellenbosch University https://scholar.sun.ac.za

89. Turnbull, B. The empirical distribution function with arbitrary groped, censored and truncated data. *J. R. Stat. Soc. Ser. B Methodol.* **38,** 290–295 (1976).

90. Braunstein, S. . *et al.* Dual testing algorithm of BED-CEIA and AxSYM Avidity Index assays performs best in identifying recent HIV infection in a sample of Rwandan sex workers. *PLOS ONE* **6,** (2011).

91. Kirkpatrick, A. R. *et al.* Development and Evaluation of a Modified Fourth Generation HIV Enzyme Immunoassay for Cross-Sectional Incidence Estimation in Clade B Populations. *AIDS Res. Hum. Retroviruses* (2016). doi:10.1089/AID.2015.0198

92. Hargrove, J. W. *et al.* Heightened HIV Antibody Responses in Postpartum Women as Exemplified by Recent Infection Assays: Implications for Incidence Estimates. *AIDS Res. Hum. Retroviruses* (2017). doi:10.1089/AID2016.0319

93. Groer, M. W. *et al.* Immunity, inflammation and infection in post-partum breast and formula feeders. *Am. J. Reprod. Immunol. N. Y. N 1989* **54,** 222–31 (2005).

94. Singh, N. & Perfect, J. R. Immune reconstitution syndrome and exacerbation of infections after pregnancy. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* **45,** 1192–9 (2007).

95. Iliff, P. J. *et al.* Early exclusive breastfeeding reduces the risk of postnatal HIV-1 transmission and increases HIV-free survival. *AIDS Lond. Engl.* **19,** 699–708 (2005).

96. Poole, J. A. & Claman, H. N. Immunology of pregnancy. Implications for the mother. *Clin. Rev. Allergy Immunol.* **26,** 161–70 (2004).

97. Moyo, S. Identifying Recent HIV Infections: From Serological Assays to Genomics (PDF Download Available). (2015). Available at:

Stellenbosch University https://scholar.sun.ac.za

https://www.researchgate.net/publication/283255189_Identifying_Recent_HIV_Infections_From_Serological_Assays_to_Genomics. (Accessed: 23rd April 2016)

98. UNAIDS/World Health Organisation. Technical update on HIV incidence assays for surveillance and monitoring purposes. (2015). doi:http://www.unaids.org/sites/default/files/media asset/HIVincidenceassayssurveillancemonitoring en.pdf

99. Incidence-Estimation. Available at: http://www.incidence-estimation.org/page/spreadsheet-tools-for-biomarker-incidence-surveys. (Accessed: 23rd March 2017)

100. Hudgens, M. G., Satten, G. A. & Longini, I. M. Nonparametric maximum likelihood estimation for competing risks survival data subject to interval censoring and truncation. *Biometrics* **57,** 74–80 (2001).

101. Gray, R. H. *et al.* Increased risk of incident HIV during pregnancy in Rakai, Uganda: a prospective study. *Lancet Lond. Engl.* **366,** 1182–1188 (2005).

102. Langford, S. E., Ananworanich, J. & Cooper, D. A. Predictors of disease progression in HIV infection: a review. *AIDS Res. Ther.* **4,** 11 (2007).

103. Kassanjee, R. *et al.* Viral load criteria and threshold optimization to improve HIV incidence assay characteristics. *AIDS Lond. Engl.* **30,** 2361–2371 (2016).

104. First Findings Released from Swaziland HIV Incidence Measurement Survey | Columbia University Mailman School of Public Health. Available at: https://www.mailman.columbia.edu/public-health-now/news/first-findings-released-swaziland-hiv-incidence-measurement-survey. (Accessed: 24th April 2016)