UCLA
UCLA Electronic Theses and Dissertations

Title
Circulating Markers of Immune Activation and Inflammation and AIDS-Associated Non-Hodgkin Lymphoma in the Multicenter AIDS Cohort Study (MACS)

Permalink
https://escholarship.org/uc/item/61x473c9

Author
Makgoeng, Solomon B.

Publication Date
2018

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA

Los Angeles

Circulating Markers of Immune Activation and Inflammation and AIDS-Associated Non-Hodgkin Lymphoma in the Multicenter AIDS Cohort Study (MACS)

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Epidemiology

by

Solomon Makgoeng

2018

© Copyright by

Solomon Makgoeng

2018

ABSTRACT OF THE DISSERTATION

Circulating Markers of Immune Activation and Inflammation and
AIDS-Associated Non-Hodgkin Lymphoma in the Multicenter AIDS Cohort Study (MACS)

by

Solomon Makgoeng

Doctor of Philosophy in Epidemiology

University of California, Los Angeles, 2018

Professor Shehnaz K. Hussain, Co-Chair

Professor Onyebuchi Aniweta Arah, Co-Chair

Background: AIDS-associated non-Hodgkin lymphoma (AIDS-NHL) remains a significant public health challenge among HIV-infected individuals. Chronic inflammation and immune activation have been documented in the literature to play a crucial role in the etiology of AIDS-NHL. We summarized results from prior work in a meta-analysis of the associations between prospectively measured circulating levels of immune biomarkers and the risk of NHL among both HIV-infected and HIV-uninfected populations. Our second study characterized the temporal variation in 24 pre-AIDS-NHL diagnosis circulating markers of inflammation and immune activation. Finally, we assessed the predictive ability of a set of 13 biomarker levels and AIDS-NHL diagnosis.

Methods: *Meta-analysis*: Our meta-analysis identified 17 relevant studies from inception of major biomedical databases (PubMed, EMBASE, and Web of Science) until January 1, 2017. We summarized published results using random-effects models for NHL and several histological subtypes of NHL. *Longitudinal study*: we summarized the slopes and means (intercepts) of biomarker trajectories using linear mixed models. *Prediction Models*: we calculated incremental discrimination ability (AUC) of models including biomarkers, individually and concurrently, relative to models including only participant characteristics and known risk factors of NHL.

Results: *Meta-analysis*: Summarizing 17 nested case–control studies, we found elevated levels of several biomarkers associated with increased odds of NHL overall: TNF-α, OR=1.18 [95% CI: 1.04, 1.34]; CXCL13, OR=1.47 [95% CI: 1.03, 2.08]; sCD23, OR=1.57 [95% CI: 1.21, 2.05]; sCD27, OR=2.18 [95% CI: 1.20, 3.98]; sCD30, OR=1.65 [95% CI: 1.22, 2.22]; AIDS-NHL showing stronger associations with IL-6, TNF-α, sCD27, and sCD30. *Longitudinal study*: prior to HAART, geometric mean biomarkers are elevated for cases relative to controls for IL-2, TNF-α, IL-6, sCD27, sIL-2Rα, IP-10, CXCL13, CRP, and pre-HAART slopes were observed to be higher for cases relative to controls for sIL-6R, sTNFR2, IL-10. Following HAART initiation, geometric mean levels are elevated to a higher degree than pre-HAART for cases relative to controls for BAFF, TNF-α, sIL-2Rα, sTNFR2, IP-10, MCP-1, CRP. *Prediction Modeling*: Models including individual biomarkers yielded modest improvements in AUC statistics above a base-case model that comprised NHL risk factors and other participant characteristics. A model including IL-6, IL-10, TNF-α, IP10, and CXCL13, concurrently performed better than all other models including individual biomarkers: 0-1: AUC=0.943 95% CI: 0.910, 0.975; 1-3: AUC=0.895 95% CI: 0.856, 0.934; >3: AUC=0.836 95% CI: 0.787, 0.885. Increments in AUC above the risk factors only model were also improved relative to individual biomarker models: 0-1: difference in

AUC=0.056 95% CI: 0.021, 0.091; 1-3: difference in AUC=0.032 95% CI: 0.007, 0.057; >3: difference in AUC=0.074 95% CI: 0.030, 0.118.

Conclusion: Each study provided further novel evidence of the association between circulating biomarkers and AIDS-NHL risk, as well as the utility of biomarkers in risk prediction. Our meta-analysis provides an overarching summary of evidence that elevated circulating levels of several markers are associated with an increased risk of NHL. Longitudinal analyses illustrate novel differences between cases and controls in aspects of the trajectories of 24 markers. Our prediction models elucidate the ability of a set of 13 marker levels to discriminate between AIDS-NHL cases versus controls. The totality of new evidence we provide supports the notion that chronic inflammation and immune activation is associated with increased AIDS-NHL risk, and that these biomarkers may have utility in the development of clinical risk prediction models.

The dissertation of Solomon Makgoeng is approved.

Roger Detels

Robert Erin Weiss

Shehnaz K. Hussain, Committee Co-Chair

Onyebuchi Aniweta Arah, Committee Co-Chair

University of California, Los Angeles

2018

TABLE OF CONTENTS

LIST OF TABLES ... viii
LIST OF FIGURES ... ix
ACKNOWLEDGMENTS ... x
VITA ... xi
1 CHAPTER I: Background & Introduction ... 1
1.1 Non-Hodgkin B-Cell Lymphoma ... 1
1.1.1 Disease Definitions and Classifications ... 1
1.1.2 Epidemiology of AIDS-NHL ... 1
1.2 Pathogenesis of AIDS-NHL ... 2
1.3 Risk Factors for Non-Hodgkin Lymphoma ... 3
1.3.1 Pathogens and Comorbid Factors ... 3
1.3.2 Demographic, Behavioral And Environmental Factors ... 4
1.4 The Multicenter AIDS Cohort Study ... 4
2 CHAPTER II: Circulating Markers of Immune Activation and Inflammation, and Non-Hodgkin Lymphoma: A Meta-Analysis of Prospective Studies ... 5
2.1 Abstract ... 5
2.2 Introduction ... 6
2.3 Materials And Methods ... 7
2.3.1 Literature Search Strategy ... 7
2.3.2 Inclusion and Exclusion Criteria ... 8
2.3.3 Data Items and Data Extraction Strategy ... 8
2.4 Data Analysis ... 9
2.4.1 Data Harmonization of Published Results ... 9
2.4.2 Estimation of Meta-Analytic Summary ORs ... 10
2.4.3 Estimation of Between-Study Heterogeneity ... 10
2.4.4 Assessment of Publication Bias & Influential Data ... 11
2.5 Results ... 12
2.5.1 Study Selection ... 12
2.5.2 Meta-Analyses ... 13
2.6 Discussion ... 17
2.7 Acknowledgements ... 24
2.8 Supporting Information ... 24
2.9 Disclosures ... 24
2.10 Tables and Figures ... 25
2.11 Supplementary Material ... 35
3 Chapter III: Temporal Variation of Serum Biomarkers of Immune Activation and Inflammation in Non-Hodgkin Lymphoma in the Multicenter AIDS Cohort Study ... 49
3.1 Abstract ... 49
3.2 Introduction ... 51
3.3 Materials And Methods ... 54
3.3.1 Study Design ... 54
3.3.2 Participant Inclusion ... 54
3.3.3 Biomarkers ... 56
3.3.4 AIDS-NHL Case Ascertainment ... 57
3.3.5 Covariates ... 58
3.3.6 Statistical Methods ... 58
3.4 Results ... 60

3.5 Discussion 62
3.6 Acknowledgements 69
3.7 Supporting Information 69
3.8 Disclosures 69
3.9 Tables 70
4 Chapter IV: Predictive Utility of Serum Biomarkers of Inflammation, Immune Activation, in AIDS-associated Lymphoma in HIV-infected Men in the MACS 78
4.1 Abstract 78
4.2 Introduction 80
4.3 Materials & Methods 82
4.4 AIDS-NHL Outcome, Biomarkers, and Covariates 82
4.5 Statistical Methods 83
4.6 Results 86
4.7 Discussion 88
4.8 Acknowledgements 92
4.9 Supporting Information 92
4.10 Disclosures 92
4.11 Tables and Figures 93
5 Chapter V: Public Health Significance & Conclusions 109
6 Chapter VI: References 111

LIST OF TABLES

Table 2.1: Characteristics of 17 Prospective Studies Included in the Meta-Analysis. 26
Table 2.2: Meta-Analysis Results for B cell NHL Overall and by HIV serostatus. 30
Table 2.3: Meta-Analysis Results for B Cell NHL Subtypes. 31
Supplementary Table 2.4: MOOSE Checklist for Meta-analyses of Observational Studies. 35
Supplementary Table 2.5: Bias Analyses: Egger's Regression P-values and Trim & Fill Analyses. 43
Supplementary Table 2.6: Results for All B-Cell NHL by Prediagnosis Time Interval: Comparing Early versus Late Biomarker Sample Collection..46
Supplementary Table 2.7: Results for All B-Cell NHL among HIV-infected: Comparing HAART exposed versus Unexposed 47
Supplementary Table 2.8: Multiple Comparisons Adjusted P-Values for Overall NHL.............48
Table 3.1 Participant Characteristics at Beginning of Follow-up (HIV-seroconversion) 70
Table 3.2: Summary Statistics of Analytes in Component Data Sets by Case Status Over Entire Follow-Up...72
Table 3.3: Geometric Mean Ratios for Cytokines: Mean Levels and Slopes. 74
Table 3.4: Geometric Mean Ratios for Soluble Receptors: Mean Levels and Slopes................75
Table 3.5: Geometric Mean Ratios for Chemokines: Mean Levels and Slopes 76
Table 3.6: Geometric Mean Ratios for Acute Phase Reactant: Mean Levels and Slopes..........77
Table 4.1: Participant Characteristics in Analytic Sample: N=358. 93
Table 4.2: Biomarker Distributions by Prediagnosis Time Point. 96
Table 4.3: Comparing AUC Estimates for Models including Risk Factors Only, versus Biomarker with Risk Factors..97
Table 4.4: Optimism Correction of AUC Estimates for Models including Biomarker with Risk Factors...99
Table 4.5: Goodness of Fit Tests Assessing Model Calibration 102

LIST OF FIGURES

Figure 2.1: Flowchart for Systematic Literature Search and Selection of Studies of Circulating Biomarkers and NHL Risk. 25
Figure 2.2: Forest Plots for Cytokines and Chemokine. 33
Figure 2.3: Forest Plots for Soluble Receptors. 34
Supplementary Figure 2.4: Forest Plots for NHL Subtype Outcomes. 37
Supplementary Figure 2.5: Funnel Plots Assessing Publication Bias. 41
Supplementary Figure 2.6: Influence Diagnostics: Leave-One-Out Analyses. 44
Figure 4.1: Calibration Plots for Models Including Biomarkers 103

ACKNOWLEDGMENTS

I would like to express gratitude to my dissertation committee members, Roger Detels, Robert Weiss, with special thanks to my co-chairs, Shehnaz Hussain and Onyebuchi Arah. In addition, Otto Martínez-Maza has been an especially valuable and much appreciated advisor. I have gained much through interaction with each one of them, and will be a much better researcher for having met them.

Chapter 2 is a version of Makgoeng SB, Bolanos RS, Jeon CY, Weiss RE, Arah OA, Breen EC, Martínez-Maza O, Hussain SK. Markers of Immune Activation and Inflammation, and Non-Hodgkin Lymphoma: A Meta-Analysis of Prospective Studies. *Submitted to JNCI Cancer Spectrum.* 2018. A special acknowledgement is due to all coauthors for conceptualizing the study, reading and contributing to writing multiple drafts, and to Rachel Bolanos for her tireless effort in assisting with data collection and verification.

I would also like to thank Zuo-Feng Zhang for his continued generous support as my mentor and PI on the National Institutes of Health/National Cancer Institute T32 Training Program in Molecular-Genetic Epidemiology of Cancer (T32CA009142). A special word of thanks is also due to the Fielding School of Public Health Department of Epidemiology, and Karin Michels for being a beacon of support at UCLA.

Lastly, I would also like to thank my fellow students, colleagues, my wife and my family who have made this journey all the more interesting.

VITA

2002 – 2006	Williams College, Williamstown, MA, Bachelor of Arts, Economics
2006 – 2007	Rochester Institute of Technology, Rochester, NY, Master of Science, Applied Statistics
2007 – 2008	Associate Analyst, National Economic Research Associates, New York, NY
2009 – 2009	Statistics Collaborative, Inc., Washington, DC, Statistical Programmer
2010 – 2013	Georgetown University, Lombardi Comprehensive Cancer Center, Washington, DC, Programmer Analyst II/Biostatistician I
2013 – 2018	University of California, Los Angeles, Fielding School of Public Health, Los Angeles, CA, PhD Student Researcher

PUBLICATIONS

1. Fulcher JA, Shoptaw S, **Makgoeng SB**, et al. Brief Report. *JAIDS J Acquir Immune Defic Syndr*. 2018;78(1):119-123.
2. Hussain SK, **Makgoeng SB**, et al. HLA and Risk of Diffuse Large B-cell Lymphoma After Solid Organ Transplantation. *Transplantation*. December 2015.
3. Stanton CA, Papandonatos GD, Shuter J, Bicki A, Lloyd-Richardson EE, de Dios MA, Morrow KM, **Makgoeng SB**, et al. Outcomes of a Tailored Intervention for Cigarette Smoking Cessation Among Latinos Living With HIV/AIDS. *Nicotine Tob Res*. 2015;17(8):975-982.

4. Hornbrook MC, Malin J, Weeks JC, **Makgoeng SB**, et al. Did Changes in Drug Reimbursement After the Medicare Modernization Act Affect Chemotherapy Prescribing? *J Clin Oncol.* 2014;32(36):4042-4049.

5. Mandelblatt JS, Huang K, **Makgoeng SB**, et al. Preliminary Development and Evaluation of an Algorithm to Identify Breast Cancer Chemotherapy Toxicities Using Electronic Medical Records and Administrative Data. *J Oncol Pract*. August 2014.

6. Mandelblatt JS, **Makgoeng SB**, et al. A planned, prospective comparison of short-term quality of life outcomes among older patients with breast cancer treated with standard chemotherapy in a randomized clinical trial vs. an observational study: CALGB #49907 and #369901. *J Geriatr Oncol*. 2013;4(4):353-361.

7. Loffredo CA, Luta, G., Wallington, S., **Makgoeng, SB**, et al. Knowledge and willingness to provide research biospecimens among foreign-born Latinos using safety-net clinics. *J Community Health*. 2013;38(4):652-659.

8. Kress M-AS, Ramakrishna N, **Makgoeng SB**, et al. Physician self-reported treatment of brain metastases according to patients' clinical and demographic factors and physician practice setting. *Radiat Oncol*. 2012;7(1):188.

9. Selsky C, Barbara K, Luta G, **Makgoeng SB**, et al. Hospice knowledge and intentions among Latinos using safety-net clinics. *J Palliat Med*. 2012;15(9):984-990.

10. Mandelblatt JS, Faul LA, Luta G, **Makgoeng SB**, et al. Patient and physician decision styles and breast cancer chemotherapy use in older women: Cancer and Leukemia Group B protocol 369901. *J Clin Oncol*. 2012;30(21):2609-2614.

11. Potosky AL, Malin JL, Kim B, Chrischilles EA, **Makgoeng SB**, et al. Use of colony-stimulating factors with chemotherapy: opportunities for cost savings and improved outcomes. *J Natl Cancer Inst*. 2011;103(12):979-982.

12. Mandelblatt JS, Luta G, Kwan ML, **Makgoeng SB**, et al. Associations of physical activity with quality of life and functional ability in breast cancer patients during active adjuvant treatment: the Pathways Study. *Breast Cancer Res Treat*. 2011;129(2):521-529.

1 CHAPTER I: Background & Introduction

1.1 Non-Hodgkin B-Cell Lymphoma

1.1.1 Disease Definitions and Classifications

Non-Hodgkin Lymphoma (NHL) comprises several histologically heterogeneous subtypes of malignant neoplasms of lymphocyte cells. The broadest taxonomic classification of these lymphoprolipherative neoplasms is whether they are derived from T-cells or B-cells, with the major of cases in the general population being of B-cell origin.[1,2] Additional levels of classification consider tumor growth rate (fast growing aggressive tumors versus slow growing indolent tumors) and location (systemic versus central nervous system) origin. In the context of AIDS-associated NHL (AIDS-NHL), NHL is further classified into AIDS-defining and non-AIDS-defining subtypes. These include diffuse large B cell lymphoma (DLBCL), primary central nervous system lymphoma (PCNSL), Burkitt's lymphoma, and primary effusion lymphoma (PEL).[3,4] The incidence of AIDS-NHL in the HAART era varies by both AIDS-defining and non-AIDS-defining subtypes,[5] although the overwhelming majority of cases were AIDS-defining with DLBCL and Burkitt lymphoma accounting for a majority of cases reported in a recent U.S. population based registry study.[6]

1.1.2 Epidemiology of AIDS-NHL

AIDS Non-Hodgkin lymphoma (AIDS-NHL) continues to be the most frequently diagnosed HIV-related malignancy in the U.S. and other developed countries even in the era of multi-agent highly active anti-retroviral therapy (HAART).[6,7] NHL also remains among the most significant causes of AIDS-related death.[7–9] The risk of developing AIDS-NHL has long been known to be significantly elevated in HIV-positive individuals relative to HIV-negative individuals an early

report showing up to a 60-fold multiple in incidence rate.[10] Although the risk of developing AIDS-NHL has decreased with effective HAART, disease risk remains elevated in those receiving HAART,[7,11,12] in addition, overall survival has improved with HAART receipt, survival outcomes still fall short of those of HIV-uninfected individuals.[13] In light of the excess risk for NHL among treated HIV positive individuals relative to the general population, research into etiologic factors contributing to HIV is needed.

1.2 Pathogenesis of AIDS-NHL

There are two primary putative pathogenic mechanisms responsible for AIDS-NHL.[3,4] The first involves the dysregulated proliferation of Epstein-Barr virus (EBV)-transformed B-cells where T-cell mediated regulation of B-cell growth has been impaired resulting in the development of EBV-positive AIDS-NHL subtypes.[3] The other mechanism involves chronic B-cell activation and resultant processes that promote oncogenic mutations and translocations.[4] There are several B-cell activation related mechanisms. First, chronic antigenic stimulation of B-cells by HIV infection itself may promote B-cell hyperactivation and transformation leading to AIDS-NHL.[4] B-cell genetic mutations due to chronic stimulation during other viral infections (including EBV, HPV and HCV), chromosomal rearrangements (BCL-6 and c-MYC) and deletions (6q), as well as mutations in RAS and p53 genes have been shown to be associated with chronic B-cell hyperactivation.[4,14–17] Another proposed mechanism involves HIV-infected macrophages contributing B-cell stimulatory signals that result in a B-cell activation and malignant B-cell growth.[14,18] Lastly, the leakage of bacteria and bacterial products from intestinal walls into the bloodstream (gut microbial translocation) is considered a potential cause of chronic immune activation, which includes increased serum levels of pro-inflammatory cytokines and chemokines and polyclonal B-cell activation, among HIV-infected individuals.[19,20] Specific mechanisms that are directly responsible for the intestinal cell depletion (enterocyte apoptosis)

that leads to the gut permeability and subsequent microbial translocation are unclear [20], although several potential hypotheses exist in the context of HIV infection that point towards increased enterocyte apoptosis commensurate with HIV viral activity including the virotoxic effect of HIV gp120 and the action of increased pro-inflammatory cytokines that are implicated in enteropathy.[20–23] Hence, microbial translocation can be considered to have a role in the etiology of NHL that is at least in part mediated through immune activation and inflammation. As such, we can expect associations between NHL and biomarkers involved in B-cell activation, such as cytokines, chemokines, their receptors, along with markers of gut microbial translocation.

1.3 Risk Factors for Non-Hodgkin Lymphoma

1.3.1 Pathogens and Comorbid Factors

Immune dysfunction and associated variables are the most consistently documented risk factors for AIDS-NHL among HAART users.[4] Low nadir CD4+ T cell count, high HIV viral load and duration of infection having been demonstrated to be associated with AIDS-NHL in the HAART era.[5,6,24] While AIDS-NHL cases attributable to oncogenic viral factors have diminished in the HAART era, there are reports of still higher virus-linked AIDS-NHL incidence relative to HIV uninfected individuals, and at least one report finding no difference in virus-linked AIDS-NHL incidence between HAART recipients and non-recipients.[3] Other morbidities reportedly associated with AIDS-NHL include obesity, which results in chronic low-level inflammation and immune responses that may influence B- and T-lymphocyte function therefore making way for NHL [25,26]; diabetes mellitus has also been shown to be associated with NHL in the general population [27,28] and is also thought to potentiate B-cell lymphoma via high levels of CCL5.[29]

1.3.2 Demographic, Behavioral And Environmental Factors

AIDS-NHL associations with demographic factors, particularly age, have been described in the general population as well as among HIV infected.[30–34] Behavioral risk factors have been explored in the MACS, with recent and 3-year lagged recreational drug use showing increased hazards of AIDS-NHL by multiples of 3-4 times for weekly or more frequent use[35]; the use of other recreational drugs was not significantly associated with AIDS-NHL in this study; tobacco use has shown no association with AIDS-NHL in an early study of the MACS[36] but recent trends in smoking rates suggest high tobacco use among lower socio-economic status subgroups[37] and smoking is a known risk factor for non-Hodgkin lymphoma in the general U.S. population.[38] There is some evidence linking environmental and occupational exposures to increased NHL risk in various populations.

1.4 The Multicenter AIDS Cohort Study

Details of the MACS have been described and published previously.[39,40] To give a brief description, the MACS is a prospective cohort study comprising men who have sex with men designed to investigate various aspects of HIV infection including the natural history of the disease, risk factors for acquiring disease, and the clinical expression of infection. Participants were recruited at four academic centers in the U.S. (Baltimore, Maryland, USA/Washington, District of Columbia; Chicago, Illinois; Los Angeles, California; Pittsburgh, Pennsylvania). The follow-up schedule consisted of semiannual study visits wherein serological, clinical, and behavioral data were collected via a variety of methods including blood samples, physical examinations, self-report in structured interviews, review of disease registries, and confirmatory medical chart review.

2 CHAPTER II: Circulating Markers of Immune Activation and Inflammation, and Non-Hodgkin Lymphoma: A Meta-Analysis of Prospective Studies

2.1 Abstract

Background: Chronic inflammation and immune activation are reported to play a key role in the etiology of non-Hodgkin lymphoma (NHL). We conducted a meta-analysis on the associations between prediagnosis circulating levels of immune stimulatory markers, interleukin 6 (IL-6), IL-10, tumor necrosis factor α (TNF-α), CXCL13, soluble CD23 (sCD23), sCD27, sCD30, and the risk of NHL.

Methods: Relevant studies were identified from PubMed, EMBASE, and Web of Science up to January 1, 2017. We calculated summary odds-ratio (OR) estimates for the association between one natural log increase in concentration of each biomarker and NHL using random-effects models for NHL as a composite outcome, and for several histological subtypes of NHL.

Results: Seventeen nested case–control studies were included. Elevated levels of several biomarkers were more strongly associated with increased odds of NHL: TNF-α, OR=1.18 [95% CI: 1.04, 1.34]; CXCL13, OR=1.47 [95% CI: 1.03, 2.08]; sCD23, OR=1.57 [95% CI: 1.21, 2.05]; sCD27, OR=2.18 [95% CI: 1.20, 3.98]; sCD30, OR=1.65 [95% CI: 1.22, 2.22]. In stratified analyses, IL-6, TNF-α, sCD27, and sCD30 were more strongly associated with NHL in HIV-infected individuals compared to HIV-uninfected individuals. Between-study heterogeneity was observed across multiple biomarkers for overall NHL, and by subtypes.

Conclusion: This meta-analysis provides evidence that elevated circulating levels of TNF-α, CXCL13, sCD23, sCD27, and sCD30 are consistently associated with an increased risk of NHL, suggesting the potential utility of these biomarkers in population risk stratification and prediction.

2.2 Introduction

Profound immune dysregulation, particularly in the setting of HIV infection or solid organ transplantation, is among the strongest risk factors for non-Hodgkin lymphoma (NHL).[41] Among HIV-infected individuals, two pathogenic mechanisms have been hypothesized to contribute to AIDS-NHL.[3,4,42] The first is the dysregulated proliferation of Epstein-Barr virus (EBV)-transformed B cells, resulting from impairment of T-cell-mediated immunity.[3] The other is chronic B cell activation and resultant downstream processes that promote oncogenic mutations and translocations.[4] In the setting of solid organ transplantation, a large fraction of NHL is attributed to EBV; however, NHL occurrence in long-term transplant survivors appears to be caused by factors other than EBV.[43–45]

Less severe immune dysregulation, in the form of autoimmune conditions and subclinical immune deficiency, has been associated with increased NHL risk.[41] Importantly, observational studies assessing associations between NHL and serologic measurements of immune markers, such as cytokines, chemokines, and soluble receptors, have provided evidence implicating alteration in these biomarkers and lymphomagenesis.[25,46–48]

Two narrative reviews have been published that descriptively summarize much of the relevant literature regarding biomarkers for NHL development,[4,49] but neither quantified the associations of immunological markers and NHL. A recent meta-analysis of associations between NHL and both sCD27 and sCD30 has been published.[50] In this study, we aim to synthesize evidence that has accumulated in the literature[4,49,50] to quantify associations of prediagnosis biomarkers of inflammation and immune activation with subsequent NHL for a select set of biomarkers. We selected immune biomarkers included in prior reviews,[4,49,50] which we hypothesize are biologically relevant to NHL etiology (interleukin [IL] 6, IL-10, CXCL13,

soluble [s]CD23, sCD27, sCD30, tumor necrosis factor [TNF] -α). Our synthesis of results through meta-analysis may contribute towards developing biomarkers for risk prediction in high risk populations.

2.3 Materials And Methods

We conducted this meta-analysis according to the guidelines stated in the Meta-Analysis of Observational Studies in Epidemiology (MOOSE) statement.[51] We provide a completed MOOSE checklist as supplementary material (Supplementary Table 2.4).

2.3.1 Literature Search Strategy

We executed a literature search in MEDLINE, EMBASE, and Web of Science to comprehensively capture publications with dates starting from inception (1966, 1946, and 1900, respectively) of the databases up to January 1, 2017. We searched the databases to identify observational studies with prospectively collected data on serological immune markers and incident NHL. Our article search strategy used controlled database vocabulary where applicable, key words, and boolean logic to apply the following search terms and logic: "'non-hodgkin lymphoma' AND ('interleukin 6' OR 'interleukin 10' OR 'tumor necrosis factor alpha' OR ‘cxcl13’ OR 'cd23 antigen' OR 'cd27 antigen' OR 'cd30 antigen')". No other restrictions were imposed on the search. We sought additional articles from the reference lists of articles identified through the database search and of recent review articles,[4,49,50] as well as from unpublished studies presented at national meetings with permission from willing investigators. A library information science specialist was consulted regarding database coverage and implementing controlled search vocabulary.

2.3.2 Inclusion and Exclusion Criteria

Studies were included in this meta-analysis if they met the following criteria: (1) studies with prospective collection of plasma or serum for measurement of immunological biomarkers; (2) original articles reporting odds ratios (OR), hazard ratios, rate ratios, or relative-risks as measures of association, or data from which an estimate of the OR could be approximated; (3) reported the association between any subset of prediagnosis serum biomarkers of interest and NHL risk or the risk of subtypes of NHL as outcomes; and (4) reported estimates adjusted or controlled for a minimum of age and sex, but not other biomarkers. For studies of HIV-infected participants, adjustment criteria included receipt of highly active antiretroviral therapy (HAART) and at least one marker of immunological function (e.g. CD4+ cell counts or duration of infection). We excluded case reports, conference abstracts, and review articles.

2.3.3 Data Items and Data Extraction Strategy

The following data were extracted from each publication: the biomarker(s) being assessed, NHL outcome including subtypes, timing of blood draw prior to NHL diagnosis (prediagnosis time lag), HIV serostatus, HAART exposure, adjustment variables, sample size (counts of cases and controls), country where the study was conducted, the first author's name, publication year, and estimates of measures of association with their corresponding 95% confidence intervals (CIs) or standard errors for each comparison evaluated, and the document identification number for the publication. We also extracted the boundaries of predictor categories when biomarkers were analyzed as categorical predictors. Two of the co-authors (RSB and SBM) extracted results and information from the manuscripts of eligible studies onto spreadsheets, but without double entry. Authors (RSB and SBM) verified the accuracy of the

collected data through cross-inspection of entered data. Discordant findings were resolved by discussion and consensus between the authors.

2.4 Data Analysis

2.4.1 Data Harmonization of Published Results

Since all studies reported ORs, we natural log-transformed the ORs and estimated the standard errors of the log-ORs by taking the natural logarithm of the upper and lower bound of the 95% CIs, then dividing the difference by 3.92 (twice the 97.5th percentile of the standard normal distribution).[52] Many publications[53–59] had analyzed their predictor biomarkers on a continuous natural logarithm unit scale, or on a continuous scale that could be rescaled to be commensurate with natural logarithm units. For publications[25,46–48,50,60–64] presenting ORs estimated with categorized predictor biomarkers, we first applied a log-transformation to the category boundaries, and calculated the intracategory mid-points. Using a published SAS macro,[65] we applied a multi-step procedure[66,67] that included fitting an inverse-variance weighted regression on the log-OR over the mid-points of biomarker categories. This allowed us to obtain an estimate of the change in log-odds of NHL for each logarithm-unit change in each biomarker, and its corresponding standard error, had the predictor not been categorized in the published analysis. For publications[47,63] that did not present the category boundaries for biomarkers categorized by percentiles, we first estimated the predictor biomarker percentiles assuming a normally distributed natural log-transformed biomarker with the mean and the standard deviation estimated from available statistics of the distribution using methods previously described.[68,69]

Considering studies that estimated associations within strata defined by prediagnosis time lag, we collapsed the strata by calculating the inverse-variance weighted average of log-

odds ratios over the time-intervals to produce estimates of biomarker-NHL associations for the composite overall NHL outcome averaged over the maximum range of prediagnosis lag time, as well as within broader categories of early prediagnosis time lag (defined as 6-10+ years prior to diagnosis), and late prediagnosis time lag (0-5 years prior to diagnosis, 0 being within the year of diagnosis). We also averaged results for NHL subtype outcomes by groups of subtypes, including diffuse large B cell lymphoma (DLBCL), chronic lymphocytic leukemia/ small lymphocytic lymphoma/ prolymphocytic leukemia (CLL/SLL/PLL), and follicular lymphoma (FL), all aggregated according to Surveillance, Epidemiology, and End Results Program (SEER) International Classification of Diseases for Oncology, third edition (ICD-O-3) morphology codes.[70]

2.4.2 Estimation of Meta-Analytic Summary ORs

Anticipating between-study heterogeneity *a priori*, we fit a restricted maximum likelihood random-effects model[71] to calculate summary ORs across studies for each biomarker. We also stratified the analyses by subgroups of HIV-serostatus and contrasted the OR estimates across serostatus subgroups by estimating a ratio-of-odds-ratios (ROR), and corresponding 95% CIs and p-values. Similarly, we calculated pairwise RORs, and their corresponding 95% CIs with p-values from z-tests to compare the OR estimates between pairs of histological subtypes of NHL. In addition, to the extent possible, we carried out stratified analyses within strata defined by HAART exposure and prediagnosis time lag ranges (0-5 years and 6-10 years prior to NHL diagnosis).

2.4.3 Estimation of Between-Study Heterogeneity

We assessed the presence of statistical heterogeneity between studies by conducting Cochran's *Q* test for statistical heterogeneity. Cochran's *Q* test statistic is computed as the sum, over all studies, of the squared deviation of each log-OR from the overall summary estimate weighted by the variance for the given log-OR.[72] The *Q* test statistic follows a Chi-squared distribution with k-1 degrees of freedom (where k was the number of studies). We chose a statistical significance threshold of a 2-sided p-value < 0.1 to indicate the presence of heterogeneity.[72] We also calculated Higgins' I^2, a measure of statistical heterogeneity, as the proportion of between-study variance relative to overall variance (overall variance being the sum of between-study and within-study variance) across the observed study log-ORs.[73] I^2 ranges from 0% for no heterogeneity to 100%, with I^2 <25% indicating low heterogeneity, 25%≤ I^2 ≤75% indicating moderate heterogeneity, and I^2 >75% signifying high heterogeneity.[74]

2.4.4 Assessment of Publication Bias & Influential Data

We assessed publication bias by visual inspection of funnel plots[75] of the meta-analytic summary estimates of ORs plotted against their respective standard errors for each biomarker included in our study. An asymmetric distribution of the plotted points exceeding the 90% pseudo-CI of the funnel plot indicate potential presence of publication bias. We also ran Egger's regression tests for each funnel plot with p-value <0.1 signaling the presence of potential publication bias.[76] Furthermore, we quantified the potential effect of publication bias on our results using trim-and-fill analyses described by Duval and Tweedie.[77,78] Trim-and-fill analyses first estimate the results of hypothetically unreported studies using the observed set of study results, such that the asymmetric part of the funnel plot is filled. Then, outlying study estimates are excluded ("trimmed") from outside of the funnel plot pseudo-CIs. Finally, meta-analytic summaries are re-estimated including the estimated hypothetically unpublished results to see if they substantially alter final summary estimates.

Lastly, we do not include formal assessments of publication quality in our analyses because, after applying our inclusion criteria, we expect limited variation in the quality of prospective studies retrieved, and such assessments of quality have been shown to have limited utility in mitigating bias in estimation of associations.[79]

We constructed the final analytic data sets in SAS® version 9.4, Cary, NC. Statistical analyses were implemented in R version 3.2.2[80] with the *meta* and *metafor* packages.[81,82]

2.5 Results

2.5.1 Study Selection

The flow diagram of our literature search is shown in Figure 2.1, with details of the included set of 17 English language papers (no foreign language papers captured by search) provided in Table 2.1. We further excluded 1 study[58] from the analyses of IL-6 and IL-10, but retained it for other analyses, because the cases and controls completely overlapped with those of another study.[54] Other included studies nested within the same parent cohorts had at most only partial, but not complete overlap of study subjects, and therefore were included here without modification. For IL-10 analyses, we further excluded another study[54] because it categorized biomarker levels as detectable versus undetectable. Our included studies comprised a total of 8,684 participants (4,047 cases, ignoring sample overlap, of which 11% were HIV-infected, and 4,637 controls, of which 13% were HIV-infected), and considered biomarkers sampled over a long range of time intervals, from within the year of diagnosis to up to 23 years prior to NHL diagnoses (Table 2.1).

2.5.2 Meta-Analyses

Meta-Analyses

Interleukin-6: Ten studies assessed associations between IL-6 levels and NHL. Each natural log-unit increase in circulating IL-6 was associated, though not statistically significantly, with a 22% increase in odds of NHL, OR=1.22 [95% CI: 0.97, 1.54] (Table 2.2, Figure 2.2). In serostatus subgroup analyses, the summary OR estimate was higher among HIV-infected subjects, OR=2.07 [95% CI: 1.19, 3.60], compared to HIV-uninfected subjects, OR=1.01 [95% CI: 0.97, 1.06], with evidence of a between the two estimates, P<0.001 (Table 2.2, Figure 2.2). When considering NHL subtypes (Table 2.3, Supplementary Figure 2.4), we find that levels of circulating IL-6 had a modest association with DLBCL and pairwise comparisons of follicular lymphoma versus DLBCL showed a modest difference (Table 2.3).

Interleukin-10: A total of 8 nested case-control studies assessed associations between circulating IL-10 levels and NHL. Our summary estimate, OR=1.24 [95% CI: 0.93, 1.63], suggests that each natural log-unit increase in circulating IL-10 is associated with a non-statistically significant increase of 24% in the odds of NHL (Table 2.2, Figure 2.2). Among HIV-infected subjects we found a moderate association with a wide confidence interval, OR=1.20 [95% CI: 0.64, 2.24], as well as among HIV-uninfected subjects, OR=1.25 [95% CI: 0.91, 1.72], with no meaningful difference between the two estimates, p=0.943 (Table 2.2, Figure 2.2). DLBCL and follicular lymphoma showed statistically significant, but modest associations with elevated IL-10 levels, and we observed no substantial differences in estimates when conducting pairwise comparisons by subtype (Table 2.3, Supplementary Figure 2.4).

Tumor Necrosis Factor-α: A set of 9 studies assessed associations between TNF-α levels and NHL. The overall summary estimate of OR=1.18 [95% CI: 1.04, 1.34] (Table 2.2, Figure 2.2) illustrates that elevated serum levels of TNF-α are associated with increased risk of NHL overall, increasing the odds by 18% per natural log unit. When comparing estimates between HIV-infected, OR=1.79 [95% CI: 1.35, 2.37], and HIV-uninfected, OR=1.12 [95% CI: 1.02, 1.23], we found evidence of a difference in ORs between HIV serostatus groups, p=0.005 (Table 2.2, Figure 2.2). Analyses within NHL subtypes showed evidence of associations between TNF-α and CLL/SLL/PLL only, with no differences found in pairwise comparisons between subtypes (Table 2.3, Supplementary Figure 2.4).

CXCL13: Five studies in total assessed associations between CXCL13 levels and NHL. A summary estimate of OR=1.47 [95% CI: 1.03, 2.08] (Table 2.2, Figure 2.2) shows that each natural log-unit increase in circulating CXCL13 is associated with a 47% increase in odds of NHL. When assessed by serostatus subgroups, the summary OR estimate among HIV-infected subjects was OR=2.56 [95% CI: 1.32, 4.96], compared to OR=1.35 [95% CI: 0.95, 1.92] among HIV-uninfected subjects with no evidence of a difference by serostatus (Table 2.2, Figure 2.2). DLBCL was the only subtype to show an association with NHL with some statistical confidence, and pairwise comparisons by subtype showed no meaningful differences (Table 2.3, Supplementary Figure 2.4).

Soluble CD23, CD27, & CD30: Soluble CD23, CD27, and CD30 had 4, 7, and 9 studies assessing its relationship with NHL, respectively. Overall, the meta-analytic estimates showed increased risk of NHL associated with higher circulating concentrations of sCD23 (OR=1.57 [95% CI: 1.21, 2.05]), sCD27 (OR=2.18 [95% CI: 1.20, 3.98]), and sCD30 (OR=1.65 [95% CI: 1.22, 2.22]) (Table 2.2,Figure 2.3). When we compared HIV-infected versus uninfected

subgroups, we observed differences in biomarker associations between NHL and both sCD27 and sCD30 (Table 2, Figure 3). Elevated levels of sCD23 were associated with DLBCL and follicular lymphoma, while all subtypes showed an association with elevated levels of sCD30 (Table 2.3, Supplementary Figure 2.4). Pairwise comparisons of sCD23 associations with follicular lymphoma versus DLBCL showed evidence of differences; similarly, for sCD30, the comparison of its association with follicular lymphoma versus its association with CLL/SLL/PLL showed evidence of a meaningful difference. No other pairwise subtype differences were notable (Table 2.3).

Prediagnosis Time Lag & HAART Exposure: We conducted analyses stratified by early (6 to 10 years prior to NHL diagnosis) versus late collection of biomarkers (0-5 years prior to NHL diagnosis) (Supplementary Table 2.6). In the early period, elevated levels of IL-10 (OR=1.10 [95% CI: 1.03, 1.17]), TNF-α (OR=1.19 [95% CI: 1.05, 1.34]), and sCD30 (OR=1.34 [95% CI: 1.00, 1.80]) were associated with NHL, while ORs and CIs for other biomarkers indicated some positive but uncertain associations with NHL. In contrast, we observed comparatively higher OR in the late period for IL-6, TNF-α, CXCL13, sCD23, sCD27, and sCD30. Formal comparisons of ORs between the two prediagnosis time strata yielded no important differences. We were able to carry out analyses stratified by HAART exposure only for IL-6, sCD23, sCD27, and sCD30, with only one study[57] providing an estimate for HAART exposed individuals (Supplementary Table 2.7). Summary estimates were generally higher among HAART unexposed individuals (estimates ranging from OR=1.75 [95% CI: 1.30, 2.36] to OR=4.72 [95% CI: 2.81, 7.93]), while the OR estimates for the HAART exposed group were generally lower, except for sCD27 for which the sample size was limited (N=9 HAART exposed cases, N=37 controls) resulting in potential sparse data bias. We also did not observe any evidence of meaningful differences in the OR estimates across HAART exposure strata.

Heterogeneity: We found substantial heterogeneity in overall and subgroup (HIV serostatus, NHL subtypes) analyses. For analyses of the overall composite NHL outcome, all Cochran *Q* tests indicated the presence of heterogeneity (i.e. all two-sided p-values less than 0.1), while Higgins's I^2 values indicated moderate to large magnitudes of heterogeneity ranging from I^2=63% [95% CI: 23%, 82%] to I^2=91% [95% CI: 85%, 95%] (Table 2.2). When we conducted subgroup analyses within HIV-serostatus strata, heterogeneity measures decreased only modestly where calculable, with most *Q* tests indicating the presence of heterogeneity (Figure 2.2, Figure 2.3, Supplementary Figure 2.4), and I^2 proportions ranging from I^2=44% [95% CI: 0%, 76%] to I^2=96% [95% CI: 90%, 99%], within the HIV-uninfected subgroup. Within the HIV-infected group sample sizes were small (at most *n*=3) rendering heterogeneity statistics unreliable. When we assessed associations by NHL histological subtypes, we found statistically detectable heterogeneity in two-thirds of comparisons (Cochran *Q* tests <0.1), but with ranges of I^2 statistics that were reduced compared to those of the composite NHL outcome (Table 2.3). We interpret these statistics with caution since the numbers of studies included in the analyses, particularly by subgroups, were limited relative to recommended sample sizes for these measures.[83]

Publication Bias & Influential Data: We provide a set of funnel plots for each analysis for our composite overall NHL outcome (Supplementary Figure 2.5). Due to small sample sizes, evidence of symmetry in the distribution of meta-analytic summary ORs is inconclusive. Egger's regression tests suggest the presence of potential publication bias for the OR estimates of NHL for IL-6, IL-10, CXCL13, sCD27, and sCD30 (p<0.1), although small samples limit the validity of this test. Trim-and-fill analyses indicated that studies predicted to be excluded from our analyses due to potential publication bias would have attenuated our estimates for all biomarkers, while maintaining the same direction of association (Supplementary Table 2.5).

Influence diagnostics show a few potentially influential studies, one study for IL-6,[54] IL-10,[25] and CXCL13[56] (Supplementary Figure 2.6).

2.6 Discussion

Two patterns become discernible from our analyses: (1) elevated expression of immune stimulatory molecules, including cytokines, chemokines, and soluble receptors, precedes an NHL diagnosis, and (2) the associated increase in risk is generally higher among HIV-infected relative to HIV-uninfected individuals. These two inferences largely corroborate what has previously been reported in prior independent reports. These results also suggest that HIV itself, due to the immune dysregulation resulting from HIV, or the subtypes that primarily emerge in the presence of HIV, are key factors in the association between immune stimulatory molecules and NHL. Further, our study findings support the use of these molecules as biomarkers for an immune environment that promotes NHL.

IL-6 is a pluripotent cytokine that can stimulate B cell proliferation and differentiation, foster cell survival, and promote tumor growth.[84,85] IL-6 has also been linked to pro-inflammatory and Th17 immune responses, which are related to autoimmunity[86,87] and are closely related to risks for NHL.[88] We found that the positive association between IL-6 and NHL was stronger among HIV-infected compared to HIV-uninfected subjects, suggesting a modifying effect of HIV infection. The stronger associations between IL-6 and NHL among HIV-infected subjects could also be influenced by the higher proportion of the DLBCL histological subtype in the presence of HIV,[42,89–91] a subtype that displayed the highest OR in our histological subtype-specific analyses for IL-6, particularly when compared to follicular lymphoma. Although these findings present with a high level of heterogeneity, they are nonetheless qualitatively consistent with the hypothesized etiologic role of IL-6 in the development of NHL.

IL-10 is a pleiotropic cytokine with stimulatory effects on B cells, and is suspected of inducing lymphomagenesis by promoting chronic B cell activation.[92–94] In a mouse model, IL-10 was required for the progression of B cell lymphoma,[95] and in humans, malignant NHL cells produce IL-10.[96,97] A growing body of literature, as described in a recent meta-analysis, showed that IL-10 gene polymorphisms, especially 3575 T/A and 1082 A/G, were associated with increased NHL risk or its subtypes, including DLBCL and follicular lymphoma.[98–101] Our analyses of NHL subtypes corroborate results from studies of genetic polymorphisms since our study also found an association between IL-10 and DLBCL, as well as follicular lymphoma, lending credence to the hypothetical function of IL-10 in lymphomagenesis.

TNF-*α* is a potent pro-inflammatory cytokine that can induce B cell activation, growth, differentiation, apoptosis, and chemotaxis.[102–104] Knock-out mouse models of *TNF,*[105] as well as genetic association studies in humans,[93,106,107] provide evidence of the involvement of TNF-*α* in lymphomagenesis. A potential mechanism through which TNF-α is involved in lymphomagenesis is enhancement of B cell survival, differentiation, and proliferation mediated by the nuclear transcription factor (NF)-κB pathway. [93,103] We found a higher summary OR estimate for NHL among the HIV-infected subgroup compared to the HIV-uninfected group, indicating that elevated levels of TNF-α confer higher risk of NHL in the context of HIV-infection. In addition, we found evidence of associations between elevated levels of TNF-α and DLBCL and follicular lymphoma subtypes. These results are consistent with a hypothesized etiologic function of elevated TNF-*α* levels prior to the onset of NHL.

CXCL13 and its receptor, CXCR5, are required for B cell homing to follicles in lymph nodes,[108] suggesting that aberrant CXCL13 expression may be involved in the pathogenesis of B cell lymphoma through abnormal chemotaxis of B cells to tissues or abnormal B cell

activation.[109] In addition, overexpression of the receptor-ligand pair CXCR5/CXCL13 has been observed in B cell chronic lymphocytic leukemia,[110] and follicular lymphoma cells have been seen to secrete CXCL13.[111] We found an association between NHL and elevated levels of CXCL13, and while our data were insufficient to reliably compare the CXCL13 and NHL associations across serostatus groups, we observed a markedly stronger association among HIV-positive versus HIV-negative individuals. In addition, DLBCL, a subtype more prevalent among HIV-infected populations, showed an association with elevated CXCL13 in our study. These results indicate a possible role for CXCL13 in lymphomagenesis, particularly in the context of HIV infection.

CD23, a cell-surface receptor for the Fc portion of IgE, can be proteolytically cleaved from the B cell surface to produce its soluble form (sCD23).[112] Through the stimulatory action of IL-4, IL-13, and infectious agents,[113] activated B cells upregulate their expression and cleavage of CD23, subsequently increasing concentrations of sCD23 in serum. Serum sCD23 effects further B cell stimulation including increases in IL-4-mediated *IgH* class switch recombination,[112,114] potentially leading to aberrant recombination which is implicated in lymphomagenesis. Additionally, sCD23 may also upregulate monocyte production of IL-6,[115] thereby increasing the likelihood of the development of various NHL subtypes in the context of autoimmune conditions.[88] Contrasting the OR estimate for NHL among the HIV-infected group versus the HIV-uninfected group, we find no substantial differences, suggesting sCD23 may be a biomarker of NHL regardless of the presence or absence of HIV. Elevated levels of sCD23 were associated with DLBCL and follicular lymphoma, with a higher OR estimate for follicular lymphoma relative to DLBCL, p=0.001 (Table 2.3), potentially suggesting a greater etiologic role for follicular lymphoma versus DLBCL.

CD27 and CD30 are members of the TNF-receptor superfamily.[116,117] CD27 is involved in the activation of both T and B cells, stimulating proliferation of T cell proliferation[118] and inducing production of immunoglobulins by B cells.[119] CD30 was first discovered, and is frequently expressed, on Hodgkin lymphoma Reed-Sternberg cells. It is also found expressed on NHL cells, particularly in anaplastic large-cell lymphoma, but is less frequently expressed in cells of other NHL subtypes.[120] CD30 is also expressed by activated T cells, which secrete cytokines that induce B cell activation, differentiation, and proliferation [121,122]. Cell membrane-associated CD27 and CD30 are proteolytically cleaved to produce the soluble forms of these molecules (sCD27 and sCD30) found in serum. Serum concentrations of both sCD27 and sCD30 have been elevated among those with viral infections and autoimmune diseases.[123,124] The potential role of sCD27 in B cell immunoglobulin production, and that of sCD30 in B cell activation, implicates these molecules in lymphomagenesis. Similarly, our study found elevated levels of both sCD27 and sCD30 to be associated with NHL overall. Broken down by HIV serostatus groups, we found larger magnitudes of ORs among HIV-positive individuals relative to those who were HIV-negative, and while the estimates were imprecise due to limited sample sizes, this result aligns with prior findings that heightened concentrations of these biomarkers precedes NHL, particularly during HIV infection.[125,126] In our analyses by subtype, we found that sCD27 was associated with DLBCL and follicular lymphoma, and sCD30 showed an association with all NHL subtypes. Evidence of differences in OR estimates for follicular lymphoma versus DLBCL, and follicular lymphoma versus CLL/SLL/PLL, for sCD23 and sCD30, respectively, suggest that higher concentrations of these biomarkers may play a greater role in the development of follicular lymphoma relative to the other markers.

Temporal variations in the association between serum biomarkers and NHL may be due to etiologic factors or prodromal effects acting at different time intervals.[54,57,126,127] We included exploratory analyses stratified by the early versus late collection of biomarkers. In the early

period, we observed evidence of associations with NHL among several biomarkers (IL-10, TNF-α, sCD30), and notably stronger associations of several biomarkers (IL-6, TNF-α, CXCL13, sCD23, sCD27, sCD30) measured nearer in time to NHL diagnosis, although there were no meaningful differences between the two time intervals (Supplementary Table 2.6). hese findings are consistent with the inference that these biomarkers are elevated several years prior to NHL, and also that further increases in concentrations of these biomarkers may occur in the tumor microenvironment as clinical detectability of malignancy approaches.

Among the studies based in HIV-infected populations, the vast majority of cases and matched controls were HAART-naïve. Recently, serum levels of several immune markers, including IL-6, were shown to be elevated in HAART naïve individuals compared to those who were HIV-negative, but normalized following HAART therapy.[128] With the advent of HAART, the etiologic effect of HIV on NHL risk appears to have been attenuated, but not eliminated.[4] In supplementary analyses, we assessed biomarker-NHL associations stratified by HAART exposure, and observed increased odds of NHL associated with higher elevations of biomarkers among the HAART unexposed relative to exposed groups for most biomarkers included in these analyses: IL-6, sCD23, sCD30 (Supplementary Table 2.7). We note that these analyses are exploratory in nature due to limited sample sizes within each stratum (N=1 for all HAART exposed; maximum N=3 for HAART unexposed).

Major strengths of our study include the comprehensive coverage of literature and biomarkers with quantitative syntheses of results, and the inclusion of studies with prospective collection of immune markers. Prior reports either included a limited set of biomarkers,[50] or were descriptive in nature, thereby lacking quantitative summaries of published estimates.[4,49] An additional strength of our study is that we included only studies that utilized a prospective-specimen collection, retrospective-blinded-evaluation (PRoBE) design with highly comparable

control groups, thereby increasing our confidence in the validity of the reports. Furthermore, the use of multiplex assays in many of the included studies allowed several biomarkers to be analyzed and reported simultaneously, without regard to statistical significance, minimizing the 'file-drawer' problem of studies hidden from publication due to results that were not statistically significant.

A weakness in our analyses is the modest number of studies for some biomarkers, which produced several limitations. First, sparse study counts limited our ability to adequately explore modifying factors across studies including prediagnosis time interval of biomarker collection, age, sex, and HAART exposure as potential modifiers of biomarker-NHL associations. We provide some exploratory analyses of associations by early versus late collection of biomarkers prior to NHL, and stratified analyses by HAART exposure, but we note the substantial limitations of these analyses. For example, in the lag-time stratified analyses, there were overlapping time intervals over which biomarkers were collected such that the definitions of "early" versus "late" collection were not strictly mutually exclusive. Secondly, estimates of heterogeneity statistics, I^2 and Q, have been documented to be biased in small samples,[83] and outliers tend to have higher influence in small samples. In addition, we did not find convincing evidence of potential publication bias partly due to the limited sample sizes that render funnel plots and Egger's regression p-values unreliable,[129] but also because simultaneous analyses of biomarkers from multiplex assays reduce the chance of non-significant associations going unpublished.

Another limitation of our study is the intrinsic variability in the biomarker quantitation among the studies in our analyses. We included studies that use various assay technologies, with biomarkers quantitated in different laboratories following different protocols and standards. Breen et al. found considerable variability between multiple laboratory sites using high-

sensitivity multiplex cytokine assays in their quantitation of 13 cytokines, both across study sites and multiplex assay technology, despite standardization of samples and laboratory protocols.[130] Noble et al. found significant variation in the quantitation of a standard cytokine provided to 11 laboratories, with the mean concentrations ranging between 67% and 136% of the grand mean.[131] An additional contributing factor to heterogeneity in results is that we were unable to differentiate between germinal cell versus non-germinal cell lymphomas. Since these subtypes differ in etiologic mechanisms and in their interactions with the immune system,[132–134] we expect these issues to contribute to the observed heterogeneity between studies, even within our subtype analyses since we were unable to further stratify by germinal cell origin.

Lastly, we acknowledge that our study is susceptible to bias due to multiple statistical testing of summary estimates, and that multiple comparison adjustments to p-values and confidence intervals widen our estimated confidence interval widths,[135,136] and attenuate the magnitudes of the p-values. However, these adjustments do not invalidate the overall qualitative message that, in general, levels of circulating markers are elevated prior to NHL diagnosis (Supplementary Table 2.8).

In conclusion, our summaries concur with the general trends in published estimates, and provide a systematic description of the variation in estimates of associations between NHL and expression of immune stimulatory molecules. Future research may further strengthen the inferences possible from a review such as ours by including larger sets of publications as the literature grows, particularly among HIV-infected populations, and pooled individual level data studies could allow for more robust control of confounding. Our findings provide support for the hypothesis that chronic immune activation is a crucial mechanism in lymphomagenesis, hence its biomarkers could, in the future, have utility in developing models for early detection.

2.7 Acknowledgements

This research was supported by the National Institutes of Health grant T32 CA09142, the Center for HIV Identification, Prevention, and Treatment (CHIPTS) NIMH grant P30MH058107; and the UCLA Center for AIDS Research (CFAR) grant 5P30AI028697, Core H. The ideas and opinions expressed herein are those of the authors, are not in any manner endorsed by the National Institutes of Health, the University of California at Los Angeles or their Contractors and Subcontractors. UCLA library Research Informationist, Bethany A. Myers, provided excellent consultation on database coverage and the search strategy.

2.8 Supporting Information

Additional Supporting Information may be found in the online version of this article including:

1. **Supplementary Table 2.4**: MOOSE Checklist for Meta-analyses of Observational Studies.
2. **Supplementary Figure 2.4**: Forest Plots for NHL Subtype Outcomes.
3. **Supplementary Figure 2.5**: Funnel Plots Assessing Publication Bias
4. **Supplementary Table 2.5**: Bias Analyses: Egger's Regression P-values and Trim & Fill Analyses.
5. **Supplementary Figure 2.6**: Influence Diagnostics: Leave-One-Out Analyses.
6. **Supplementary Table 2.6:** Results for All B-Cell NHL by Prediagnosis Time Interval: Comparing Early versus Late Biomarker Sample Collection
7. **Supplementary Table 2.7:** Results for All B-Cell NHL among HIV-infected: Comparing HAART exposed versus Unexposed
8. **Supplementary Table 2.8:** Multiple Comparisons Adjusted P-Values for Overall NHL

2.9 Disclosures

The authors of the manuscript have no conflicts of interest to disclose

2.10 Tables and Figures

Figure 2.1: Flowchart for Systematic Literature Search and Selection of Studies of Circulating Biomarkers and NHL Risk.

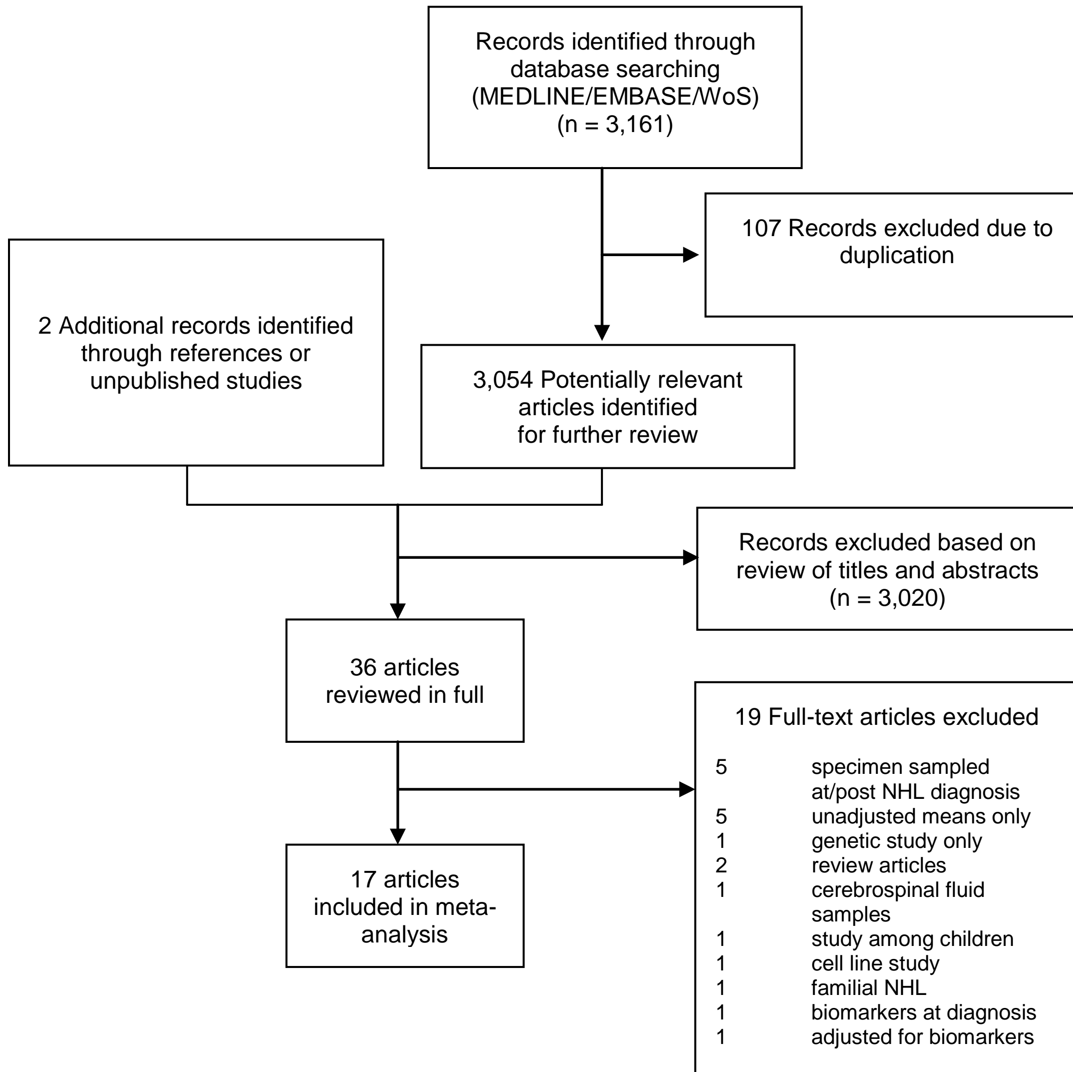


Table 2.1: Characteristics of 17 Prospective Studies Included in the Meta-Analysis.

Source	Year [a]	Location, Cohort, Enrollment Years [b]	Sex	Age [c]	Biomarker(s)	Relevant NHL Subtypes	Cases	Controls	HIV Sero-status	Pre-NHL Time Interval [d]	Covariates [e]
Purdue [60]	2009	United States, PLCO, 1993–2001	M/F	55-74	sCD30	B-NHL, CLL/SLL, DLBCL, FL	234	234	HIV-	1-10	Matched: Age (Baseline), Sex, Race, Blood Draw Date (Baseline), Center;
Gu [61]	2010	United States, NYHS, 1985–1991	F	35–65	IL-10, IL-6, TNF-α	B-NHL	92	184	HIV-	0-15+	Matched: Age, Race, Blood Draw Date; BMI, Alcohol Intake, Smoking.
Hosnijeh [48]	2010	Italy, EPIC Italy, 1993–1998	M/F	35-65	IL-10, IL-6, TNF-α	B-NHL	86	86	HIV-	0-10	Matched: Age (Diagnosis), Age (Baseline), Recruitment (Baseline) Date, Sex, Center; BMI, Alcohol Intake;
Breen [54]	2011	United States, MACS, 1984–1985/1987–1991	M	24-60	IL-6, sCD23, sCD27, sCD30	B-NHL, DLBCL	179	179	HIV+	0-5	Matched: Duration Of HIV Infection/Duration Since Study Entry Date, Expected Sample Availability; Age, CD4+ T-cells/mm^3.
Purdue [62]	2011	United States, PLCO, 1993–2001	M/F	55-74	IL-10, IL-6, TNF-α, sCD27	B-NHL, CLL/SLL, DLBCL, FL	297	297	HIV-	1-10	Matched: Age (Baseline), Sex, Race, Blood Draw Date (Baseline), Center;
Rabkin [55]	2011	United States, NCI, 1985–2004	M/F	29-44	IL-10, IL-6, TNF-α	B-NHL	63	181	HIV+	0.1-2	Matched: Age (Diagnosis), Race, Sex, Blood Draw Date (Period), CD4+ T-cells/mm^3 (Diagnosis), Cohort, Sample Type;
Vermeulen [63]	2011	Italy, EPIC, 1993–1998	M/F	35-70	sCD30	B-NHL	35	36	HIV-	2-6+	Age (Baseline), Sex; BMI

Source	Year [a]	Location, Cohort, Enrollment Years [b]	Sex	Age [c]	Biomarker(s)	Relevant NHL Subtypes	Cases	Controls	HIV Sero-status	Pre-NHL Time Interval [d]	Covariates [e]
De Roos [56]	2012	United States, WHI OS, 1994–1998	F	50-79	CXCL13, sCD23, sCD27, sCD30	B-NHL, CLL/SLL/PLL, DLBCL, FL	491	491	HIV-	0-13	Matched: Age (Birth Year), Blood Draw Date (Baseline), Region;
Conroy [25]	2013	United States, MEC Biospecimen Subcohort, 2001–2006	M/F	45-75	IL-10, IL-6, TNF-α	B-NHL, DLBCL, FL	272	541	HIV-	0-11.5	Matched: Age, Sex, Race, Region (State), Blood Draw Date & Time, Fasting Hours (Pre-Blood Draw);
Hussain [57]	2013	United States, WIHS, 1994–1995/2001–2002	F	<30 - ≥50	CXCL13, IL-6, sCD23, sCD27, sCD30	B-NHL	22	78	HIV+	0.1-4.7	Matched: Age, Race, CD4+ T-cells/mm^3, Duration Since Seroconversion; HIV Viral Load, HAART, Smoking, HCV, Education.
Purdue [47]	2013	United States, PLCO, 1993–2001	M/F	55-74	CXCL13, IL-10, IL-6, TNF-α	B-NHL, CLL/SLL, DLBCL, FL	301	301	HIV-	5-13	Matched: Age (Baseline), Sex, Race, Center, Blood Draw Date & Time;
Edlefsen [59]	2014	United States, WHI OS, 1994–1998	F	50-79	IL-10, IL-6, TNF-α	B-NHL, CLL/SLL/PLL, DLBCL, FL	491	491	HIV-	<3-13	Matched: Age, Blood Draw Date, Region;
Vendrame [58]	2014	United States, MACS, 1984–1985/1987–1991	M	24-70	IL-10, IL-6, TNF-α	B-NHL	176	176	HIV+	0-5	Matched: Duration Of HIV Infection/Duration Since Study Entry Date, expected sample availability; Age, CD4+ T-cells/mm^3.

Source	Year [a]	Location, Cohort, Enrollment Years [b]	Sex	Age [c]	Biomarker(s)	Relevant NHL Subtypes	Cases	Controls	HIV Sero-status	Pre-NHL Time Interval [d]	Covariates [e]
Bassig [46]	2015	Shanghai, SWHS, 1996–2000; Shanghai, SCS, 1986–1989; Singapore, SCHS, 1993–1998	M/F	40-74	sCD27, sCD30	B-NHL	218	218	HIV-	0-10+	Matched: SCS: Age (Baseline), Sex, Blood Draw Date, Region (Neighborhood); SCHS: Age (Baseline), Sex, Baseline Date, Biospecimen Collection Date, Dialect; SWHS: Age (Baseline), Blood Draw Date; Age, Smoking
Purdue [64]	2015	Finland, ATBC, 1985–1988	M	50-69	sCD23, sCD27, sCD30	B-NHL, CLL/SLL, DLBCL	272	325	HIV-	2-23	Matched: Age (Baseline), Blood Draw Date, Number of Prior Specimen Thaws; Smoking.
Hosnijeh [50]	2016	Italy, EPIC Italy, 1993–1998; Sweden, NSHDS/VIP, 1985–2008	M/F	35-70	sCD27, sCD30	B-NHL, CLL/SLL, DLBCL, FL	218	218	HIV-	1-17	Matched: Age (Baseline), Sex, Blood Draw Date, Cohort, Center;
Epstein [53]	2017	United States, NHS 1989–1990; HPFS, 1993–1994	M/F	30-75	CXCL13, IL-10, IL-6, TNF-α, sCD30	B-NHL, CLL/SLL, DLBCL, FL	600	601	HIV-	0-10+	Age (Blood Draw), Race, Blood Draw Time of Day, Cohort

Abbreviations (alphabetical): Alpha-Tocopherol, Beta Carotene Cancer Prevention (ATBC); B-NHL, B cell NHL; BMI, Body Mass Index (in kg/m^2); CLL/SLL/PLL, chronic lymphocytic leukemia/small lymphocytic lymphoma/ prolymphocytic leukemia; DLBCL, diffuse large B cell lymphoma; EPIC Italy, Italian European Prospective Investigation into Cancer and Nutrition cohort; FL, Follicular Lymphoma; HAART, highly active antiretroviral therapy; HPFS, Health Professionals Follow-up Study; MACS, Multicenter AIDS Cohort Study; NCI, U.S. National Cancer Institute; NHL, Non-Hodgkin lymphoma; NHS, Nurses' Health Study; NSAID, Nonsteroidal anti-inflammatory drug; NSHDS, Northern Sweden Health and Disease Study; SCHS, Singapore Chinese Health Study, SCS, Shanghai Cohort Study, SWHS, Shanghai Women's Health Study; VIP, Västerbotten Intervention program; WHI OS, Women's Health Initiative Observational Study component.

a. Year original article was published.

b. Country or city, nesting cohort study name, and enrollment period of nesting cohort study. Years reported for Rabkin [55], were years of NHL diagnosis in combined NCI cohort data.

c. Age at enrollment into the nesting cohort study. Where enrollment age not reported, age range from article descriptive statistics provided.

Source	Year [a]	Location, Cohort, Enrollment Years [b]	Sex	Age [c]	Biomarker(s)	Relevant NHL Subtypes	Cases	Controls	HIV Sero-status	Pre-NHL Time Interval [d]	Covariates [e]

d. Pre-NHL time interval refers to the range of time intervals, in years, prior to NHL diagnosis wherein venipuncture and blood sample collection was conducted. Lower bound of 0 means within the year of, but prior to NHL diagnosis.

e. Matching factors listed defining matching sets used in conditional regression; additional covariates included in models listed after semicolon. Otherwise, covariates for unconditional logistic regression listed for some studies [53,63]. Covariates listed are for the analyses of the composite NHL outcome. Analyses for subtype outcomes may have used different models (e.g. polytomous logistic regression), and adjusted for additional sets of covariates.

f. Subjects for Rabkin [55] comprised a combination of three HIV-infected cohorts followed at the U.S. National Cancer Institute (NCI).

Table 2.2: Meta-Analysis Results for B cell NHL Overall and by HIV serostatus.

	All Subgroups					HIV Serostatus Subgroups					
				Q Test [b]		HIV+		HIV-		Meta-Regression ROR [c] for	
Biomarker	**N**	**OR 95% CI**	I^2 [a]	***Q***	**P**	**OR 95% CI**	I^2	**OR 95% CI**	I^2	**HIV+ / HIV-**	**P**[c]
IL-6	10	1.22 [0.97, 1.54]	80 [65, 89]	45.61	<0.001	2.07 [1.19, 3.60]	82 [44, 94]	1.01 [0.97, 1.06]	0 [0, 69]	1.96 [1.53, 2.50]	<0.001
IL-10	8	1.24 [0.93, 1.63]	82 [65, 90]	38.43	<0.001	1.20 [0.64, 2.24]	--	1.25 [0.91, 1.72]	84 [69, 92]	0.96 [0.33, 2.83]	0.943
TNF-α	9	1.18 [1.04, 1.34]	63 [23, 82]	21.46	0.035	1.79 [1.35, 2.37]	0 [-,-]	1.12 [1.02, 1.23]	44 [0, 76]	1.58 [1.15, 2.18]	0.005
CXCL13	5	1.47 [1.03, 2.08]	89 [78, 95]	37.00	<0.001	2.56 [1.32, 4.96]	--	1.35 [0.95, 1.92]	91 [79, 96]	1.89 [0.69, 5.23]	0.218
CD23	4	1.57 [1.21, 2.05]	90 [77, 96]	29.24	<0.001	1.59 [1.23, 2.06]	0 [-,-]	1.58 [0.93, 2.66]	0 [0, 69]	1.00 [0.54, 1.87]	0.996
CD27	7	2.18 [1.20, 3.98]	92 [87, 96]	79.67	<0.001	4.93 [3.00, 8.08]	0 [-,-]	1.61 [0.89, 2.93]	84 [69, 92]	3.35 [1.05, 10.71]	0.041
CD30	9	1.65 [1.22, 2.22]	90 [84, 94]	83.01	<0.001	3.69 [2.40, 5.69]	11 [-,-]	1.40 [1.11, 1.76]	44 [0, 76]	2.55 [1.38, 4.73]	0.003

a. Higgins' I^2 statistic measuring the proportion of the observed variance between studies relative to the total variance of a set of studies.
b. Q Test assessing the degree to which study effect sizes are concordant.
c. The ratio of odd-ratios compares the odds-ratio for the HIV+ subgroup with that of the HIV- subgroup (HIV+/HIV-). The corresponding p-values test the null hypothesis of no difference between the serostatus groups.
d. Double dash "--" and "-" denote Higgins' I^2 statistics, and confidence intervals, that were not calculated due to inadequate sample size, *n*=1, and *n*=2, respectively.

Table 2.3: Meta-Analysis Results for B Cell NHL Subtypes.

							Comparison of Summary ORs [c]			
				Q-Test [b]		Summary OR	DLBCL		Follicular Lymphoma	
Biomarker	Outcome	N	I^2 [a]	*Q*	P	OR [95% CI]	ROR [95% CI]	P	ROR [95% CI]	P
IL-6	CLL/SLL/PLL	4	0 [0, 0]	0.19	0.996	0.98 [0.92, 1.06]	1.15 [0.99,1.34]	0.074	0.97 [0.87,1.09]	0.652
	DLBCL	6	0 [0, 74]	4.80	0.570	1.13 [0.99, 1.30]	1.00		0.85 [0.72,1.00]	0.044
	Follicular lymphoma	5	9 [0, 81]	4.41	0.492	0.96 [0.88, 1.05]			1.00	
IL-10	CLL/SLL/PLL	4	78 [41, 92]	13.76	0.008	1.09 [0.88, 1.34]	1.04 [0.83,1.29]	0.747	1.01 [0.81,1.26]	0.955
	DLBCL	5	45 [0, 80]	7.28	0.201	1.13 [1.06, 1.21]	1.00		0.97 [0.87,1.07]	0.485
	Follicular lymphoma	5	66 [13, 87]	11.93	0.036	1.09 [1.02, 1.18]			1.00	
TNF-α	CLL/SLL/PLL	4	0 [0, 66]	1.34	0.854	1.15 [1.04, 1.27]	0.91 [0.73,1.14]	0.410	1.21 [0.89,1.65]	0.214
	DLBCL	5	62 [0, 86]	10.41	0.064	1.04 [0.85, 1.28]	1.00		1.34 [0.94,1.90]	0.107
	Follicular lymphoma	5	66 [12, 87]	11.82	0.037	1.39 [1.04, 1.86]			1.00	
CXCL13	CLL/SLL/PLL	4	77 [36, 91]	12.81	0.012	1.43 [0.97, 2.11]	1.18 [0.65,2.12]	0.584	1.20 [0.61,2.37]	0.604
	DLBCL	4	85 [61, 94]	19.43	0.001	1.69 [1.08, 2.62]	1.00		1.02 [0.50,2.08]	0.964
	Follicular lymphoma	3	86 [60, 95]	14.38	0.002	1.71 [0.98, 3.00]			1.00	
sCD23	CLL/SLL/PLL	2	99 [97, 99]	69.59	0.000	2.62 [0.74, 9.19]	0.48 [0.14,1.69]	0.253	0.75 [0.21,2.71]	0.664
	DLBCL	3	49 [0, 85]	3.90	0.272	1.25 [1.11, 1.41]	1.00		1.57 [1.19,2.08]	0.001
	Follicular lymphoma	1	--	0.00	1.000	1.97 [1.53, 2.53]			1.00	
sCD27	CLL/SLL/PLL	3	95 [89, 98]	39.81	<0.001	2.03 [0.73, 5.64]	1.06 [0.29,3.83]	0.927	1.08 [0.22,5.16]	0.927
	DLBCL	4	89 [74, 95]	26.90	<0.001	2.15 [0.99, 4.67]	1.00		1.01 [0.25,4.18]	0.985
	Follicular lymphoma	2	94 [81, 98]	16.56	<0.001	2.18 [0.67, 7.16]			1.00	
sCD30	CLL/SLL/PLL	4	76	12.70	0.013	1.23	1.38	0.205	1.89	0.028

		[35, 91]			[1.05, 1.44]	[0.84,2.26]	[1.07,3.35]	
DLBCL	5	88 [74, 94]	32.94	<0.001	1.69 [1.06, 2.71]	1.00	1.37 [0.67,2.82]	0.387
Follicular lymphoma	3	87 [64, 96]	15.66	0.001	2.33 [1.35, 4.01]		1.00	

a. Higgins's I^2 statistic measuring the proportion of the observed variance between studies relative to the total variance of a set of studies.
b. Q Test assessing the degree to which study effect sizes are concordant.
c. ORs and P-values for comparisons of estimates between outcomes for each biomarker. Each ratio of odds ratios (ROR) compares the odds-ratio for the column biomarker to that of the row biomarker as reference, for example OR_{DLBCL}/ $OR_{CLL/SLL/PLL}$=1.15, with corresponding Wald-type confidence interval using the square root of the sum of the OR variances.
d. Double dash "--" denotes statistics that were not calculated due to inadequate sample size.

Figure 2.2: Forest Plots for Cytokines and Chemokine.

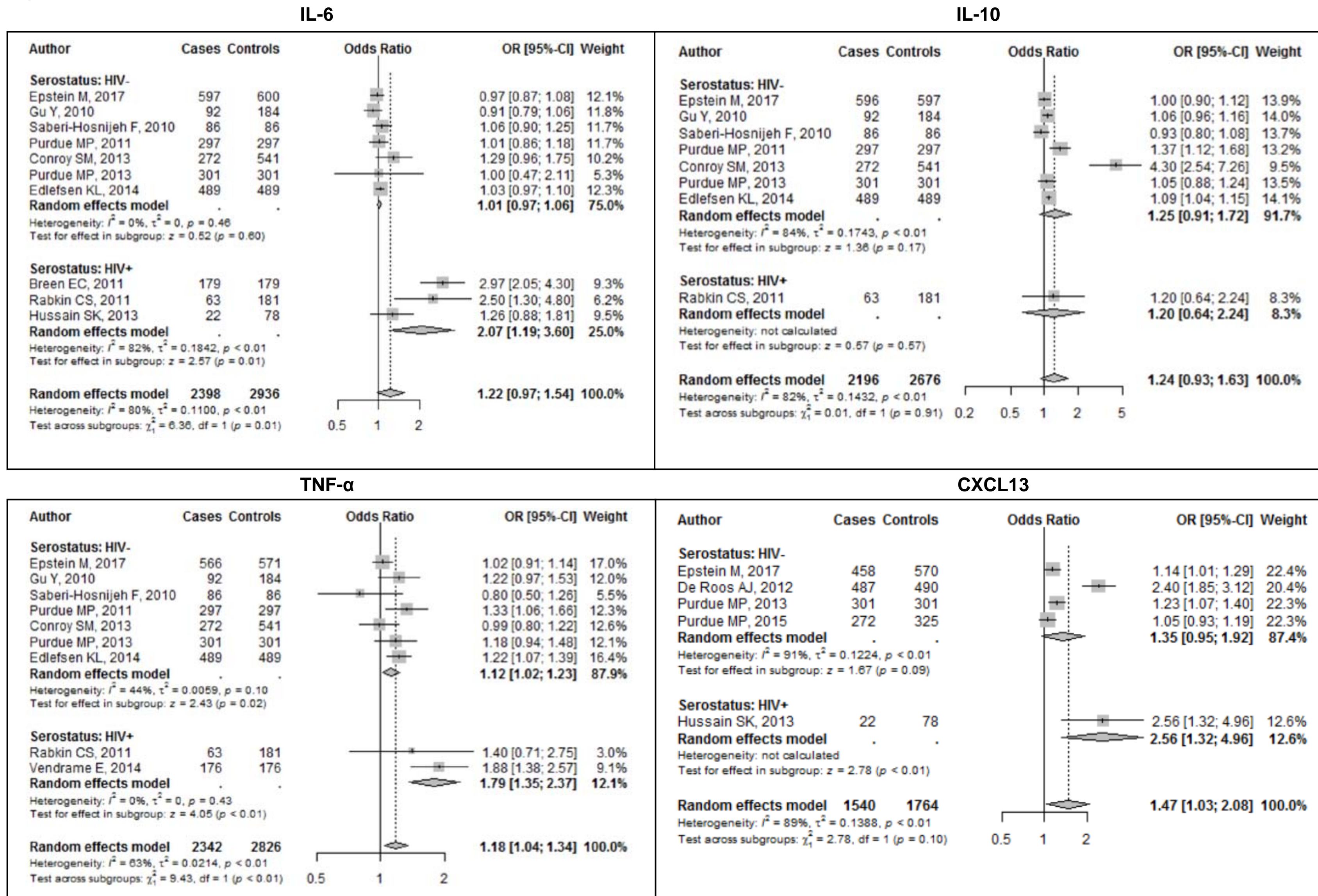


Figure 2.2 Legend:
OR represented with gray squares with error bars indicate 95% CIs; size of the squares indicates the precision weight of each study in the random-effects meta-analysis. Diamonds indicate the summary ORs, with the width denoting the 95% CIs.

Figure 2.3: Forest Plots for Soluble Receptors.

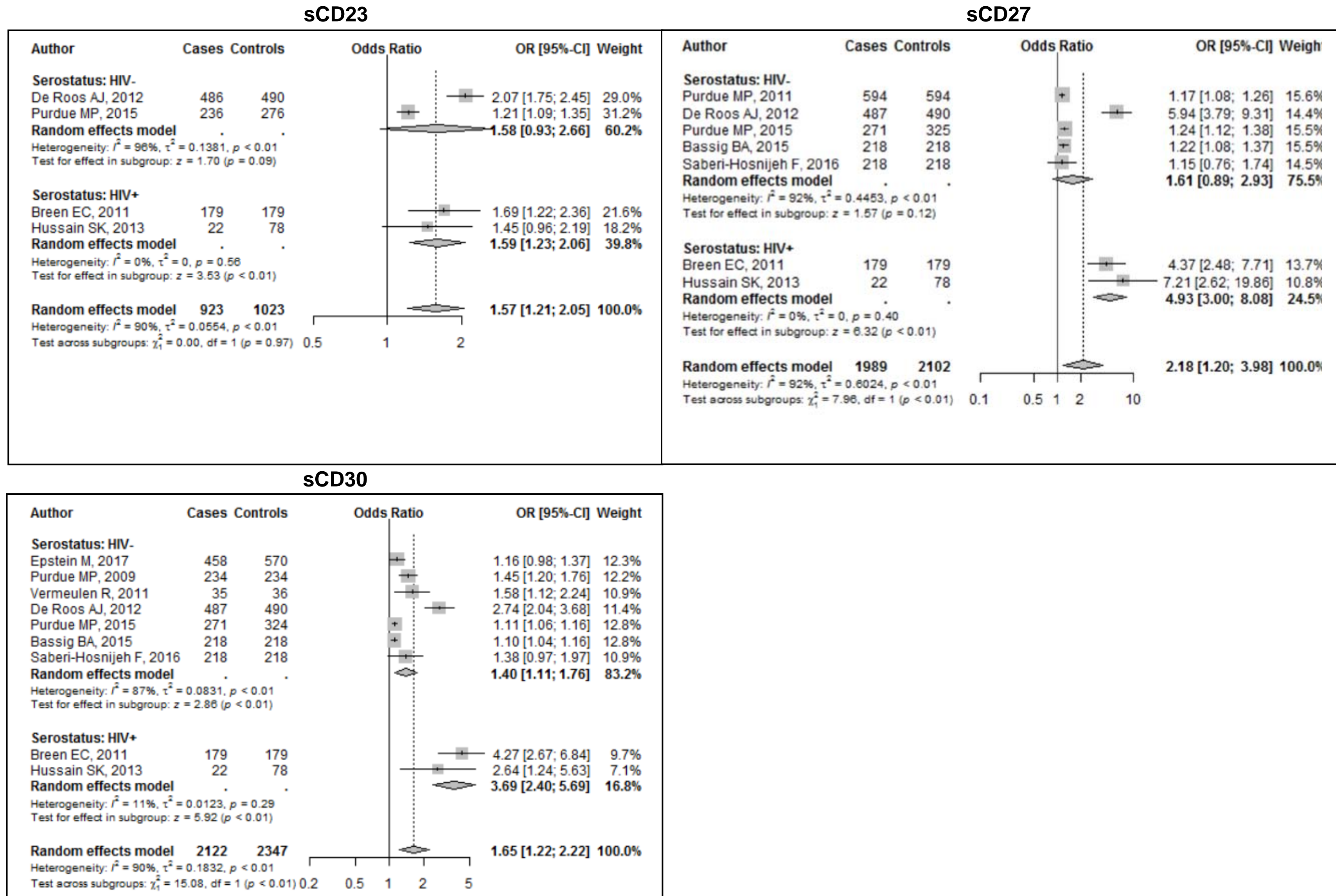


Figure 2.3 Legend:
OR represented with gray squares with error bars indicate 95% CIs; size of the squares indicates the precision weight of each study in the random-effects meta-analysis. Diamonds indicate the summary ORs, with the width denoting the 95% CIs.

2.11 Supplementary Material

Supplementary Table 2.4: MOOSE Checklist for Meta-analyses of Observational Studies.

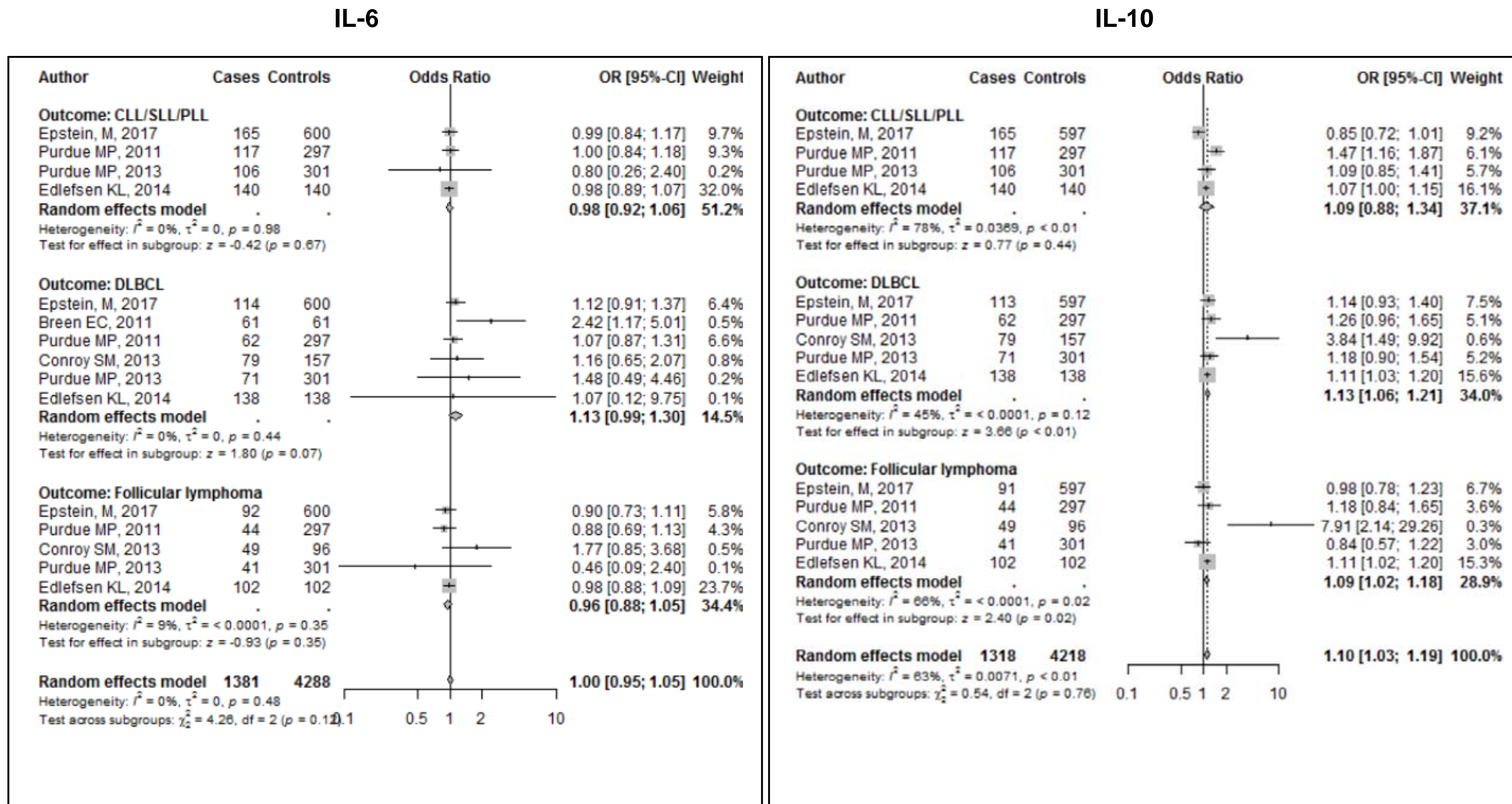
Item No	Recommendation	Reported on Page No
Reporting of background should include		
1	Problem definition	3
2	Hypothesis statement	3
3	Description of study outcome(s)	3
4	Type of exposure or intervention used	3
5	Type of study designs used	4
6	Study population	4-5
Reporting of search strategy should include		
7	Qualifications of searchers (e.g., librarians and investigators)	1,4
8	Search strategy, including time period included in the synthesis and key words	4
9	Effort to include all available studies, including contact with authors	4
10	Databases and registries searched	4
11	Search software used, name and version, including special features used (e.g., explosion)	4
12	Use of hand searching (e.g., reference lists of obtained articles)	4
13	List of citations located and those excluded, including justification	28, Fig 1
14	Method of addressing articles published in languages other than English	--
15	Method of handling abstracts and unpublished studies	4
16	Description of any contact with authors	4
Reporting of methods should include		
17	Description of relevance or appropriateness of studies assembled for assessing the hypothesis to be tested	3-4
18	Rationale for the selection and coding of data (e.g., sound clinical principles or convenience)	5-6
19	Documentation of how data were classified and coded (e.g., multiple raters, blinding and interrater reliability)	5-6
20	Assessment of confounding (e.g., comparability of cases and controls in studies where appropriate)	--
21	Assessment of study quality, including blinding of quality assessors, stratification or regression on possible predictors of study results	--
22	Assessment of heterogeneity	6
23	Description of statistical methods (e.g., complete description of fixed or random effects models, justification of whether the chosen models account for	6-7

	predictors of study results, dose-response models, or cumulative meta-analysis) in sufficient detail to be replicated	
24	Provision of appropriate tables and graphics	28+
Reporting of results should include		
25	Graphic summarizing individual study estimates and overall estimate	Fig 2, 3, Supplement
26	Table giving descriptive information for each study included	28
27	Results of sensitivity testing (e.g., subgroup analysis)	31, 32, Supplement
28	Indication of statistical uncertainty of findings	31, 32, Supplement

From: Stroup DF, Berlin JA, Morton SC, et al, for the Meta-analysis Of Observational Studies in Epidemiology (MOOSE) Group. Meta-analysis of Observational Studies in Epidemiology. A Proposal for Reporting. *JAMA*. 2000;283(15):2008-2012. doi: 10.1001/jama.283.15.2008.

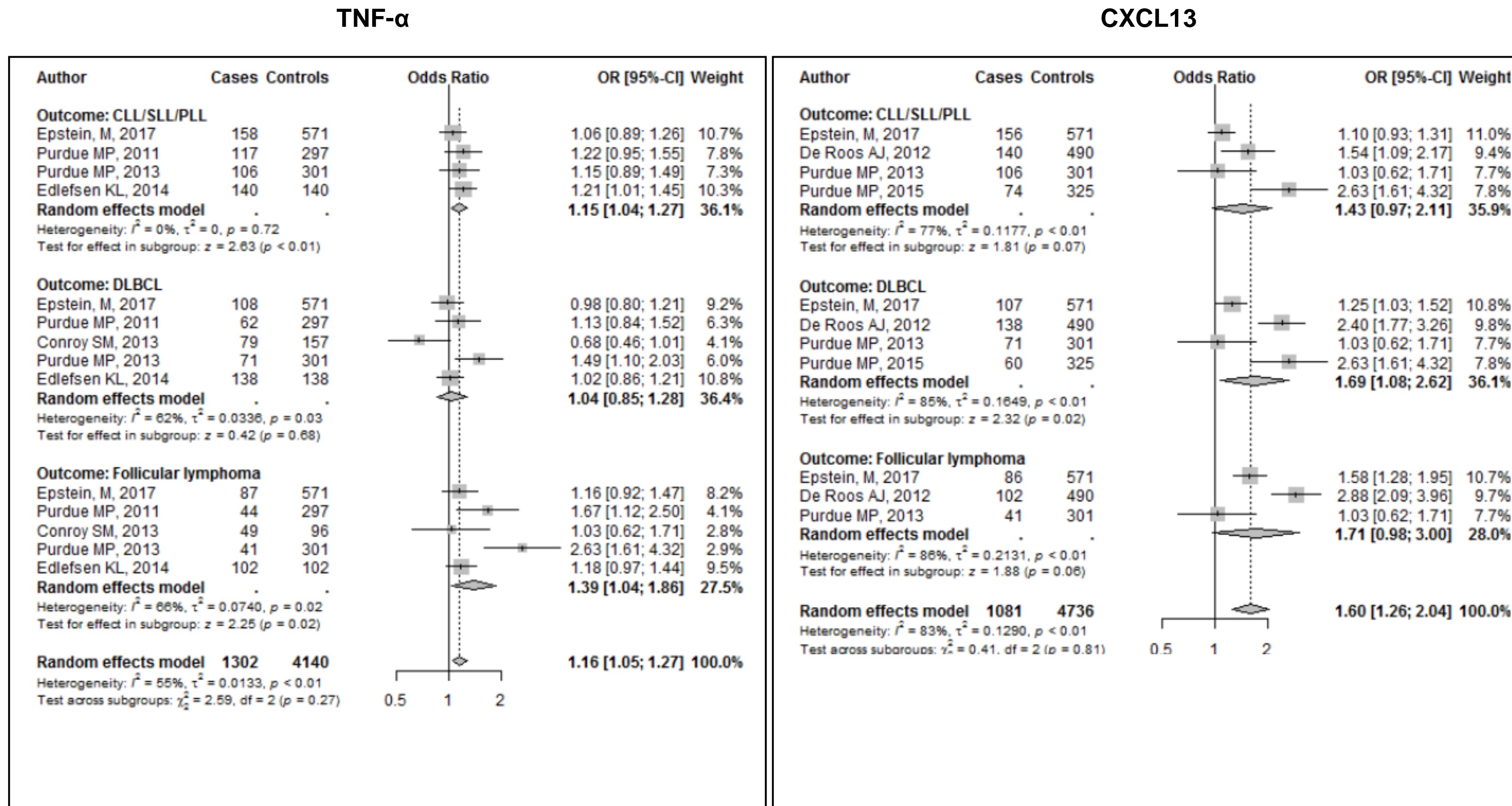
Transcribed from the original paper within the NEUROSURGERY® Editorial Office, Atlanta, GA, United Sates. August 2012.

Supplementary Figure 2.4: Forest Plots for NHL Subtype Outcomes.



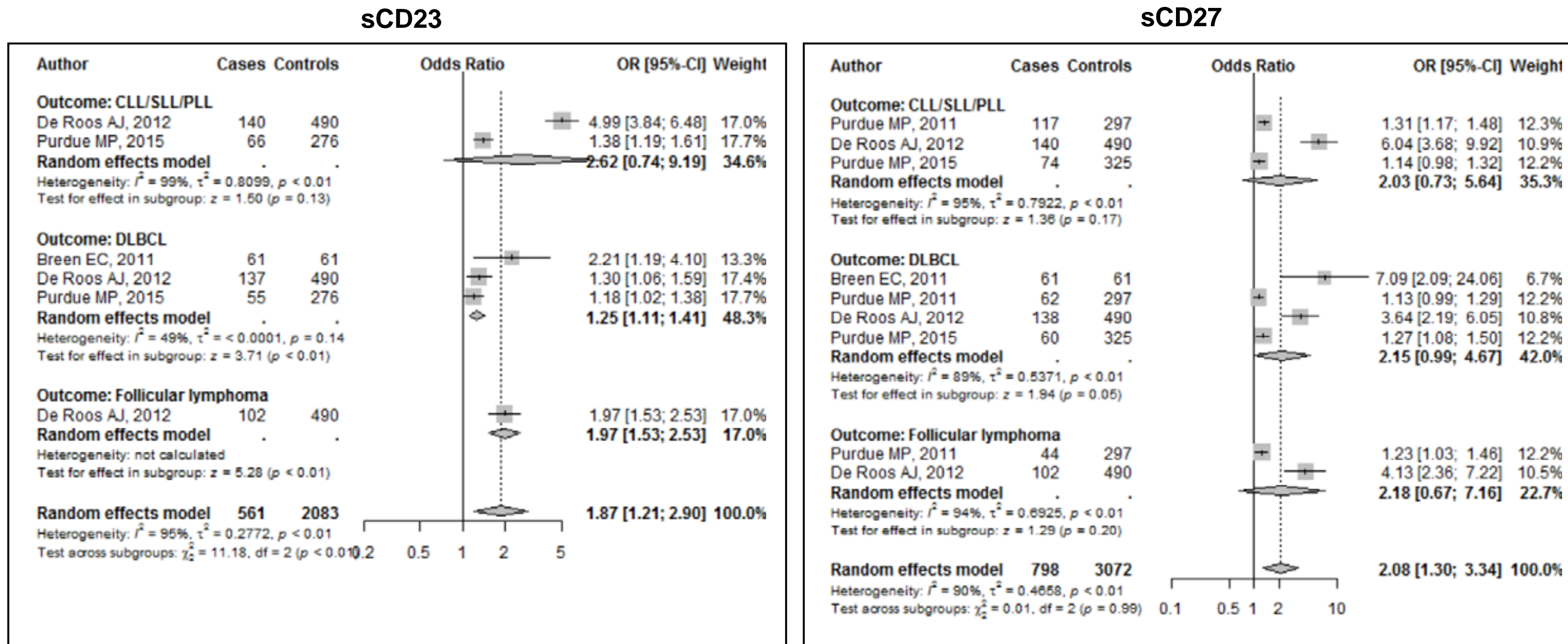
Supplementary Figure 2.4 Legend: Error bars indicate 95% CIs; size of the squares indicates the precision weight of each study in the random-effects meta-analysis. Diamonds indicate the summary ORs. Case and control sample sizes for were estimated for Purdue, 2009[22].

Supplementary Figure 2.4 continued: Forest Plots for NHL Subtype Outcomes.



Supplementary Figure 2.4 Legend: Error bars indicate 95% CIs; size of the squares indicates the precision weight of each study in the random-effects meta-analysis. Diamonds indicate the summary ORs. Case and control sample sizes for were estimated for Purdue, 2009[22].

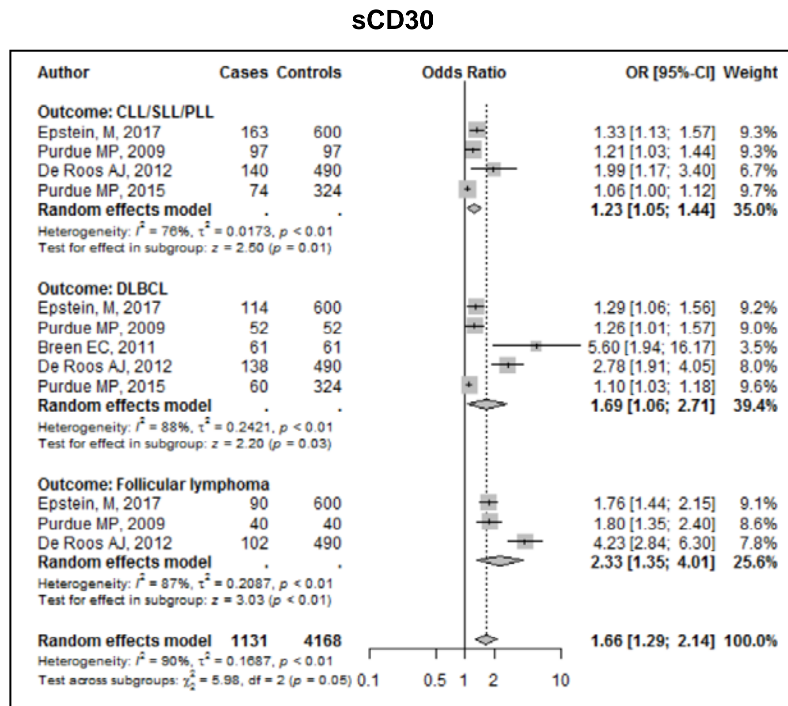
Supplementary Figure 2.4 continued: Forest Plots for NHL Subtype Outcomes.



Supplementary Figure 2.4 Legend: Error bars indicate 95% CIs; size of the squares indicates the precision weight of each study in the random-effects meta-analysis. Diamonds indicate the summary ORs. Case and control sample sizes for were estimated for Purdue, 2009[22].

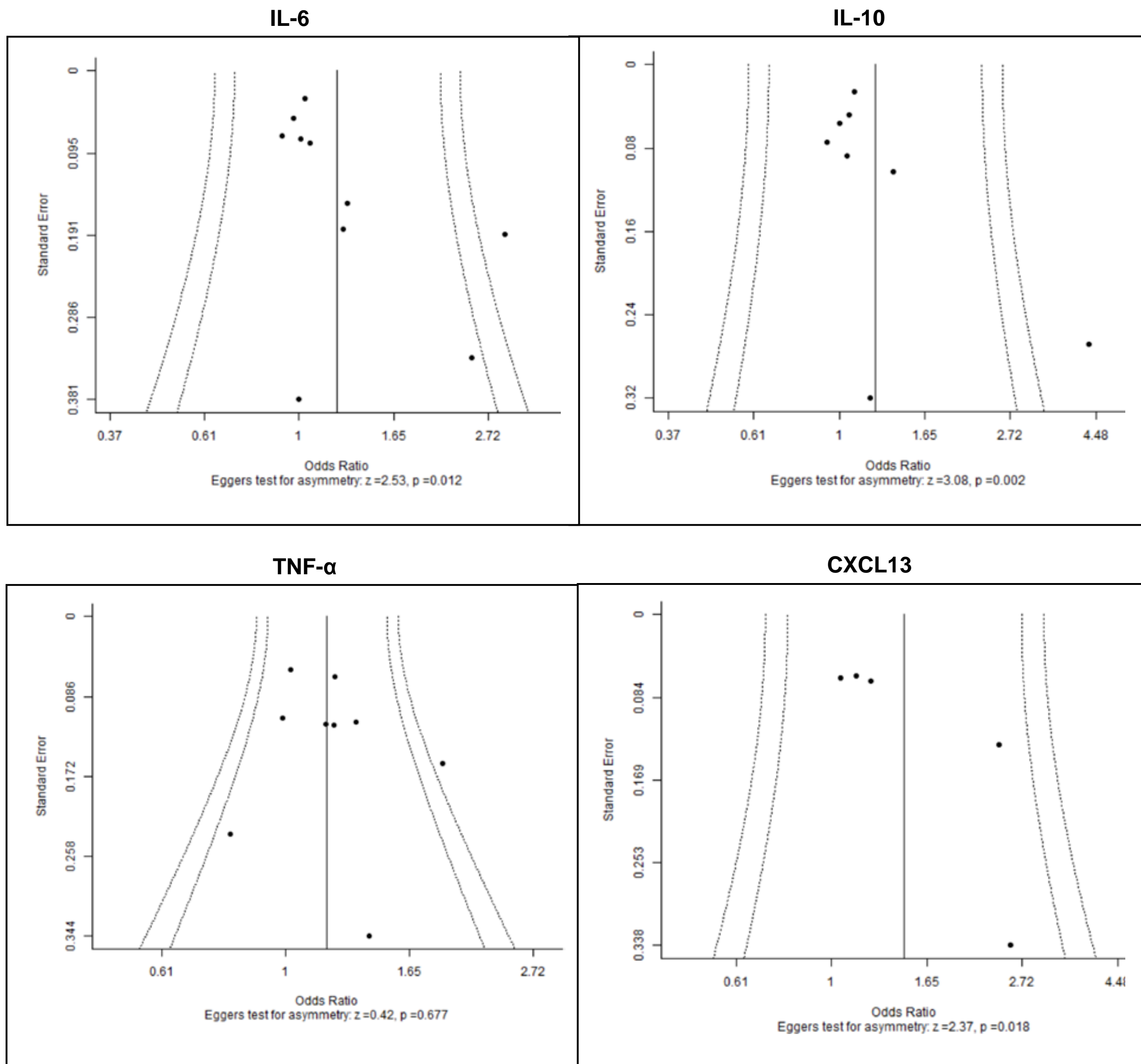
Supplementary Figure 2.4 continued: Forest Plots for NHL Subtype Outcomes.

sCD30



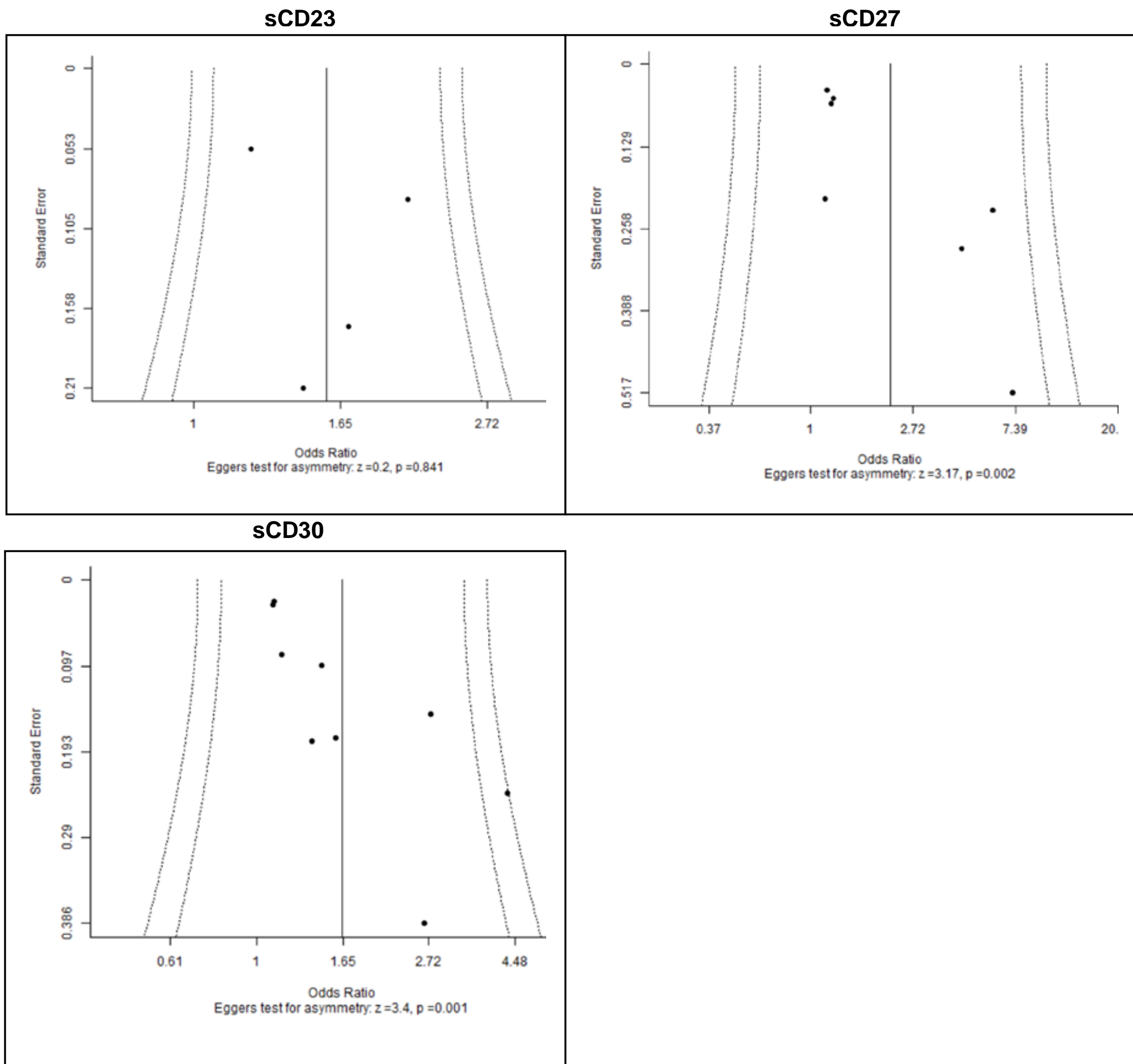
Supplementary Figure 2.4 Legend: Error bars indicate 95% CIs; size of the squares indicates the precision weight of each study in the random-effects meta-analysis. Diamonds indicate the summary ORs. Case and control sample sizes for were estimated for Purdue, 2009[22].

Supplementary Figure 2.5: Funnel Plots Assessing Publication Bias.



Supplementary Figure 2.5 Legend: Funnels indicate pseudo 90% (outer) and 95% (inner) confidence intervals accounting for between study heterogeneity estimated from the random effects models. Each plot assesses the potential influence of publication bias on the association between a given biomarker and the overall NHL outcome. Vertical reference lines drawn at point estimate for each analysis. Analyses and assessments of publication bias are not independent since most publications examined multiple biomarkers.

Supplementary Figure 2.5 continued: Funnel Plots Assessing Publication Bias.



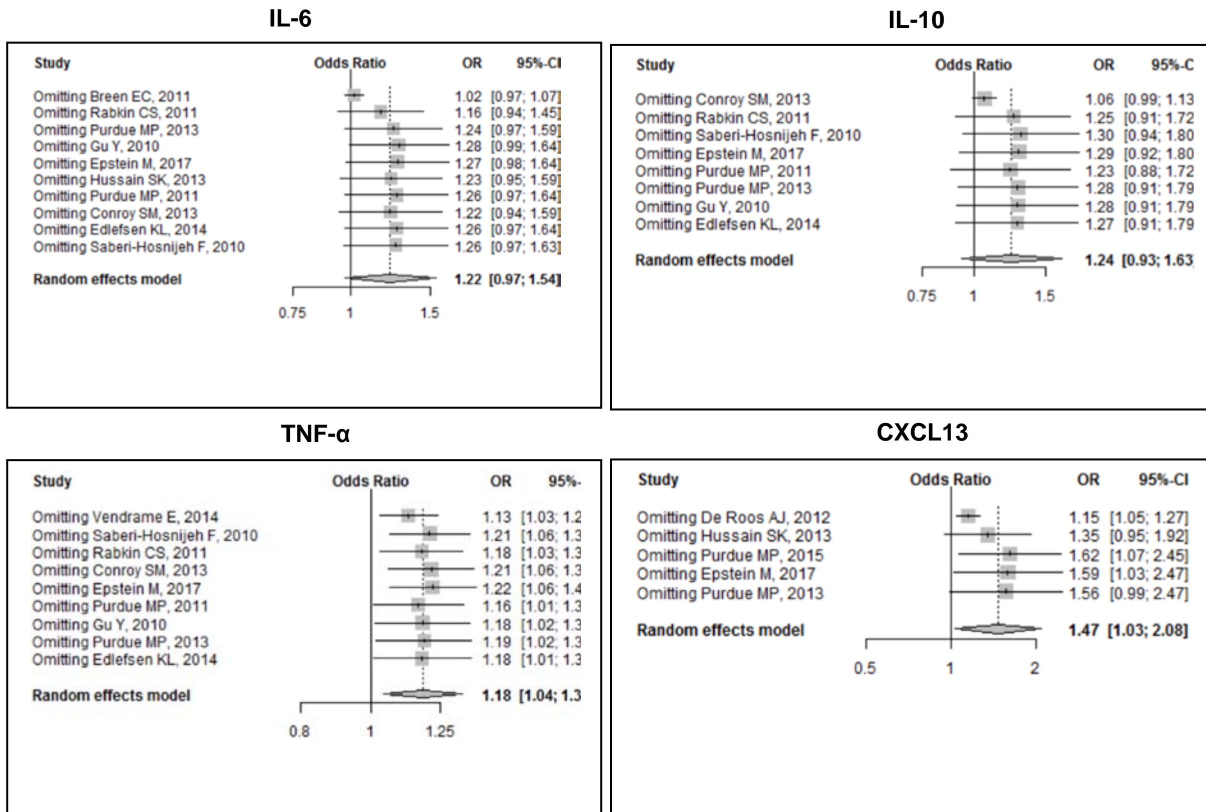
Supplementary Figure 2.5 Legend: Funnels indicate pseudo 90% (outer) and 95% (inner) confidence intervals accounting for between study heterogeneity estimated from the random effects models. Each plot assesses the potential influence of publication bias on the association between a given biomarker and the overall NHL outcome. Vertical reference lines drawn at point estimate for each analysis. Analyses and assessments of publication bias are not independent since most publications examined multiple biomarkers.

Supplementary Table 2.5: Bias Analyses: Egger's Regression P-values and Trim & Fill Analyses.

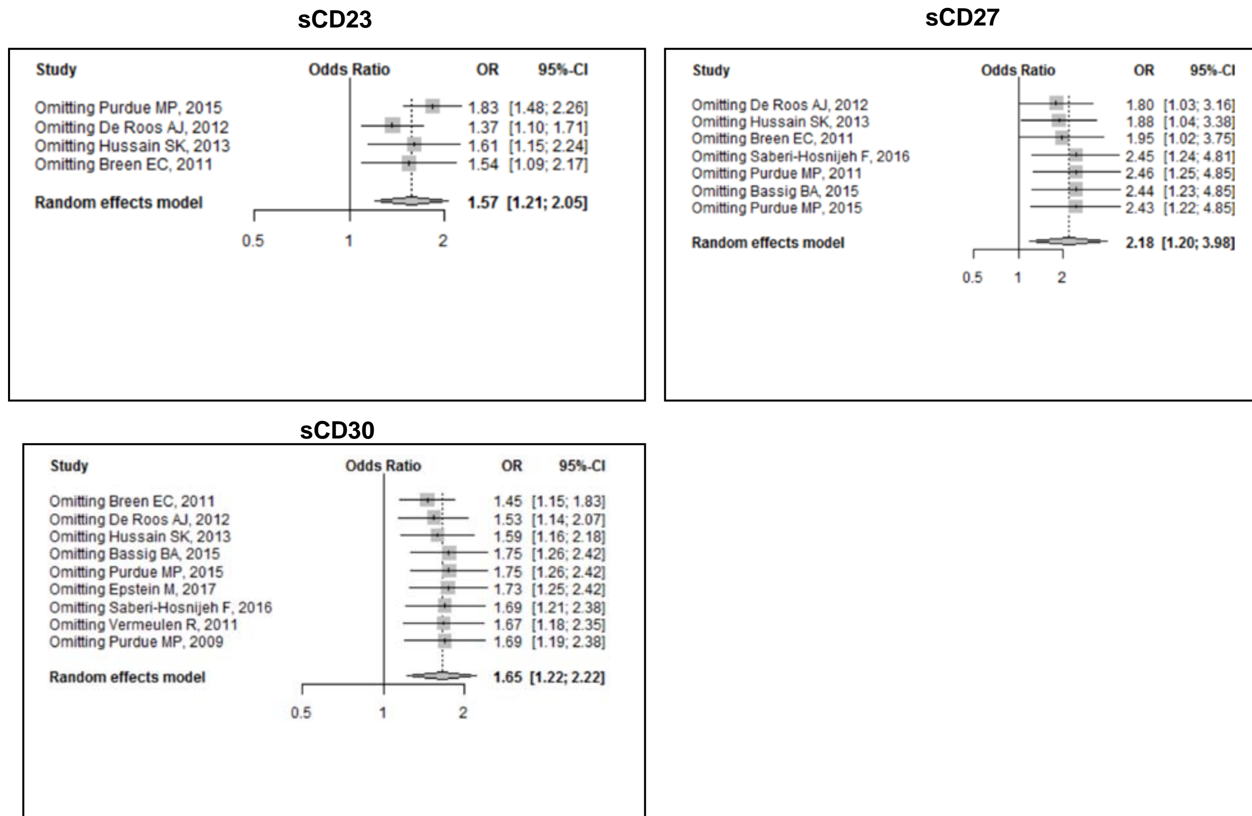
	Summary of Observed Data			After Trim-and-Fill *	
Biomarker	N	OR 95% CI	Egger's P	N	OR 95% CI
IL-6	10	1.22 [0.97, 1.54]	0.012	13	1.02 [0.76, 1.37]
IL-10	8	1.24 [0.93, 1.63]	0.002	9	1.09 [0.72, 1.65]
TNF-α	9	1.18 [1.04, 1.34]	0.677	10	1.12 [0.95, 1.33]
CXCL13	5	1.47 [1.03, 2.08]	0.018	6	1.32 [0.89, 1.94]
CD23	4	1.57 [1.21, 2.05]	0.841	6	1.25 [0.90, 1.74]
CD27	7	2.18 (1.20, 3.98)	0.002	9	1.39 (0.63, 3.06)
CD30	9	1.65 [1.22, 2.22]	<0.001	14	1.13 [0.78, 1.64]

*Trim-and-fill analyses impute data from studies predicted to have possibly been excluded due to publication bias and re-estimate the summary OR estimates. OR for trim-and-fill analyses represent the summary estimates including these hypothetically missing studies.

Supplementary Figure 2.6: Influence Diagnostics: Leave-One-Out Analyses.



Supplementary Figure 2.6 continued: Influence Diagnostics: Leave-One-Out Analyses.



Supplementary Figure 2.6 Legend: Leave-one-out diagnostic analyses showing the effect of removing a given study on the summary meta-analytic estimate. Random effects model refers to the summary OR estimate calculated in meta-analyses.

Supplementary Table 2.6: Results for All B-Cell NHL by Prediagnosis Time Interval: Comparing Early versus Late Biomarker Sample Collection

	Prediagnosis Time Interval							
		Early (6-10 yrs) *			Late (0-5 yrs) *		Meta-Regression Comparison of	
Analyte	N	OR 95% CI	I^2 95% CI	N	OR 95% CI	I^2 95% CI	Early vs. Late	P
IL-6†	2	1.04 [0.89, 1.21]	0 [-,-]	6	1.44 [1.00, 2.08]	87 [74, 93]	0.74 [0.39, 1.40]	0.352
IL-10	3	1.10 [1.03, 1.17]	0 [0, 34]	6	1.12 [0.98, 1.28]	44 [0, 78]	0.98 [0.84, 1.15]	0.817
TNF-α	2	1.19 [1.05, 1.34]	0 [-,-]	6	1.25 [1.01, 1.54]	85 [70, 93]	0.96 [0.67, 1.36]	0.813
CXCL13	2	1.33 [0.67, 2.62]	96 [88, 99]	2	2.69 [2.20, 3.28]	0 [-,-]	0.50 [0.24, 1.06]	0.070
sCD23	2	1.41 [0.98, 2.01]	92 [74, 98]	4	1.62 [1.21, 2.15]	89 [75, 95]	0.87 [0.54, 1.39]	0.559
sCD27	4	1.50 [0.96, 2.35]	88 [71, 95]	6	2.64 [1.34, 5.21]	96 [93, 97]	0.58 [0.24, 1.45]	0.247
sCD30	4	1.34 [1.00, 1.80]	87 [69, 95]	7	1.89 [1.28, 2.79]	94 [90, 96]	0.73 [0.42, 1.27]	0.262

Notes:

* The early category included studies that had categories that had upper bounds of the time intervals that were greater than 10 years (e.g. intervals such as >7, 5–13, 9–13, 8–15, 15–23 years prior to diagnosis) , while the late (0-5 year) category included studies with intervals that exceeded the upper bound of 5 years (e.g. 2–6, <7 years prior to diagnosis). For these analyses, estimates of associations covering both intervals completely, or nearly completely, were excluded.

† The IL-6 analyses included Vendrame, 2014 and Breen, 2011 which contain completely overlapping study subjects, but different assay technologies. We include them here, but not in the manuscript because the results are not substantially different with or without exclusion, and given the small sample size, the additional information dominates the small bias due to lack of independence for our assessment of associations by prediagnosis time periods and their differences.

Supplementary Table 2.7: Results for All B-Cell NHL among HIV-infected: Comparing HAART exposed versus Unexposed

	HAART Exposure Status							
		Unexposed*			Exposed*		Meta-Regression Comparison Of Unexposed vs. Exposed	
Analyte	N	OR 95% CI	I^2 95% CI	N	OR 95% CI	I^2 95% CI		P
IL-6 ‡	3	2.08 [1.17, 3.70]	75 [92, 18]	1	1.37 [0.75, 2.52]	0 [-,-]	1.52 [0.46, 5.01]	0.493
sCD23	2	1.75 [1.30, 2.36]	0 [-,-]	1	1.18 [0.61, 2.30]	0 [-,-]	1.48 [0.72, 3.08]	0.288
sCD27	2	4.72 [2.81, 7.93]	0 [-,-]	1	13.07 [1.87, 91.52] †	0 [-,-]	0.36 [0.05, 2.71]	0.322
sCD30	2	4.14 [2.71, 6.32]	11 [-,-]	1	1.55 [0.46, 5.26]	0 [-,-]	2.67 [0.73, 9.73]	0.137

Notes:
* The unexposed group comprises studies comprised 1 study including HAART naive participants. The other 2 studies with HAART unexposed participants adjusted for HAART, implying the estimate presented is among the reference group of HAART unexposed individuals, which we include in these analyses [54,137,138].
† In the publication assessing HAART exposed individuals [57], there were 9 HAART exposed cases and 37 controls, which led to an inflated estimate.
‡ These analyses included Vendrame, 2014 and Breen, 2011 which contain completely overlapping study subjects, but different assay technologies. We include them here, but not in the manuscript because the results are not substantially different with or without exclusion, and given the small sample size, the additional information dominates the small bias due to lack of independence for our assessment of HAART exposure associations and their differences. Analyses presented only if at least one stratum had 2 or more studies.

Supplementary Table 2.8: Multiple Comparisons Adjusted P-Values for Overall NHL

Multiple Comparisons		P-Value Adjustment		
Group	**Biomarker**	**Unadjusted**	**Stepdown Bonferroni***	**Dependent FDR***
Overall	CXCL13	0.0320	0.0959	0.1160
	IL-10	0.1375	0.1651	0.3564
	IL-6	0.0825	0.1651	0.2497
	TNF-α	0.0100	0.0498	0.0498
	CD23	0.0008	0.0055	0.0102
	CD27	0.0110	0.0498	0.0498
	CD30	0.0011	0.0068	0.0102
HIV-	CXCL13	0.0944	0.4429	0.4265
	IL-10	0.1741	0.4429	0.5267
	IL-6	0.6025	0.6025	1.0000
	TNF-α	0.0151	0.0908	0.1374
	CD23	0.0886	0.4429	0.4265
	CD27	0.1175	0.4429	0.4265
	CD30	0.0043	0.0298	0.0772
HIV+	CXCL13	0.0054	0.0161	0.0195
	IL-10	0.5683	0.5683	1.0000
	IL-6	0.0101	0.0201	0.0304
	TNF-α	<.0001	0.0003	0.0003
	CD23	0.0004	0.0017	0.0019
	CD27	<.0001	<.0001	<.0001
	CD30	<.0001	<.0001	<.0001
Meta-Regression	CXCL13	0.2181	0.6543	0.7917
	IL-10	0.9428	1.0000	1.0000
	IL-6	<.0001	<.0001	<.0001
	TNF-α	0.0052	0.0262	0.0317
	CD23	0.9959	1.0000	1.0000
	CD27	0.0413	0.1651	0.1873
	CD30	0.0029	0.0177	0.0268

Notes:

*Methods used described in Holm, 1979 and Benjamini & Yekutieli, 2001, were deemed appropriate for tests that may contain statistical dependence since biomarker biological functions are not independent.

3 Chapter III: Temporal Variation of Serum Biomarkers of Immune Activation and Inflammation in Non-Hodgkin Lymphoma in the Multicenter AIDS Cohort Study

3.1 Abstract

Background: Chronic inflammation and immune activation are thought to play a crucial role in the etiology of AIDS-associated non-Hodgkin lymphoma (AIDS-NHL). Temporal variation in biomarkers of inflammation and immune activation may indicate the evolution of latent etiologic processes during the development of non-Hodgkin lymphoma (NHL). We investigated the temporal variation in 24 biomarkers in the Multicenter AIDS Cohort Study (MACS) to explore the manner in which biomarker signatures may show association with AIDS-NHL.

Methods: Serum concentrations of 24 biomarkers of inflammation and immune activation were quantitated prospectively from blood samples of MACS participants. In longitudinal analyses of these biomarkers using linear mixed-models, we describe the temporal variation in these 24 biomarkers in the sera of HIV-infected individuals in the MACS who develop AIDS-NHL versus those who do not.

Results: We follow 1,918 HIV-infected men from seroconversion over 11,220 longitudinal person-visits in the Multicenter AIDS Cohort Study (MACS) from 1984 to 2009. During HAART naïve visits, we observe elevated geometric mean biomarkers for cases relative to controls for several biomarkers (IL-2, TNF-α, IL-6, sCD27, sIL-2Rα, IP-10, CXCL13, CRP), and higher rates of change (slopes) for cases relative to controls (sIL-6R, sTNFR2, IL-10). Following HAART initiation, we find that geometric mean levels are elevated for cases relative to controls (BAFF, TNF-α, sIL-2Rα, sTNFR2, IP-10, MCP-1, CRP) with estimates generally higher than those observed prior to HAART. No differences in slopes were detectable in our data following HAART.

Conclusion: Our results suggest that differences in the trajectories of several biomarkers are detectable between individuals who eventually develop AIDS-NHL relative to those who do not. These findings suggest that these biomarkers have potential clinical utility in risk stratification and risk prediction in HIV-infected populations.

3.2 Introduction

HIV associated non-Hodgkin lymphoma (AIDS-NHL) is consistently ranked as the HIV-related malignancy with the highest incidence both in the U.S. and other developed countries even in the era of wide-spread access to multi-agent highly active anti-retroviral therapy (HAART) [6,7]. As a cause of morality, AIDS-NHL accounts for the highest proportion of AIDS-related death [7–9]. The disparity in NHL incidence between HIV-infected individuals relative to the general population has long been documented to be large and persistent. AIDS-NHL risk in the pre-HAART era was documented to be up to a 60-fold multiple in incidence rate [10,139]. While the risk of developing AIDS-NHL has diminished with effective HAART [140], it remains elevated even among individuals receiving treatment [7,11,12]. HAART has been reported to improve overall survival outcomes, but disparities in survival outcomes are still observed in HIV-infected versus HIV-uninfected individuals [13]. In this study, we seek to describe potential AIDS-NHL etiologic factors in the context of HIV infection.

AIDS-NHL comprises several histologically heterogeneous subtypes. These include diffuse large B cell lymphoma (DLBCL), primary central nervous system lymphoma (PCNSL), Burkitt's lymphoma, and primary effusion lymphoma (PEL) [3,4]. The incidence of AIDS-NHL in the HAART era varies by both AIDS-defining and non-AIDS-defining subtypes [5], although the overwhelming majority of cases have been AIDS-defining, with DLBCL and Burkitt lymphoma accounting for a majority of cases reported in a recent U.S. population based registry study [6].

Among HAART receiving HIV-infected individuals, immune dysfunction and associated factors are the most consistently reported risk factors for AIDS-NHL [4]. These include low nadir CD4+ T cell count, high HIV viral load, and duration of HIV-infection [5,6,24]. Although AIDS-NHL cases attributable to viral factors are reduced in the HAART era, virus-linked AIDS-NHL incidence

continues to be elevated for HIV-infected persons on HAART relative to the general population [3]. Other morbidies associated with AIDS-NHL include obesity [25,26], and diabetes mellitus [27–29]. The association between behavioral and demographic risk factors and NHL have been documented in the MACS and in the general U.S. population, including recreational drug use [35] and tobacco use [36–38] as well as other demographic factors, most especially age [30–34].

There are two hypothesized pathogenic mechanisms responsible for AIDS-NHL [3,4]. One mechanism involves the dysregulated proliferation of Epstein-Barr virus (EBV)-transformed B-cells wherein impaired T-cell mediated regulation of B-cell growth results in B-cell hyperproliferation leading to EBV-positive AIDS-NHL subtypes [3]. The other mechanism involves chronic B-cell hyperactivation and associated processes that promote oncogenic mutations and translocations [4]. As such, biomarkers involved in B-cell activation, such as cytokines, chemokines, their receptors, and other molecules involved in immune activation and inflammation, likely serve as mediators of the multiple risk factors for AIDS-NHL that interact with the immune system.

Several serum biomarkers of immune activation and inflammation are altered prior to AIDS - NHL diagnosis including sIL-1R1 [14], IL-2 [141], IL-4 [138,141], IL-5 [138], IL-6 [54,141], IL-10 [54,141,142], IL-11 [14], IL-29 [14], CCL19, CXCL10/IP-10 [14,141], CXCL11, CXCL13, sCD23 [14,54,143], sCD30 [54,126], sCD44 [144], neopterin [141], IgA, M, G and κ and λ free light chains [145], CRP [146–148], and TNF-α [14,141], MCP-2, MIP-1δ/MIP-5/CCL15, IFN-α, GM-CSF [14]. A recent study in the MACS evaluating the effect of HAART on the biomarkers included in the current study found that within a year of initiating HAART therapy and achieving successful virologic suppression, many of these biomarkers return towards levels observed in uninfected individuals [128]. In fact, serum biomarker concentrations stabilized in the years following HAART initiation (for example IL-2, IL-6, IL-10, CXCL13), although there was evidence of persistent residual immune activation due to elevated

levels of CXCL10, CRP, sCD14, sTNFR2, TNF-α, sCD27, sGP130, IL-8, CCL13, BAFF, GM-CSF, and IL-12p70 [128].

The current study is an investigation of prospectively quantitated serum biomarkers in the MACS to characterize the temporal variation in the mean levels and slopes of temporal trajectories of serum biomarkers of immune activation and inflammation. We hypothesize that temporal trajectories of biomarkers of immune activation and inflammation will show differences in mean levels and slopes between eventual AIDS-NHL cases versus HIV-infected controls, even among individuals receiving HAART. To our knowledge, this is the first longitudinal assessment of the association of biomarker trajectories with AIDS-NHL in the MACS.

3.3 Materials And Methods

3.3.1 Study Design

The biomarker data included in this study were generated from two prior sub-studies in the MACS, a longitudinal subcohort [128] and a multi-time-point nested case-control study [146]. Details of the MACS have been described and published previously [39,40]. Briefly, the MACS is a prospective cohort study comprising men who have sex with men designed to investigate various aspects of HIV infection including the natural history of the disease, risk factors for acquiring disease, and the clinical expression of infection. Participants were recruited at four academic centers in the U.S. (Baltimore, Maryland, USA/Washington, District of Columbia; Chicago, Illinois; Los Angeles, California; Pittsburgh, Pennsylvania). The follow-up schedule consisted of semiannual study visits wherein serological, clinical, and behavioral data were collected via a variety of methods including blood samples, physical examinations, self-report in structured interviews, review of disease registries, and confirmatory medical chart review.

3.3.2 Participant Inclusion

For the subcohort, MACS participants were sampled from the main cohort for a cohort-wide study assessing the role of biomarkers of inflammation, immune activation, and microbial translocation in the progression of HIV infection and associated sequelae described elsewhere [128]. Participants were included if (1) they had known seroconversion dates, and available serum samples at annual visits (2) they were seroprevalent with available serology specimens within two years prior to HAART initiation. For all HAART recipients, serum samples were collected from study visits immediately before and after treatment initiation, and subsequently every two years post-treatment initiation. In our study, we included (1) all available AIDS-NHL cases with

biomarker measurements, and (2) all remaining subcohort members who were seroconverters or seroprevalent with at least one visit after known, or estimated, HIV-infection date wherein serum biomarkers were measured. All 24 biomarkers were measured among these participants.

For the nested case-control study [146], serum samples were collected from MACS participants enrolled between 1984–85 or 1987–1991. Cases (n=179) were defined as HIV-infected MACS participants diagnosed with AIDS-NHL before April, 2003, with at least one serum sample at a visit prior to AIDS-NHL diagnosis. For each AIDS-NHL case, an HIV-infected control was selected from among HIV-infected MACS subjects with no lymphoma diagnosis as of April 2003 (1:1 match). Each case was matched to a control on (1) the duration of HIV-infection if the date of HIV-seroconversion was known, or date of entry (within a year) into the MACS cohort if they were HIV-seroprevalent, and (2) the expected availability of their serum samples at time points comparable to those of cases within a year of collection. Serum samples were collected from three possible time intervals prior to NHL diagnosis at 0–1, 1–3, and >3 (at approximately 4) years prior to NHL, and comparable time-points for controls. A subset of 13 of the 24 biomarkers we consider here were measured among these participants: sCD27, CRP, CXCL13, GMCSF, IFN-γ, IL-1β, IL-2, IL-6, IL-8/CXCL8, IL-10, IL-12p70, CXCL10/IP-10, TNF-α.

When we combined records from both studies, if there were duplicated visits, we preferred the visit from the case-control study because of the higher likelihood of complete biomarker data due to the use of sample availability as a matching criterion for that study sample. In our combined analytic data set with duplicated visits removed, we have a total of 1,918 MACS participants among which 202 developed AIDS-NHL with 11,220 biomarker visits prior to AIDS-NHL diagnoses (end of follow-up for controls). Our data comprise a total of 13 biomarkers measured in both component data sets, and an additional 11 measured only in the subcohort.

3.3.3 Biomarkers

For the subcohort, biomarker assays have been described in detail elsewhere [128]. In summary, two multiplex assay platforms were used to quantify 24 serologic markers of inflammation, and immune activation. Serologic markers measured on the Meso Scale Discovery (MSD, Gaithersburg, MD) system included CCL 2/MCP1, CCL4/MIP-1β, CCL11/Eotaxin-1, CCL13/MCP-4, CCL17/TARC; CXCL10/ IP-10, IL-8/CXCL8 (Ultra-Sensitive Human Chemokine 7-Plex Kit); interferon gamma (IFN-γ) IL-1β, IL-2, IL-6, IL-8, IL-10, IL-12 p70, TNF-α, granular-macrophage colony-stimulating factor (GM-CSF) (UltraSensitive Human Pro-Inflammatory 9-Plex Kit). Biomarkers measured on the Luminex platform (Luminex, Austin, TX) included CXCL13/BLC-BCA1, B-cell activating factor (BAFF), sCD14, sCD27, sGP130, sIL-2Rα, sIL-6R, sTNFR2 according to the manufacturer's protocol (R&D Systems, Minneapolis, MN). A high-sensitivity immunonephelometric assay was used to quantitate complement reactive protein (CRP) in a reference laboratory.

To minimize plate-to-plate variability, samples from each individual were measured on the same plate; all biomarkers were assessed in centralized laboratories. For CCL11/Eotaxin-1, CCL13/MCP-4, CCL17/TARC, IL-8/CXCL8, CXCL10/IP-10, CCL2/MCP1, IL-10, IL12p70, IL-6, and TNF-α, intra-assay coefficient of variation (CV) medians ranged from 3.3-14.9, while inter-assay CVs ranged from 6.6 to 40.5; for BAFF/BLyS, CXCL13/BLC-BCA1, sCD14, sCD27, sgp130, sIL2ra, and sTNFR2 the medians of intra-assay CVs ranged from 3.9 to 5.2 while the inter-assay variability ranged from 8.8 to 29. All platforms showed acceptable performance for chemokines, soluble receptors, and some cytokines with high levels of percentages detected and variability measures [128].

Biomarker quantitation for the case-control study serum biomarker levels is described in a previous study [146]. Briefly, several biomarkers were measured using enzyme-linked immunosorbent assay (ELISA). sCD27 was quantitated using the PeliKine-compact ELISA kit and Toolset according to the manufacturer's protocol (CLB/Sanquin, Netherlands); complement reactive protein (CRP) was measured using the manufacturer prescribed high sensitivity protocol (Virgo CRP 150, Hemagen, Columbia, Maryland); serum CXCL13 was measured according to manufacturer's protocol (R&D Systems, Minneapolis, MN); IL-6 was measured using an ultrasensitive assay (Biosource/Invitrogen, Carlsbad, California); IL-10 was measured using a human IL10- specific assay (Biosource International, Camarillo, California).

Another set of immune markers were assessed using Luminex-based multiplexed immunoassays. IL-1β, IL-2, IL-8, IL-12p70, GM-CSF, IFN-γ, TNF-α were determined using the Fluorokine® MAP Human Inflammation Kit (R&D Systems, Minneapolis, MN, USA). IP-10/CXCL10 was determined using the Procarta Cytokine Assay (Affymetrix, Santa Clara, CA, USA)[58].

3.3.4 AIDS-NHL Case Ascertainment

In accordance with MACS protocol, all cancer diagnoses were ascertained at each follow-up visit through structured interviews of subjects followed by medical chart abstraction, pathology reports from autopsy, and/or matching to state cancer registries [11]. As of 2006, all current and historically diagnosed cancers were classified by site and histology codes of the International Classification of Diseases for Oncology, 3rd edition (ICD-O-3).

3.3.5 Covariates

Covariates included measured variables that were known *a priori* to be confounders in assessments of biomarker associations with AIDS-NHL [54,57,141], and any additional variables that are known to be risk factors for AIDS-NHL, as described above, while also plausibly potentially affecting biomarker levels either directly or through other factors. As such, potential time-dependent confounders included age at each visit as a continuous variable; self-reported tobacco, cocaine, and stimulant drug use (methamphetamine) in the time interval preceding the current visit; self-reported frequency of alcohol use (captured as no use, less than twice weekly, and more than twice weekly since the last visit); body mass index (categorized as BMI>30 kg/m^2 at each visit); HCV infection status at each current visit (negative versus positive) [128,149]. CD4+ T cell count was also included as a time-dependent confounder at each visit. Potential time-invariant confounders included race categorized as non-Hispanic white versus non-white, due to category frequencies, and MACS center. Finally, year of HIV infection was considered a covariate to control for diagnoses occurring during different treatment and medical practice pattern eras, and to address differences in duration of HIV infection when included with age.

Given that we combined data from a matched nested case-control with a subcohort, we constructed a covariate indicating whether an observation came from the case-control study or subcohort. In addition, we consider the risk set comprising matched cases and controls (with subcohort data considered as a single risk set) as a random effect covariate to be included in our models to account for the complex matched nature of the included case-control data[150].

3.3.6 Statistical Methods

We followed participants from date of first known HIV infection until AIDS-NHL diagnosis date, or last visit date with biomarker data. For participants who seroconverted after enrollment into

the cohort, the start of follow-up was HIV seroconversion date. For participants who were enrolled into the cohort HIV seroprevalent the start of follow-up was cohort entry date, meaning their observed time-at-risk is left truncated. We summarized and tabulated salient characteristics of the analytic data set for descriptive purposes and to list potential confounders. Covariates were considered potential confounders on the basis of *a priori* considerations of variables that are known risk factors for the AIDS-NHL that may also plausibly differentially affect biomarker levels, and included all such measured covariates in our subsequent regression modeling. We also graphically explored the outcome biomarker variable distributions for concordance with the normal probability distribution. Goodness of fit tests showed that the distributions were non-normal, even upon transformation.

For each biomarker we fit a linear mixed model with the natural log-transformed biomarker as the outcome, and disease groups (AIDS-NHL cases versus HIV+ controls) as the predictor of primary interest. Our models included Kenward-Rogers corrections for degrees of freedom that have been shown to confer robustness to departures from normality for data sets larger than 30 observations [151,152]. We parameterized our regression models as follows: first, we included a time-invariant indicator variable for case versus control status that provides an estimate of the difference in mean log transformed biomarker levels at the beginning of follow-up. Then, to estimate changes in biomarker mean levels following HAART relative to pre-treatment levels, we included an indicator variable marking all visits that followed the estimated HAART receipt date. For estimates of the slopes of temporal trajectories of the biomarker outcomes, we included a time variable (scaled to 10's of years) with the origin (t=0) at date of first known HIV infection. Finally, we included spline terms, implemented as an interaction of time by case status, and time by cases status by HAART, that allowed calculation of contrasts estimating differences in temporal slopes by case status prior to and following HAART receipt. Since biomarkers were treated as natural log transformed outcome variables in regression models, we

present exponentiated regression parameters that represent geometric mean ratios of the biomarker slopes and mean levels between comparator and reference levels.

Given the large number of comparisons in our analyses, we applied a false discovery rate correction (α=0.05) that is robust to dependency to all the p-values of the regression parameter estimates that we present in our results tables [136].

Analyses were performed using SAS procedures in SAS 9.2 Cary, NC.

3.4 Results

We describe participant characteristics of our analytic sample in Table 3.1, at first known HIV-positive visit, or closest visit. Biomarker availability and summary statistics at the same time-point appear in Table 3.1. We combined the 179 case-control pairs (n=358) from the nested case-control study with HIV-infected participants from the subcohort (n=1,623), and removed duplicated person-visits for a combined total of 1,918 participants with at least one post-seroconversion (or seroprevalent) visit wherein biomarker concentrations were quantified. The data set included 202 AIDS-NHL cases among these participants with a total of 638 person visits, and 10,582 person visits among controls. We followed participants up to a maximum of 31.5 years post seroconversion, with a median follow up during HIV infection of 7 years (IQR: 4-11 years). The majority of AIDS-NHL cases were of DLBCL subtype (52%). AIDS-NHL cases were found to be of comparable age 37 (IQR: 32-43 years) to controls, 39 (IQR: 33-44 years). A higher proportion cases were of white race (83%) relative to HIV+ controls (62%), with slightly lower tobacco use (33%) relative to controls (41%). A higher proportion of AIDS-NHL cases tended to have lower CD4+ T-cell counts, 51% with <400 cells/mL versus 38% for HIV+ controls.

Results from our longitudinal analyses are summarized in subsequent tables (Table 3.3 through Table 3.6). Each of the regression parameter estimates is exponentiated so that the results are geometric mean ratios for cases relative to controls. For each biomarker we present the ratio of geometric means among HAART naïve visits, that is, the difference between cases and controls in regression intercept; then we compare cases and controls on by their biomarker slopes prior to HAART receipt. We then present the ratio of biomarker geometric means at HAART receipt, that is, the difference between cases and controls in regression intercepts at HAART receipt, along with corresponding estimates of the differences in the slopes for cases relative to controls following HAART.

Estimates of mean levels: Among cytokines (Table 3.3), when we compare geometric mean levels of biomarkers among HAART naïve visits, we find that IL-2 (MR: 1.09 95%CI: 1.00, 1.19), TNF-α (MR: 1.17 95%CI: 1.07, 1.30) and IL-6 (MR: 1.12 95%CI: 1.01, 1.23) levels are elevated for eventual AIDS-NHL cases relative to HIV-infected controls. Still considering cytokines (Table 3.3), following HAART exposure, we find that visits among AIDS-NHL cases have higher geometric means levels of BAFF (MR: 1.41 95%CI: 1.12, 1.78) and TNF-α (MR: 1.41 95%CI: 1.02, 1.94) relative to controls.

Table 3.4 summarizes geometric mean comparisons for serum soluble receptors. We find elevated levels of the soluble receptor sCD27 (MR: 1.11 95%CI: 1.03, 1.21) and sIL-2Rα (MR: 1.16 95%CI: 1.02, 1.32) during HAART naïve visits for AIDS-NHL cases relative to HIV-positive controls. Biomarker means following HAART are elevated for sIL-2Rα (MR: 1.47 95%CI: 1.14, 1.91), and sTNFR2 (MR: 1.34 95%CI: 1.03, 1.73) among cases relative to controls.

In Table 3.5 we compared chemokine geometric means for HAART naïve visits by disease groups and found elevated levels of IP-10 (MR: 1.23 95%CI: 1.07, 1.42) and CXCL13 (MR: 1.30

95%CI: 1.18, 1.43) for AIDS-NHL cases relative to HIV-positive controls. Following HAART exposure, higher geometric mean levels of IP-10 (MR: 1.91 95%CI: 1.26, 2.92) and MCP-1 (MR: 1.31 95%CI: 1.01, 1.70) were observed for AIDS-NHL cases relative to controls.

Lastly, CRP geometric mean levels (Table 3.6) were elevated for AIDS-NHL cases relative to controls during HAART naïve visits (MR: 1.14 95%CI: 1.00, 1.29). CRP levels remained elevated for cases versus controls following HAART exposure (MR: 1.66 95%CI: 1.08, 2.53; Table 3.6).

Estimates of slopes over time: We estimate biomarker slopes (changes per 10 years) for HAART naïve visits and found cases having higher slopes than controls for trajectories of sIL-6R (MR: 1.14 95%CI: 1.00, 1.29), sTNFR2 (MR: 1.25 95%CI: 1.04, 1.51), and IL-10 (MR: 1.57 95%CI: 1.19, 2.06). Following HAART receipt, visits for AIDS-NHL cases are comparable to those of HIV-infected controls and we find no differences in slopes.

3.5 Discussion

In this study of prospectively collected serum biomarkers of immune activation and inflammation we found two general patterns. First, several biomarker geometric mean levels (IL-2, TNF-α, IL-6, sCD27, sIL-2Rα, IP-10, CXCL13) are elevated for eventual AIDS-NHL cases relative to HIV-infected controls prior to HAART receipt, and these differences tend to widen following HAART receipt. Second, we find that temporal rates of changes (slopes) of biomarker levels are different by case status during the HAART naïve period, but no differences in rates of change are observed following HAART receipt. These two patterns in our results combined mean that the trajectories of circulating markers of immune activation and inflammation are altered among those who develop go on to develop NHL relative to HIV-infected controls who do not, with

AIDS-NHL cases often having higher biomarker levels and higher rates of temporal change relative to HIV-infected controls.

We considered 8 cytokines and described the changes in aspects of their trajectories over time, with notable variation in the trajectories of BAFF, IL-2, TNF-α and IL-6 (Table 3.3). BAFF is a member of the TNF ligand family and plays important roles in B-cell maturation functions and is implicated in the development and perpetuation of malignancies [153]. It has been shown that tumor cells of all NHL histological subtypes express several known BAFF receptors [154], and in the context of HIV-infection in the MACS, administration of successful HAART tends to attenuate BAFF levels [128]. In our sample, prior to HAART receipt, we observe no significant differences in the mean level and rate of change aspects of the trajectory of BAFF levels when comparing cases to controls. However, following HAART, cases are observed to have higher BAFF levels relative to controls. This finding suggests that those individuals who eventually become AIDS-NHL cases may not experience reversion of BAFF to normal levels in response to HAART, which in turn may increase the risk of lymphoma.

IL-2 is a pro-inflammatory cytokine that stimulates the proliferation of activated B cells [155]. IL-2 primarily functions to mount a cytotoxic immune response to potentially malignant cells through the promoting proliferation of natural killer and lymphokine-activated killer cells [155]. These two facts taken together imply that we could expect to observe elevated IL-2 levels among individuals that go on to develop NHL due to chronic B-cell stimulation, as well as increased serum concentrations in IL-2 in response to cancerous cells that may lower the incidence of malignancies. In accordance with both these facts, results from a prior epidemiologic study among treated HIV-infected individuals found an inverse association between IL-2 levels and NHL [48], while another prospective study found elevated levels of IL-2 associated with NHL among HIV-uninfected subjects [61]. We find elevated levels of IL-2 among cases relative to

controls prior to HIV treatment with HAART, but no other differences in the temporal trajectory of IL-2, which supports the notion that increased levels of IL-2 are associated with NHL through B-cell stimulatory effects.

There are several epidemiologic studies of TNF-α, and IL-6, that we summarize in a meta-analysis [156]. These studies indicate that TNF-α and IL-6 levels are elevated prior to NHL diagnosis. We find that both biomarkers are elevated during HAART naïve visits for cases relative to controls. The slopes of both biomarkers exhibit no differences by case status pre- or post- treatment. Post-HAART mean levels of TNF-α are elevated for cases relative to controls. Taken together, these observations indicate that TNF-α and IL-6 levels remain different between eventual AIDS-NHL cases relative to controls from HIV-seroconversion without much variation, even though the post-HAART case-control contrast for IL-6 does not quite reach statistical significance. It is thus plausible that a cumulative effect of consistently higher levels of TNF-α and IL-6 may generate or support lymphomagenic processes or may indicate non-modifiable immunologic differences between AIDS-NHL cases relative to controls.

We compared the trajectories of a total of 6 soluble receptors (sCD14, sCD27, sGP130, sIL2R-α, sIL6R, sTNFR2; Table 3.4) and found that aspects of the trajectories of sCD27, sIL2R-α, sIL6R, and sTNFR2 differed by case-control status. We assessed temporal slopes and levels of sCD27, a soluble form of CD27 [116,117]. sCD27 concentrations increase in circulation when CD27 is proteolytically cleaved from cell membranes to produce its soluble form. CD27 is involved in B cell activation [118] and B-cell secretion of immunoglobulins [119] thereby being implicated in B-cell hyperproliferation-related malignancy. We find elevated levels of sCD27 among AIDS-NHL cases relative to controls during HAART naive visits, with no differences in any aspects of the trajectory following HAART receipt. A possible explanation for this observation is that since HAART has been observed to reduce sCD27 levels in successfully treated HIV-infected

individuals relative to those who are untreated [128], there is a threshold elevation in sCD27 concentrations, attained only pre-treatment, that confers increased NHL risk.

sIL-2R- α is recognized as a tumor marker for malignant lymphomas in general, including the histological subtypes that comprise NHL [155,157]. The α chain (CD25) of the IL-2 receptor on the cell membrane surface is cleaved by enzymatic proteolytic processes [158–160] and is ultimately detected in sera as sIL-2R-α [155,157,158]. Lymphoma cells expressing IL-2R combined with proteinases in tumor microenvironment have been shown to be correlated with elevated levels of sIL-2R in sera [157]. In B-cell non-Hodgkin lymphomas, serum sIL-2R is seen to be elevated in HIV-uninfected patients with DLBCL and follicular lymphoma (FL) [157,158,161–163] and has been studied for its utility as a prognostic factor for both DLBCL and FL [158]. We find that levels of sIL-2R-α are elevated for cases relative to controls during HAART naïve visits, as well as, to a greater extent, post-HAART. This indicates that the response of sIL-2R-α levels to HAART is not adequate to reduce the risk resulting from chronically elevated serum concentrations of this biomarker.

The sIL-6R/IL-6 complex is involved in B cell proliferation and differentiation, cell survival, and tumor growth [84,85]. We find that the trajectories of sIL-6R differ between cases and controls only in slopes prior to treatment for HIV. These observations can be taken as further evidence that the sIL-6R/IL-6/GP130 B-cell signaling processes may be involved in NHL etiology over extended periods prior to diagnosis as has previously been reported [164].

TNFR2 is a soluble fragment of TNF receptor 2 (TNFR2) that mediates the pro-inflammatory effects of TNF-α [47]. TNFR2 is expressed by immune cells and signals activation of nuclear factor κ-B pathway (NF-κB) which in turn mediates B-cell maturation, proliferation, and survival [47]. Abberant NF-κB activation can promote chronic lymphocyte proliferation and survival and

has been recognized as an important pathogenetic factor in lymphoma that plays an important role in NHL pathogenesis [47,165]. Tenuous associations between serum sTNFR2 and NHL have been described in the NYU-WHS cohort [61], the PLCO study [166], and a recent study reported a strong sTNFR2 - NHL association that was detectable 8-13 years post serum sample collection and was consistent across NHL subtypes among uninfected individuals [47]. We find significantly increased slopes of sTNFR2 concentrations for cases relative to controls prior to HAART initiation, and that mean levels remain elevated for cases relative to controls following HAART initiation, implicating the residual elevation [128] in sTNFR2 in NHL etiology. Our results comport with prior literature and provide support for a role of sTNFR2 in lymphomagenesis [47,165].

We assessed the levels of nine chemokines (IL-8, IP-10, MCP-1, MCP-4, MIP-1β, TARC, CXCL13, EOTX, IL-10) and the acute phase reactant, complement reactive protein, and found differences by case status in the trajectories of IP-10, MCP-1, CXCL13, and IL-10. We considered IP-10 (CXCL10), which acts upon activated T cells and macrophages, and has also be shown to stimulate HIV replication in monocyte-derived macrophages and lymphocytes, and reduces the amount of MIP-1α, MIP-1β, and RANTES produced by activated peripheral blood lymphocytes [167]. In this way, IP-10 can thus be linked to B-cell lymphomagenesis through the chronic immune activation engendered by HIV-infection [168,169]. In our study sample, we find increased geometric means among NHL cases relative to HIV-infected controls during HAART naïve visits, and an even greater difference following HAART.

We also assessed trajectories of chemokines of the C-C motif including MCP-1 (CCL2), MCP-4 (CCL13), MIP-1β (CCL4), and TARC (CCL17) and only found MCP-1 levels to be higher following therapy for HIV-infection relative to visits with no treatment. We detected no other differences by case status in aspects of the trajectories of C-C motif chemokines. While the relationship between NHL and these chemokines has not been investigated, indirect evidence

for the role of chemokines of the C-C motif and their receptors in the pathogenesis of NHL exists in the form of modest NHL-polymorphism associations [170]. The differences in the trajectory of MCP-1 by case status may signify underlying immunologic and pathogenic processes that are yet to be elucidated fully.

For CXCL13, it is thought that aberrant expression may be involved in the pathogenesis of B cell lymphoma through abnormal chemotaxis of B cells to tissues or abnormal B cell activation [109]. Overexpression of the receptor-ligand pair CXCR5/CXCL13 has been observed in B cell chronic lymphocytic leukemia [110], and follicular lymphoma cells have been seen to secrete CXCL13 [111], indicating the possibility that CXCL13 can be a biomarker for NHL and these subtypes. In the context of HIV-infection, aberrant expression of the receptor-ligand pair CXCR5/CXCL13 has been seen in B cells [171]. Increased serum CXCL13 levels have been observed among HIV-positive relative to seronegative individuals [127], and it is reported that AIDS-NHL tumors express the CXCL13 receptor CXCR5 [109]. We observe elevated levels of CXCL13 among cases relative to controls prior to HAART receipt. These results indicate a possible etiologic role for CXCL13 in lymphomagenesis particularly in the context of untreated or uncontrolled HIV infection.

We assessed IL-10, a pleiotropic cytokine, that has stimulatory effects on B cells and is linked to the development of NHL through promoting chronic B cell activation [92–94] . Malignant NHL cells are also known to secrete IL-10 [96,97]. We have previously summarized results from epidemiologic studies showing elevated levels of IL-10 preceding NHL [156]. In this study, we observe IL-10 levels exhibiting divergent slopes between cases (steeper slopes) and controls in the period prior to HAART initiation, and no detectable differences between cases and controls in the trajectory of this biomarker after treatment for HIV, implying a reduction in the effect of this biomarker on NHL risk in response to HAART.

A limitation of our study is that since some of our biomarker data, comprising the majority of controls, were repurposed to assess a novel outcome, our biomarker measurements are not strictly collected at regular intervals for all participants meaning that we rely the longitudinal regression model to interpolate a trend over missing data. Relatedly, sampling probabilities are unknown for observations included in our study from the superset of MACS observations. This limits our inferences to the sample we include here, limiting generalizations to the main cohort or beyond. Having mentioned these weaknesses, there are also several strengths of our study. We were able to prospectively assess the experience of a group of men over a substantial follow up which reduces biases due to the independent collection of biomarker exposures and outcomes. The centralized nature of the specimen testing meant that random variability in the measures was reduced; the quality and sensitivity of the assay instrumentation meant that we were able to receive measures with the highest possible detection capability. We also included a comprehensive set of potentially confounding variables with which we could adjust our estimates, lending robustness of our results.

In conclusion, our prospective investigation into 24 markers of immune activation, and inflammation provides evidence that levels of several biomarkers are altered relative to controls over extended periods of time, and while HAART tends to have a normalizing effect on the biomarker levels, there still remain differences between cases and controls that may be consequential for AIDS-NHL disease processes.

3.6 Acknowledgements

This research was supported by the National Institutes of Health grant T32 CA09142, the Center for HIV Identification, Prevention, and Treatment (CHIPTS) NIMH grant P30MH058107; and the UCLA Center for AIDS Research (CFAR) grant 5P30AI028697, Core H. The ideas and opinions expressed herein are those of the authors, are not in any manner endorsed by the National Institutes of Health, the University of California at Los Angeles or their Contractors and Subcontractors.

3.7 Supporting Information

No additional Supporting Information

3.8 Disclosures

The authors of the manuscript have no conflicts of interest to disclose.

3.9 Tables

Table 3.1 Participant Characteristics at Beginning of Follow-up (HIV-seroconversion)

Variable	Category	Case Status		
		NHL+ (n=202)	**NHL- (n=1,716)**	**Total (n=1,918)**
Age	Mean (SD)	38 (8)	39 (8)	39 (8)
	Median (Q1-Q3)	37 (32-43)	39 (33-44)	38 (33-44)
Calendar Duration (Years)	Mean (SD)	5 (3)	9 (6)	8 (6)
	Median (Q1-Q3)	4 (4-5)	7 (4-12)	7 (4-11)
Person Visits (Total)	Mean (SD)	3 (3)	6 (5)	6 (5)
	Median (Q1-Q3)	3 (2-3)	5 (3-8)	4 (3-7)
HAART Received	No	169 (84)	374 (22)	543 (28)
	Yes	33 (16)	1,342 (78)	1,375 (72)
Race	Non-white	35 (17)	649 (38)	684 (36)
	White	167 (83)	1067 (62)	1234 (64)
BMI (kg/m3)	Overweight/normal/thin (<30.0)	189 (95)	1481 (87)	1670 (87)
	Obese (30.0+)	10 (5)	227 (13)	237 (12)
	Missing	3	8	11 (1)
NHL Histological Subtype	Diffuse large B-cell lymphoma	106 (52)		106 (52)
	Malignant lymphoma, non-Hodgkin, NOS	62 (31)		62 (31)
	Burkitt lymphoma/Burkitt cell leukemia	21 (10)		21 (10)
	Lymphoplasmacytic lymphoma	3 (1)		3 (1)
	B-cell chronic lymphocytic leukemia/small lymphocytic lymphoma	2 (1)		2 (1)
	Follicular lymphoma	2 (1)		2 (1)
	Malignant lymphoma, mixed small and large cell, diffuse	1 (1)		1 (1)
	Primary effusion lymphoma	1 (1)		1 (1)
	Peripheral T-cell lymphoma, NOS	1 (1)		1 (1)
	Anaplastic large cell lymphoma, T- or Null-cell	1 (1)		1 (1)
	Unknown	2 (1)		2 (1)
CD4 cell count/mm3 (Categories)	<400	100 (51)	633 (38)	733 (38)
	400-600	46 (23)	423 (25)	469 (24)

Variable	Category	Case Status NHL+ (n=202)	NHL- (n=1,716)	Total (n=1,918)
	>600	51 (26)	609 (37)	660 (34)
	Missing	5	51	56 (3)
CD4 cell count/mm3	Mean (SD)	440 (274)	533 (318)	523 (315)
	Median (Q1-Q3)	399 (231-605)	497 (307-719)	488 (300-709)
Hepatitis C Infection	HCV Unexposed	179 (89)	1523 (89)	1702 (89)
	HCV Exposed	23 (11)	191 (11)	214 (11)
	Missing	0	2	2 (0)
Smoking Status	Non-smoker	130 (67)	966 (59)	1096 (57)
	Smoker	64 (33)	681 (41)	745 (39)
	Missing	8	69	77 (4)
No. Alcohol Drinks Per Week	None	28 (14)	273 (16)	301 (16)
	1 to 3 drinks/week	80 (41)	717 (43)	797 (42)
	4 to 13 drinks/week	69 (35)	503 (30)	572 (30)
	More than 13 drinks/week	19 (10)	163 (10)	182 (9)
	Missing	6	60	66 (3)
Uncontrolled Diabetes Mellitus	No	202 (100)	1663 (97)	1865 (97)
	Yes	0	52 (3)	52 (3)
	Missing	0	1	1 (0)
Stimulant Use	No	146 (74)	1186 (72)	1332 (69)
	Yes	50 (26)	466 (28)	516 (27)
	Missing	6	64	70 (4)
MACS Site	Baltimore	51 (25)	408 (24)	459 (24)
	Chicago	50 (25)	392 (23)	442 (23)
	Los Angeles	76 (38)	560 (33)	636 (33)
	Pittsburgh	25 (12)	356 (21)	381 (20)

Table 3.2: Summary Statistics of Analytes in Component Data Sets by Case Status Over Entire Follow-Up.

	Sub-Cohort					Nested Case-Control				
		Cases		Controls			Cases		Controls	
	% Detect-able	Geometric Mean (SD)	Min-Max	Geometric Mean (SD)	Min-Max	% Detect-able	Geometric Mean (SD)	Min-Max	Geometric Mean (SD)	Min-Max
Cytokines										
BAFF	100	7.91 (0.47)	7.11 - 10.35	7.79 (0.40)	5.61 - 10.94	--	--	--	--	--
GMCSF	35	0.79 (0.63)	0.18 - 5.26	0.86 (0.67)	0.18 - 5.99	90	0.29 (0.31)	0.01 - 2.13	0.22 (0.27)	0.01 - 2.41
IFNg	43	1.07 (0.56)	0.34 - 3.93	1.07 (0.54)	0.23 - 6.72	90	0.46 (0.47)	0.00 - 3.60	0.34 (0.42)	0.00 - 2.57
IL12P70	87	1.59 (1.22)	0.34 - 7.55	1.45 (1.06)	0.22 - 7.82	90	0.22 (0.51)	0.03 - 3.08	0.26 (0.60)	0.03 - 4.76
IL1B	17	0.50 (0.48)	0.09 - 4.27	0.54 (0.55)	0.08 - 7.27	90	0.35 (0.57)	0.01 - 5.13	0.31 (0.36)	0.01 - 2.46
IL2	73	0.70 (0.38)	0.22 - 2.67	0.66 (0.43)	0.10 - 5.47	90	0.45 (0.47)	0.00 - 2.24	0.29 (0.45)	0.00 - 2.87
TNFA	100	2.63 (0.5)	1.74 - 5.39	2.52 (0.52)	0.65 - 7.15	90	2.41 (0.56)	0.08 - 4.23	2.10 (0.60)	0.04 - 4.53
IL6	99	0.79 (0.47)	0.18 - 3.57	0.80 (0.55)	0.06 - 7.71	94	0.89 (0.48)	0.08 - 3.63	0.67 (0.36)	0.08 - 2.33
Soluble Receptors										
sCD14	100	14.81 (0.32)	13.75 - 15.95	14.74 (0.31)	11.97 - 16.12	--	--	--	--	--
sCD27	100	9.74 (0.52)	8.68 - 11.70	9.50 (0.47)	7.78 - 14.09	99	6.24 (0.45)	4.94 - 8.31	6.03 (0.38)	4.98 - 7.47
sGP130	100	12.45 (0.18)	11.84 - 12.98	12.50 (0.25)	4.83 - 14.77	--	--	--	--	--
sIL2RA	100	7.88 (0.49)	6.88 - 10.26	7.49 (0.45)	4.75 - 10.13	--	--	--	--	--
sIL6R	99	10.95 (0.25)	10.34 - 11.66	10.85 (0.29)	9.07 - 13.13	--	--	--	--	--
sTNFR2	100	8.31 (0.49)	7.02 - 10.59	8.05 (0.45)	6.09 - 11.09	--	--	--	--	--

Chemokines										
IL8	100	2.86 (1.00)	1.44 - 7.79	2.91 (0.93)	0.06 - 7.82	90	2.87 (0.91)	0.87 - 8.07	2.67 (0.90)	0.25 - 7.51
IP10	100	6.02 (0.73)	4.58 - 8.14	5.69 (0.77)	3.12 - 8.36	59	7.98 (0.65)	5.55 - 9.90	7.43 (0.64)	5.55 - 8.94
MCP1	100	6.53 (0.40)	5.67 - 7.82	6.36 (0.42)	4.07 - 8.47	--	--	--	--	--
MCP4	100	6.61 (0.51)	4.01 - 8.20	6.64 (0.47)	3.64 - 8.47	--	--	--	--	--
MIP1	100	4.68 (0.71)	1.85 - 6.75	4.78 (0.68)	1.76 - 9.15	--	--	--	--	--
TARC	100	6.14 (0.68)	4.73 - 8.53	6.19 (0.72)	1.31 - 9.18	--	--	--	--	--
CXCL13	99	5.88 (0.48)	4.60 - 7.76	5.75 (0.37)	2.69 - 9.02	100	4.83 (1.10)	0.00 - 7.68	4.35 (1.03)	0.00 - 7.38
EOTX	100	7.59 (0.38)	6.24 - 8.72	7.50 (0.49)	2.18 - 9.20	--	--	--	--	--
IL10	99	2.02 (1.10)	0.76 - 7.77	1.68 (0.92)	0.2 - 7.77	90	1.32 (0.81)	0.01 - 5.46	0.95 (0.70)	0.01 - 3.72
Acute Phase Reactant										
CRP	95	1.10 (0.76)	0.18 - 5.05	1.04 (0.72)	0.18 - 5.15	99	1.16 (0.79)	0.12 - 5.13	0.95 (0.61)	0.12 - 4.78

Notes:
a. All biomarker concentration units were pg/mL.

Table 3.3: Geometric Mean Ratios for Cytokines: Mean Levels and Slopes.

	BAFF	GMCSF	IFN-γ	IL-12p70	IL-1β	IL-2	TNF-α	IL-6
	MR (95%CI)	MR (95%CI)	MR (95%CI)	MR (95%CI)	MR (95%CI)	MR (95%CI)	MR (95%CI)	MR (95%CI)
HAART Naïve Visits								
Means								
Cases	1.02 (0.91, 1.15)	0.93 (0.82, 1.05)	0.97 (0.87, 1.07)	1.01 (0.81, 1.25)	0.97 (0.87, 1.07)	1.09 (1.00, 1.19)*	1.17 (1.07, 1.30)*	1.12 (1.01, 1.23)*
Controls (ref.)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Slopes								
Cases	1.10 (0.93, 1.31)	1.22 (1.00, 1.49)	1.12 (0.95, 1.31)	0.99 (0.72, 1.36)	1.01 (0.85, 1.21)	1.02 (0.89, 1.16)	1.02 (0.87, 1.19)	1.02 (0.87, 1.19)
Controls (ref.)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
HAART Exposed Visits								
Means								
Cases	1.41 (1.12, 1.78)*	0.94 (0.59, 1.49)	1.28 (0.87, 1.90)	0.97 (0.46, 2.05)	0.85 (0.51, 1.43)	1.00 (0.74, 1.35)	1.41 (1.02, 1.94)*	1.15 (0.83, 1.61)
Controls (ref.)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Slopes								
Cases	1.00 (0.80, 1.25)	1.06 (0.70, 1.60)	0.96 (0.68, 1.34)	1.05 (0.52, 2.12)	1.07 (0.69, 1.67)	1.03 (0.78, 1.35)	0.97 (0.73, 1.31)	0.95 (0.70, 1.29)
Controls (ref.)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Notes:

* Statistically significant at dependent FDR corrected p-value < 0.05.

§ Results from linear mixed models with each natural log transformed biomarker as outcome, covariates included age, race, BMI, CD4+ T cell counts, HCV infection, smoking, alcohol intake, stimulant use, and MACS site.

Table 3.4: Geometric Mean Ratios for Soluble Receptors: Mean Levels and Slopes

	sCD14	**sCD27**	**sGP130**	**sIL-2Rα**	**sIL-6R**	**sTNFR2**
	MR (95%CI)	MR (95%CI)	MR (95%CI)	MR (95%CI)	MR (95%CI)	MR (95%CI)
HAART Naïve Visits						
Means						
Cases	1.06 (0.97, 1.16)	1.11 (1.03, 1.21)*	0.94 (0.87, 1.02)	1.16 (1.02, 1.32)*	0.95 (0.87, 1.04)	1.03 (0.91, 1.17)
Controls (ref.)	1.00	1.00	1.00	1.00	1.00	1.00
Slopes						
Cases	1.01 (0.88, 1.15)	1.04 (0.92, 1.18)	1.05 (0.94, 1.18)	1.17 (0.97, 1.42)	1.14 (1.00, 1.29)*	1.25 (1.04, 1.51)*
Controls (ref.)	1.00	1.00	1.00	1.00	1.00	1.00
HAART Exposed Visits						
Means						
Cases	1.00 (0.83, 1.20)	1.05 (0.81, 1.36)	0.95 (0.81, 1.11)	1.47 (1.14, 1.91)*	1.15 (0.97, 1.35)	1.34 (1.03, 1.73)*
Controls (ref.)	1.00	1.00	1.00	1.00	1.00	1.00
Slopes						
Cases	1.19 (1.00, 1.43)	1.25 (0.98, 1.59)	1.10 (0.94, 1.28)	1.20 (0.94, 1.55)	1.07 (0.91, 1.27)	1.21 (0.94, 1.55)
Controls (ref.)	1.00	1.00	1.00	1.00	1.00	1.00

Notes:

* Statistically significant at dependent FDR corrected p-value < 0.05.
§ Results from linear mixed models with each natural log transformed biomarker as outcome, covariates included age, race. BMI, CD4+ T cell counts, HCV infection, smoking, alcohol intake, stimulant use, and MACS site.

Table 3.5: Geometric Mean Ratios for Chemokines: Mean Levels and Slopes

	IL-8	IP-10	MCP-1	MCP-4	MIP-1β	TARC	CXCL13	EOTX-1	IL-10
	MR (95%CI)	MR (95%CI)	MR (95%CI)	MR (95%CI)	MR (95%CI)	MR (95%CI)	MR (95%CI)	MR (95%CI)	MR (95%CI)
HAART Naïve Visits									
Means									
Cases	0.84 (0.64, 1.11)	1.23 (1.07, 1.42)*	1.08 (0.94, 1.23)	0.93 (0.80, 1.08)	1.07 (0.86, 1.32)	0.95 (0.76, 1.20)	1.30 (1.18, 1.43)**	1.05 (0.91, 1.21)	1.05 (0.87, 1.26)
Controls (ref.)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Slopes									
Cases	1.35 (0.90, 2.03)	1.10 (0.89, 1.35)	1.08 (0.89, 1.30)	1.23 (1.00, 1.51)	0.86 (0.64, 1.17)	1.03 (0.75, 1.40)	1.10 (0.95, 1.28)	1.02 (0.83, 1.24)	1.57 (1.19, 2.06)*
Controls (ref.)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
HAART Exposed Visits									
Means									
Cases	0.96 (0.54, 1.71)	1.91 (1.26, 2.92)*	1.31 (1.01, 1.70)*	1.22 (0.92, 1.61)	1.23 (0.81, 1.88)	1.29 (0.84, 1.97)	1.20 (0.88, 1.64)	0.80 (0.61, 1.05)	1.31 (0.74, 2.34)
Controls (ref.)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Slopes									
Cases	1.16 (0.68, 1.96)	1.00 (0.67, 1.49)	1.01 (0.78, 1.30)	0.93 (0.70, 1.22)	0.77 (0.51, 1.15)	0.79 (0.52, 1.20)	1.15 (0.86, 1.54)	1.22 (0.94, 1.60)	1.34 (0.78, 2.32)
Controls (ref.)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Notes:

* Statistically significant at dependent FDR corrected p-value < 0.05.
§ Results from linear mixed models with each natural log transformed biomarker as outcome, covariates included age, race. BMI, CD4+ T cell counts, HCV infection, smoking, alcohol intake, stimulant use, and MACS site.

Table 3.6: Geometric Mean Ratios for Acute Phase Reactant: Mean Levels and Slopes.

	CRP
	MR (95%CI)
HAART Naïve Visits	
Means	
Cases	1.14 (1.00, 1.29)*
Controls (ref.)	1.00
Slopes	
Cases	1.01 (0.82, 1.24)
Controls (ref.)	1.00
HAART Exposed Visits	
Means	
Cases	1.66 (1.08, 2.53)*
Controls (ref.)	1.00
Slopes	
Cases	0.79 (0.53, 1.17)
Controls (ref.)	1.00

Notes:

* Statistically significant at dependent FDR corrected p-value < 0.05.

§ Results from linear mixed models with each natural log transformed biomarker as outcome, covariates included age, race. BMI, CD4+ T cell counts, HCV infection, smoking, alcohol intake, stimulant use, and MACS site.

4 Chapter IV: Predictive Utility of Serum Biomarkers of Inflammation, Immune Activation, in AIDS-associated Lymphoma in HIV-infected Men in the MACS

4.1 Abstract

Background: Long term immune dysregulation has been documented to be etiologically linked to AIDS-associated non-Hodgkin lymphoma (AIDS-NHL). Prospectively measured biomarkers of inflammation and immune activation may thus aid in predicting the incidence of AIDS-NHL events in high risk population characterized by a chronic immune challenge. We investigated the predictive utility of 13 biomarkers in forecasting AIDS-NHL events in the Multicenter AIDS Cohort Study (MACS).

Methods: We prospectively ascertained the serum concentrations of 13 biomarkers of inflammation and immune activation from stored serum samples of MACS participants from a 1:1 matched case-control substudy (N=358). At 0-1, 1<-3, and >3 years prior to NHL diagnosis, we fit 13 separate random logistic regression models for the prediction of AIDS-NHL with each biomarker included in addition to a set of subject characteristics as covariates. We assessed the discrimination of the models by calculating area under receiver operating characteristic curve (AUC) values to assess the ability of the logistic regression models to discriminate between AIDS-NHL cases versus controls. We then fit another logistic regression model including multiple predictor biomarkers selected using the rankings of AUC statistics in individual biomarker models to select the best combination of biomarkers to include concurrently in a model. We also calibrated our models by graphing calibration plots and conducting goodness-of-fit tests that indicated the differences in predicted disease risks relative to observed disease proportions. Finally, we performed bootstrap validation to examine the internal validity of our prediction models.

Results: Models including individual biomarkers yielded modest improvements in AUC statistics above a base-case model that comprised NHL risk factors and other participant characteristics. A model including IL-6, TNF-α, IP10, CXCL13, and sCD27, concurrently performed better than all other models including individual biomarkers: 0-1: AUC=0.943 95% CI: 0.910, 0.975; 1-3: AUC=0.895 95% CI: 0.856, 0.934; >3: AUC=0.836 95% CI: 0.787, 0.885. Increments in AUC above the base-case model were also better than any of the models extended using individual biomarkers: 0-1: difference in AUC=0.056 95% CI: 0.021, 0.091; 1-3: difference in AUC=0.032 95% CI: 0.007, 0.057; >3: difference in AUC=0.074 95% CI: 0.030, 0.118.

Conclusion: Biomarkers of immune activation and inflammation provide modesty improved discrimination over other participant characteristics and risk factors. Combining biomarkers, IL-6, TNF-α, IP10, IL-10, CXCL13 in a single model provides the best discrimination and calibration over extended follow-up times. Risk prediction models with biomarkers may have future clinical utility in identifying patients at high risk for AIDS-NHL.

4.2 Introduction

Non-Hodgkin lymphoma (NHL) continues to present a significant public health challenge among HIV-infected individuals even with the wide-spread availability of multi-agent highly active anti-retroviral therapy (HAART) [7,42]. Immune dysfunction, particularly chronic immune activation, is consistently reported to be a risk factor for AIDS-NHL among HAART users [4,172]. Several serum biomarkers of immune activation and inflammation are documented to be altered prior to AIDS-NHL diagnosis including some we consider in this study: IL-6 [54,141], IL-10 [54,141,142], CCL19, CXCL10/IP-10 [14,141], CXCL11, CXCL13, sCD27, CRP [54,147,148], and TNF-α [14,141], MCP-2, MIP-1δ/MIP-5/CCL15, IFN-α, GM-CSF [14]. In a prior report we summarize evidence from the literature linking pre-diagnosis signatures of circulating markers of inflammation and immune activation the risk of NHL [156]. In another study, we compared the levels and slopes of temporal trajectories of 24 prospectively collected biomarkers among individuals who develop AIDS-NHL relative to those who do not.

In this study, we develop exploratory risk prediction models that include biomarkers previously described [54,128], in combination with a rich set of prospectively measured participant characteristics that serve as predictors. The rationale for this study is to explore the use of these biomarker measures in prediction models that provide easily interpretable prediction results with potential future clinical utility. While prior literature has shown associations between serum biomarkers and AIDS-NHL, there are no reports that develop models that could have direct utility as risk prediction tools for identifying patients of high risk and facilitate early detection and treatment of AIDS-NHL. We note the importance of the distinction between estimating measures of associations versus estimating measures of classification accuracy. It has been shown in the literature that a very strong association is a necessary but not sufficient condition for accurate classification, meaning that a high association between a covariate and the outcome under

study do not necessarily imply better classification of the outcome[173,174]. Risk prediction models are statistical models developed using several risk factors or patient characteristics thought to be associated with a health outcome of interest; they are used to predict the probability that an individual with a given set of risk factors or characteristics will experience the outcome of interest. The health outcome of interest could be the presence of disease (diagnosis) or the future progression of a disease (prognosis) [175–178]. Examples of such risk prediction tools for several other disease outcomes using serum immunologic factors include the use of serum FLCs and immunoglobulins in the prediction of multiple myeloma [179–181], serum biomarkers and ischemic stroke [182], and multiple circulating immune factors in prediction models for resected NSCLC [183].

This study aims to evaluate whether our set of 13 biomarkers improve the *discrimination* between AIDS-NHL cases versus controls relative to a reference model comprising other individual subject characteristics or known risk factors for NHL, while exhibiting adequate *calibration* to the data. Discrimination quantifies how well a risk prediction model distinguishes between subjects who will develop disease outcomes of interest from those who will not experience the events [184,185]. Various measures have been proposed to capture discrimination. However, the most widespread measure of discrimination accuracy of prediction models is the use of AUC [184,186]. AUC is the area under the receiver operating characteristic curve. It measures the degree to which a risk prediction model can discriminate between those who do and do not have a disease of interest. Since the ROC curve plots sensitivity versus 1−specificity across a range of cut points for a continuous predictor, higher AUC signifies higher discriminative ability of a risk prediction model [187]. Calibration is an attribute of a risk calculator -- in our case, a logistic regression model. A risk calculator is said to be well calibrated when the frequency of disease outcomes among subjects with predictor values $X = x$ is equal to the calculator's measure of the probability (risk) of occurrence of that disease in the subject

population [178]. More loosely, it is a quantity that measures how close predicted outcomes from a model are to observed outcomes in the data set. Calibration is evaluated by graphical methods such as calibration slope plots or goodness-of-fit tests, and involves comparing the observed event rates versus average estimated risk values within subgroups of subjects defined by combinations of values of predictors included in the risk prediction model [178].

4.3 Materials & Methods

Details of the MACS have been described and published previously [39,40]. We include data from a case-control substudy nested within the MACS [54]. In this substudy, serum samples were extracted from MACS participants enrolled between 1984–85 or 1987–1991. Cases were defined as HIV-infected MACS participants diagnosed with AIDS-NHL before April, 2003, with at least one serum sample at a visit prior to AIDS-NHL diagnosis (n=179). For each AIDS-NHL case, an HIV-infected control was selected from among HIV-infected MACS subjects with no lymphoma diagnosis as of April 2003. Cases were matched to controls (1:1 ratio for a total sample size of n=358) on the duration of HIV-infection if the date of HIV-seroconversion was known, or date of entry (within a year) into the MACS cohort if they were HIV-seroprevalent, and also matched on expected availability of their serum samples at time points comparable to cases within a year. Serum samples were collected from three possible time intervals prior to NHL diagnosis (prediagnosis time point) at 0–1 year (nearest visit to 0.5), 1–3 years (nearest visit to 2), and >3 years (nearest visit to 4) prior to AIDS-NHL, with each participant contributing at most one sample per interval.

4.4 AIDS-NHL Outcome, Biomarkers, and Covariates

AIDS-NHL diagnoses were ascertained at each follow-up visit through interviews of subjects and medical chart abstraction, pathology reports from autopsy, and finally queries of state

cancer registries [11]. Since 2006, all diagnosed cancers were classified by site and histology codes of the International Classification of Diseases for Oncology, 3rd edition (ICD-O-3). Our primary outcome was AIDS-NHL diagnosis. Biomarkers were our primary predictors of interest, and details of assays to measure their concentration have been described in detail elsewhere.[128] We considered a comprehensive set of covariates and their corresponding functional forms using prior literature [3,4,36,188] and data availability. We included time-dependent variables CD4+ T cell counts at each visit as a continuous variable; self-reported tobacco use, self-reported stimulant drug use (define as either cocaine or methamphetamine use); self-reported frequency of alcohol use (captured as no use, less than twice weekly, and more than twice weekly since the last visit); body mass index (BMI, categorized as above normal, BMI>25 kg/m^2, versus normal or lower at each visit); HCV exposure status at each current visit (no exposure versus any exposure); uncontrolled diabetes mellitus at each visit based on classification of A1c>6.5% or fasting glucose >126mg/dL (7.0 mmol/l) [128,149]; and estimated duration of HIV-infection. Time-invariant confounders included age at NHL diagnosis, race categorized as non-Hispanic white versus non-white, and MACS center.

4.5 Statistical Methods

We followed the approach described by Steyerberg [175,176,189,190] for documenting the steps involved in developing the prediction models, including data inspection and coding, model specification and estimation, and finally assessment of model performance and validation.

Data Inspection: We tabulated descriptive statistics for variables in our study sample and explored the distributions of the predictors included in our models at each prediagnosis time point of interest. We also presented summary statistics for the distribution of the biomarker values and associated levels of missingness.

Model Specification and Estimation: At each of 3 prediagnosis time points (0-1, 1<-3, and >3 years prior to NHL diagnosis) we fit logistic regression models using maximum likelihood estimation. For the base-case regression model, we included a set of subject characteristics and risk factors for NHL, categorized as described in section 4.4, as covariates namely, age at AIDS-NHL diagnosis (age at matched time point for controls), race, MACS center, and at the specific AIDS-NHL prediagnosis time point of prediction: duration of HIV-infection, CD4+ T cell counts, smoking status, stimulant drug use, alcohol intake, BMI, and HCV infection exposure. We then estimated logistic regression models extending the base-case model with each extended model additionally including one of 13 biomarkers. A final extended regression model included multiple predictor biomarkers selected by assessing the top 5 ranking biomarkers as measured by improvement in AUC. [191,192]

Assessment of Model Performance and Validation: In assessing the performance of our prediction models we were interested in two aspects of predictive accuracy: (1) discrimination, and (2) calibration. We tabulated AUCs a measure of discrimination for each of our models and estimate the difference between the AUC of the base-case model versus the AUC of the extended model including the biomarker. It has been shown that for models containing known risk factors exhibiting good discrimination, very large associations of biomarkers with the outcome are required in order to observe even a modest improvement in AUC. We therefore take a descriptive approach and rank the improvement in AUC between markers. We assessed calibration by conducting goodness of fit tests using the Hosmer-Lemeshow test,[193] which assesses whether observed proportions of AIDS-NHL are concordant with predicted AIDS-NHL risk in subgroups defined by the deciles of the predicted risk values. In addition, we cross-check the Hosmer-Lemeshow goodness of fit tests with a model based version that is shown to have better power to detect departures from concordance in small samples.[194] We used locally

estimated scatterplot smoothing (loess) to graph calibration plots assessing the degree to which our estimated predicted probabilities of AIDS-NHL deviated from smoothed observed proportions of AIDS-NHL cases.[195] For each plot, a 'span' was defined around each subject such that all subjects within that window were used to fit a local regression used to obtain a prediction of observed risk for each subject which was then compared to the risk predicted by the logistic regression model estimated with risk factors and biomarkers included together as predictors.

Finally, we assessed the internal validity of our models via correction of statistical 'optimism' via bootstrap techniques as described in Steyeberg.[177,196] Optimism is a measure of data overfitting when calculating performance statistics such as the AUC. The following steps were used in calculating and correcting for optimism[196]:

1. Logistic regressions were fit to the original data set (n=358). AUC statistics render "apparent" performance.
2. Bootstrap samples (B=100) were drawn with replacement from the original sample, each n=358.
3. Logistic regressions, specified with the same variables as in step 1, were fit to each of the bootstrap samples. These AUC statistics measure "bootstrap" performance.
4. The regression parameters from step 3 were used to calculate AUC statistics using the original data used in step 1. These AUC statistics measure "test" performance.
5. Optimism is calculated as the difference between bootstrap performance and test performance for a total of B=100 optimism estimates.
6. B=100 optimism corrected estimates are obtained by subtracting each AUC from step 1 with corresponding optimism estimates in step 5. A final optimism estimated estimate is presented as the median of the optimism corrected estimates. Uncertainty intervals for the final optimism corrected estimates are presented as the 2.5th and 97.5th quantiles of the distribution of the optimism corrected estimates.

We also include leave-one-out cross validation AUC statistics as an alternative correction for optimism.

All analyses were performed using SAS procedures in SAS 9.2 Cary, NC.

4.6 Results

Descriptive Statistics

We describe participant characteristics of our analytic sample (n=358) in Table 4.1 by prediagnosis time point. We observed similar distributions of age and race between cases and controls at each time-point. Controls exhibited higher prevalences of overweight and obese BMIs relative to controls across time points, while CD4 cell counts were consistently depressed among cases relative to controls with the trend intensifying nearing the time of NHL diagnosis. A similar trend is observed in behavioral risk factors: smoking, heavy drinking, and stimulant use are lower among cases relative to controls, and the trend becomes more pronounced as clinical diagnosis of NHL approaches.

We present biomarker median levels and interquartile range boundaries by case status and prediagnosis time point in Table 4.2. We consistently observed higher median concentrations among cases relative to controls across all time-points, with the only exception being IL-12p70 where there appear to be no meaningful differences by case status.

Model Predictive Performance

For each biomarker, we present AUC statistics, and corresponding Wald 95% confidence intervals, assessing the predictive performance of the model, and the increment in AUC over that of the base-case model including only participant characteristics and NHL risk factors (Table 4.3). We find that in general, the models including the biomarkers improve the AUC statistics only marginally (<10%) in absolute terms, frequently with a U-shaped trend in the AUC difference statistics from the furthest time-point (3+ years prior to NHL), to the prediagnosis time point during the year of diagnosis of AIDS-NHL. The models that had the top 5 incremental predictive performance averaged across time points were IL-6 (0-1 years: difference in AUC= 0.031 95% CI: 0.002, 0.061; 1-3 years: difference in AUC=0.006 95% CI: -0.005, 0.016; >3 years: difference in AUC= 0.043 95% CI: 0.013, 0.074), TNF-α (0-1 years: difference in AUC= 0.020 95% CI: -0.006, 0.046; 1-3 years: difference in AUC= 0.025 95% CI: 0.004, 0.046; >3 years: difference in AUC= 0.018 95% CI: -0.008, 0.043), IP10 (0-1 years: difference in AUC= 0.020 95% CI: -0.001, 0.041; 1-3 years: difference in AUC=0.014 95% CI: -0.002, 0.030; >3 years: difference in AUC= 0.022 95% CI: -0.006, 0.050), IL-10 (0-1 years: difference in AUC= 0.013 95% CI: -0.006, 0.031; 1-3 years: difference in AUC=0.012 95% CI: -0.003, 0.028; >3 years: difference in AUC= 0.022 95% CI: -0.003, 0.046), CXCL13 (0-1 years: difference in AUC= 0.000 95% CI: -0.008, 0.008; 1-3 years: difference in AUC= 0.017 95% CI: 0.002, 0.032; >3 years: difference in AUC= 0.028 95% CI: -0.002, 0.054) (Table 4.3). A model including these biomarkers concurrently performed better, considering average AUC across time points, than all other models that included only individual biomarkers: 0-1: AUC=0.943, 95% CI: 0.910, 0.975; 1-3: AUC=0.895, 95% CI: 0.856, 0.934; >3: AUC=0.836, 95% CI: 0.787, 0.885 (Table 4.3). The increments in AUC above the base-case model were also better than any of the models extended using individual biomarkers: 0-1: difference in AUC=0.056 95% CI: 0.021, 0.091; 1-3: difference in AUC=0.032 95% CI: 0.007, 0.057; >3: difference in AUC=0.074 95% CI: 0.030, 0.118 (Table 4.3).

We present our assessments of model calibration using goodness of fit tests (Table 4.5), and calibration plots (Figure 4.1). Goodness of fit tests indicate that the models including biomarkers are generally well calibrated consistently across time-points (all p-values > 0.05, indicating adequate calibration). Conversely, the calibration plots indicate that while the overall levels of calibration may be adequate, for some biomarkers (CRP, CXCL13, GM-CSF, INF-γ, IP12p70, IL1B,IL-2, IL-6, IP10), there is some variability in calibration for the 1-3 year time point, and to a lesser extent the 3+ years time-point, when compared to the 0-1 year time-point. This result indicates we have the best concordance between predicted and observed NHL risks in the time intervals closest to clinical diagnosis of AIDS-NHL. The model comprising the combined biomarkers exhibits good calibration overall, although the calibration level varies considerably for the >3 years time point in an erratic manner, indicating relatively poor concordance between predicted and observed risks, particularly when we consider the calibration levels for the time-points nearer to NHL diagnosis.

Model Validation

We corrected our AUC estimates in our observed data set and present the updated values (Table 4.5). We find that the difference between the 'apparent' AUC estimates and those corrected for statistical optimism are all within 5% of each other, which suggests that our estimates of AUC are likely to have reasonable applicability to new data sets.

4.7 Discussion

There are two trends that become apparent from our analyses. Firstly, that while biomarker models provide improved discrimination over other participant characteristics and risk factors, they do so only modestly, and the combination of biomarkers (IL-6, IL-10, TNF-α, IP10,

CXCL13) provides the best discrimination. Secondly, the improvement in discrimination performance due to the addition of biomarkers as predictors tends to be exhibit a "U-shape", with greatest performance at the earliest prediagnosis time point, and then another spike in performance in the year of AIDS-NHL diagnosis. These findings have scientific import because biomarker-disease association does not always translate to measurable disease prediction performance.

We considered a rich set of cytokines as potential predictors in our prediction models including GMCSF, IFN-γ, IL12p70, IL1β, IL-2, TNF-α, and IL-6. Among these cytokines, logistic regression models including IL-6 and TNF-α had the best overall discriminatory capability across the prediagnosis time points, with discrimination improving with the approach of AIDS-NHL diagnosis. IL-6 is a pluripotent cytokine that stimulates B cell proliferation and differentiation and promotes tumor growth [84,85] both of which lead to increased risk of NHL. In addition, IL-6 is associated with pro-inflammatory and Th17 immune responses both of which play a role in autoimmunity [86,87], which in turn is related to risks for NHL[88]. TNF-α is a pro-inflammatory cytokine that plays a role in B cell activation, growth, differentiation, apoptosis, and chemotaxis [102–104], all of which enhance B-cell hyperactivation associated lymphomagenic processes. This is particularly the case in the context of HIV-infection, where even HIV proteins imitate the TNF-α signaling pathway thereby increasing the viral reservoir [197], and potentially further encouraging a lymphomagenic environment marked by chronic immune activation and inflammation. While the absolute values of the AUC statistics were modest, both these cytokines do add some value to prediction models for AIDS-NHL, and would be useful to further evaluate in a larger study.

We also considered a set of chemokines (IL8, IP10, CXCL13, IL-10) as potential predictors of AIDS-NHL among which IP10, IL-10, and CXCL13 had the best performance. The chemokine

IP-10 (CXCL10) acts upon activated T cells and macrophages, and has also be shown to stimulate HIV replication in monocyte-derived macrophages and lymphocytes, and reduces the amount of MIP-1α, MIP-1β, and RANTES produced by activated peripheral blood lymphocytes [167]. With these considerations, IP-10 can be linked to B-cell lymphomagenesis through the chronic immune activation engendered by HIV-infection. Consistent with this prior evidence, we find that IP10 levels can help discriminate between AIDS-NHL cases and controls over and above known risk factors. IL-10, a pleiotropic cytokine with stimulatory effects on B cells, promotes chronic B cell activation and is likely etiologically linked to NHL[92–94]. We found that circulating IL-10 levels provide some improvement in predictive ability of our models in accordance with its purported etiological role for NHL. We also considered CXCL13 for which aberrant expression is thought to be involved in the pathogenesis of B cell lymphoma through abnormal chemotaxis of B cells to tissues or abnormal B cell activation.[109] Overexpression of the receptor-ligand pair CXCR5/CXCL13 has been observed in B cell chronic lymphocytic leukemia,[110] and follicular lymphoma cells have been seen to secrete CXCL13,[111] indicating the possibility that CXCL13 can be a biomarker for NHL and these subtypes. We observed that levels of CXCL13 improve the discrimination between AIDS-NHL cases and controls indicating that CXCL13 could have some utility as a biomarker for AIDS-NHL.

Models that included a combination of the biomarkers (IL-6, TNF-α, IP10, IL-10, CXCL13) that conferred the best improvements in AUC above the base-case model, outperformed all the models including individual biomarkers. This finding indicates that a larger study is warranted to investigate which biomarkers among our included set, or some larger set of biomarkers, might provide the best predictive performance in predicting AIDS-NHL. In addition, more data would allow incorporation of more complex predictive modeling strategies than we could employ here to assess the precise combinations and functional forms of biomarkers that could yield the best clinical prediction. It would also be ideal to have access to an additional independent data set

that could be used for external validation of these or similar prediction models. Such validation would bring results from prediction model building closer to translation into clinical practice.

There are several data related limitation in our study. First, we were using data from a matched nested case-control design, with matching factors being duration of HIV-infection (or date of entry (within a year) for HIV-seroprevalent), and expected availability of serum samples. We used duration of HIV-infection in our prediction models in logistic regression models that did not condition on matched risk set, but did not have the data assessing expected availability of serum samples, therefore could not include it leading to potential introduction of bias in our estimates. In addition, 'breaking the match' may also introduce biased estimates of risk prediction measures in a case-control.[198] Secondly, the number of participants included at each time point were not equivalent across time intervals meaning that the comparisons of the prediction measures were compromised by differences in amounts of information used for each estimate.

Our study possesses several strengths. First, we were able to prospectively assess the AIDS-NHL event history of a group of men over a substantial follow up period which reduces biases due to the independent collection of biomarker exposures and outcomes, and also allowed us to comprehensively assess the predictive performance of the biomarkers over a significant time-frame. Centralized labs were utilized in specimen extraction and quantitation meaning that random variability in the measures was reduced, and the high quality and sensitivity of the assay instrumentation meant that we were able to receive measures with the highest possible detection capability. We also included a comprehensive set of risk factors and patient characteristics that have either been shown to be associated with NHL risk or could plausibly have predictive value.

In conclusion, our prospective investigation into 13 markers of immune activation, and inflammation provided some evidence that elevated levels of several biomarkers could be used to improve models that predict AIDS-NHL incidence in a high-risk population for up to more than three years prior to diagnosis. These biomarkers could be considered in larger studies that further develop fine-tuned risk prediction models for AIDS-NHL.

4.8 Acknowledgements

Funding sources

This research was supported by the National Institutes of Health grant T32 CA09142, the Center for HIV Identification, Prevention, and Treatment (CHIPTS) NIMH grant P30MH058107; UCLA Center for AIDS Research (CFAR) grant 5P30AI028697, Core H, and the American Recovery and Reinvestment Act. The ideas and opinions expressed herein are those of the authors, are not in any manner endorsed by the National Institutes of Health, the University of California at Los Angeles or their Contractors and Subcontractors.

4.9 Supporting Information

No additional Supporting Information

4.10 Disclosures

The authors of the manuscript have no conflicts of interest to disclose.

4.11 Tables and Figures

Table 4.1: Participant Characteristics in Analytic Sample: N=358.

		Year Prior to NHL Diagnosis					
		0-1		**1-3**		**>3**	
Variable	**Category**	**Control**	**Case**	**Control**	**Case**	**Control**	**Case**
N	Total	107	106	149	151	147	150
Age	Median (Q1-Q3)	38 (34-43)	41 (36-45)	37 (33-41)	39 (34-44)	36 (32-40)	37 (32-42)
Race	White, non-Hispanic	92 (86)	93 (88)	131 (88)	130 (86)	126 (86)	123 (82)
	Hispanic	8 (7)	10 (9)	9 (6)	14 (9)	10 (7)	19 (13)
	Black, non-Hispanic	4 (4)	3 (3)	6 (4)	7 (5)	9 (6)	8 (5)
	Other race	3 (3)	0 (0)	3 (2)	0 (0)	2 (1)	0 (0)
BMI (kg/m3)	Normal/Underweight (<25.0)	59 (58)	83 (84)	88 (61)	109 (76)	85 (61)	102 (71)
	Overweight/Obese (25.0+)	43 (42)	16 (16)	56 (39)	34 (24)	54 (39)	41 (29)
	Missing	5	7	5	8	8	7
NHL Histological Subtype	Burkitt lymphoma/Burkitt cell leukemia	--	12 (12)	--	17 (12)	--	14 (10)
	Diffuse large B-cell lymphoma	--	62 (60)	--	85 (58)	--	84 (57)
	Follicular lymphoma	--	2 (2)	--	2 (1)	--	2 (1)
	Lymphoplasmacytic lymphoma	--	2 (2)	--	2 (1)	--	1 (1)
	Malignant lymphoma, mixed small and large cell, diffuse	--	1 (1)	--	1 (1)	--	1 (1)

		Year Prior to NHL Diagnosis					
		0-1		1-3		>3	
Variable	**Category**	**Control**	**Case**	**Control**	**Case**	**Control**	**Case**
	Malignant lymphoma, non-Hodgkin, NOS	--	24 (23)	--	40 (27)	--	45 (31)
	Unknown	--	3	--	4	--	3
CD4 cell count/mm3	Median (Q1-Q3)	468 (280-629)	79 (23-258)	513 (349-688)	210 (81-356)	562 (437-759)	385 (243-550)
CD4 cell category, count/mm3	<400	46 (43)	90 (87)	47 (32)	118 (81)	31 (21)	79 (54)
	400-600	25 (24)	10 (10)	41 (28)	17 (12)	55 (37)	37 (25)
	>600	35 (33)	3 (3)	57 (39)	11 (8)	61 (41)	30 (21)
	Missing	1	3	4	5	0	4
Duration of HIV Infection (years)	Median (Q1-Q3)	5 (3-9)	6 (4-9)	4 (2-7)	4 (2-7)	3 (1-5)	3 (1-5)
Year of HIV Infection	1984-1985	86 (80)	87 (82)	121 (81)	124 (82)	122 (83)	124 (83)
	1986-1987	8 (7)	10 (9)	10 (7)	13 (9)	12 (8)	14 (9)
	1988-1989	8 (7)	3 (3)	11 (7)	5 (3)	8 (5)	5 (3)
	1990-1991	3 (3)	4 (4)	5 (3)	7 (5)	3 (2)	5 (3)
	1992-1994	2 (2)	2 (2)	2 (1)	2 (1)	2 (1)	2 (1)
Hepatitis C Infection	HCV Unexposed	96 (90)	89 (84)	135 (92)	127 (85)	137 (93)	130 (87)
	HCV Exposed	11 (10)	17 (16)	12 (8)	22 (15)	10 (7)	19 (13)
	Missing			2	2	0	1
Smoking Status	Smoker	41 (39)	20 (19)	51 (36)	34 (23)	49 (36)	45 (31)
	Non-smoker	65 (61)	83 (81)	91 (64)	112 (77)	88 (64)	98 (69)
	Missing	1	3	7	5	10	7
No. Alcohol Drinks Per Week	None	20 (19)	26 (25)	15 (10)	22 (15)	12 (9)	18 (13)
	1 to 3 drinks/week	43 (40)	52 (51)	55 (38)	75 (51)	56 (40)	57 (40)

Variable	Category	Year Prior to NHL Diagnosis					
		0-1		1-3		>3	
		Control	Case	Control	Case	Control	Case
	4 to 13 drinks/week	33 (31)	21 (21)	56 (39)	41 (28)	57 (41)	52 (36)
	More than 13 drinks/week	11 (10)	3 (3)	18 (13)	8 (5)	14 (10)	16 (11)
	Missing	0	4	5	5	8	7
Uncontrolled Diabetes Mellitus	No	106 (99)	106 (100)	147 (100)	149 (100)	147 (100)	149 (100)
	Yes	1 (1)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
	Missing	0	0	2	2	0	1
Stimulant Use	No	83 (78)	92 (90)	108 (75)	123 (85)	92 (67)	111 (77)
	Yes	23 (22)	10 (10)	36 (25)	21 (15)	46 (33)	34 (23)
	Missing	1	4	5	7	9	5
MACS Site	Baltimore	18 (17)	28 (26)	29 (19)	45 (30)	32 (22)	39 (26)
	Chicago	16 (15)	16 (15)	25 (17)	31 (21)	25 (17)	38 (25)
	Los Angeles	46 (43)	45 (42)	62 (42)	58 (38)	62 (42)	58 (39)
	Pittsburgh	27 (25)	17 (16)	33 (22)	17 (11)	28 (19)	15 (10)

Table 4.2: Biomarker Distributions by Prediagnosis Time Point.

	Year Prior to NHL Diagnosis					
	0-1		1-3		>3	
Variable	**Control**	**Case**	**Control**	**Case**	**Control**	**Case**
N	107	106	149	151	147	150
Biomarkers	**Median (Q1-Q3)**					
GMCSF	0.17 (0.01-0.32)	0.32 (0.17-0.52)	0.07 (0.01-0.32)	0.26 (0.01-0.40)	0.11 (0.01-0.33)	0.17 (0.01-0.35)
IFN-γ	0.21 (0.12-0.31)	0.43 (0.25-0.76)	0.21 (0.10-0.33)	0.34 (0.19-0.57)	0.21 (0.10-0.31)	0.25 (0.11-0.42)
IL12p70	0.03 (0.03-0.03)	0.03 (0.03-0.03)	0.03 (0.03-0.10)	0.03 (0.03-0.03)	0.03 (0.03-0.03)	0.03 (0.03-0.03)
IL1β	0.17 (0.01-0.29)	0.24 (0.03-0.47)	0.21 (0.06-0.34)	0.21 (0.03-0.39)	0.26 (0.08-0.41)	0.21 (0.06-0.43)
IL2	0.00 (0.00-0.46)	0.55 (0.00-0.83)	0.00 (0.00-0.46)	0.46 (0.00-0.81)	0.00 (0.00-0.33)	0.05 (0.00-0.53)
TNF-α	2.09 (1.67-2.44)	2.55 (2.22-2.86)	2.17 (1.74-2.46)	2.46 (2.14-2.81)	2.11 (1.74-2.41)	2.38 (2.02-2.64)
IL6	0.58 (0.41-0.83)	0.96 (0.78-1.39)	0.58 (0.41-0.73)	0.77 (0.63-1.05)	0.51 (0.35-0.72)	0.68 (0.48-0.89)
sCD27	6.00 (5.81-6.29)	6.22 (5.90-6.58)	6.03 (5.77-6.22)	6.19 (5.93-6.53)	5.91 (5.73-6.13)	6.12 (5.89-6.39)
IL8	2.62 (2.22-3.04)	3.05 (2.60-3.61)	2.48 (2.14-2.93)	2.79 (2.36-3.21)	2.53 (2.15-2.97)	2.63 (2.23-3.06)
IP10	7.61 (7.15-8.06)	8.22 (7.70-8.76)	7.52 (6.91-8.00)	8.11 (7.68-8.54)	7.28 (6.88-7.79)	7.72 (7.37-8.16)
CXCL13	4.39 (3.97-4.74)	5.04 (4.60-5.60)	4.37 (3.88-4.84)	5.01 (4.57-5.42)	4.31 (3.86-4.77)	4.74 (4.39-5.24)
IL10	0.73 (0.49-1.04)	1.45 (0.98-2.05)	0.64 (0.43-0.95)	1.17 (0.81-1.72)	0.58 (0.38-1.04)	1.01 (0.58-1.43)
CRP	0.84 (0.57-1.18)	1.19 (0.67-1.69)	0.78 (0.46-1.26)	1.02 (0.60-1.61)	0.64 (0.41-1.11)	0.89 (0.48-1.34)

Notes:

a. Biomarkers measured on a natural log scale, with units of pg/mL.

Table 4.3: Comparing AUC Estimates for Models including Risk Factors Only, versus Biomarker with Risk Factors.

Biomarker	Prediagnosis Period (Years)	Model AUC§ Risk Factors Only	Biomarker + Risk Factors‡	Difference
sCD27	0-1	0.896 (0.852, 0.940)	0.900 (0.857, 0.943)	0.004 (-0.011, 0.020)
	1-3	0.863 (0.820, 0.905)	0.878 (0.838, 0.917)	0.015 (0.000, 0.030)
	>3	0.771 (0.715, 0.826)	0.793 (0.740, 0.847)	0.023 (-0.003, 0.048)
CRP	0-1	0.896 (0.852, 0.940)	0.900 (0.856, 0.943)	0.004 (-0.006, 0.013)
	1-3	0.863 (0.820, 0.905)	0.867 (0.826, 0.909)	0.005 (-0.004, 0.014)
	>3	0.771 (0.715, 0.826)	0.800 (0.748, 0.852)	0.030 (0.000, 0.059)
CXCL13	0-1	0.897 (0.853, 0.941)	0.897 (0.853, 0.940)	-0.000 (-0.008, 0.008)
	1-3	0.864 (0.822, 0.906)	0.881 (0.842, 0.920)	0.017 (0.002, 0.032)
	>3	0.773 (0.718, 0.828)	0.801 (0.749, 0.852)	0.028 (0.002, 0.054)
GM-CSF	0-1	0.886 (0.837, 0.934)	0.895 (0.848, 0.941)	0.009 (-0.007, 0.026)
	1-3	0.865 (0.821, 0.908)	0.866 (0.823, 0.909)	0.001 (-0.005, 0.008)
	>3	0.768 (0.710, 0.825)	0.772 (0.715, 0.829)	0.005 (-0.005, 0.014)
IFN-γ	0-1	0.886 (0.837, 0.934)	0.901 (0.856, 0.946)	0.016 (-0.006, 0.038)
	1-3	0.865 (0.821, 0.908)	0.867 (0.824, 0.910)	0.002 (-0.006, 0.010)
	>3	0.768 (0.710, 0.825)	0.783 (0.727, 0.838)	0.015 (-0.006, 0.036)
IL10	0-1	0.886 (0.837, 0.934)	0.898 (0.854, 0.943)	0.013 (-0.006, 0.031)
	1-3	0.865 (0.821, 0.908)	0.877 (0.836, 0.918)	0.012 (-0.003, 0.028)
	>3	0.768 (0.710, 0.825)	0.789 (0.734, 0.844)	0.022 (-0.003, 0.046)
IL12P70	0-1	0.886 (0.837, 0.934)	0.886 (0.837, 0.935)	0.000 (-0.007, 0.008)
	1-3	0.865 (0.821, 0.908)	0.870 (0.827, 0.913)	0.006 (-0.005, 0.016)
	>3	0.768 (0.710, 0.825)	0.777 (0.721, 0.834)	0.010 (-0.006, 0.026)
IL1β	0-1	0.886 (0.837, 0.934)	0.899 (0.854, 0.944)	0.014 (-0.007, 0.034)
	1-3	0.865 (0.821, 0.908)	0.866 (0.822, 0.909)	0.001 (-0.004, 0.006)
	>3	0.768 (0.710, 0.825)	0.772 (0.715, 0.829)	0.005 (-0.005, 0.014)
IL2	0-1	0.886 (0.837, 0.934)	0.894 (0.847, 0.940)	0.008 (-0.008, 0.025)
	1-3	0.865 (0.821, 0.908)	0.870 (0.827, 0.912)	0.005 (-0.005, 0.015)

	>3	0.768 (0.710, 0.825)	0.776 (0.719, 0.832)	0.008 (-0.006, 0.023)
IL6	0-1	0.888 (0.840, 0.936)	0.919 (0.880, 0.958)	0.031 (0.002, 0.061)
	1-3	0.860 (0.817, 0.904)	0.866 (0.823, 0.909)	0.006 (-0.005, 0.016)
	>3	0.770 (0.714, 0.827)	0.814 (0.762, 0.866)	0.043 (0.013, 0.074)
IL8	0-1	0.886 (0.837, 0.934)	0.896 (0.850, 0.942)	0.011 (-0.008, 0.030)
	1-3	0.865 (0.821, 0.908)	0.868 (0.825, 0.912)	0.004 (-0.005, 0.013)
	>3	0.768 (0.710, 0.825)	0.771 (0.714, 0.828)	0.003 (-0.006, 0.013)
IP10	0-1	0.888 (0.839, 0.937)	0.909 (0.865, 0.952)	0.020 (-0.001, 0.041)
	1-3	0.865 (0.820, 0.909)	0.879 (0.837, 0.920)	0.014 (-0.002, 0.030)
	>3	0.764 (0.706, 0.823)	0.786 (0.730, 0.842)	0.022 (-0.006, 0.050)
TNF-α	0-1	0.886 (0.837, 0.934)	0.905 (0.863, 0.948)	0.020 (-0.006, 0.046)
	1-3	0.865 (0.821, 0.908)	0.890 (0.851, 0.929)	0.025 (0.004, 0.046)
	>3	0.768 (0.710, 0.825)	0.785 (0.730, 0.841)	0.018 (-0.008, 0.043)
Multiple†	0-1	0.887 (0.837, 0.937)	0.943 (0.910, 0.975)	0.056 (0.021, 0.091)
	1-3	0.863 (0.819, 0.908)	0.895 (0.856, 0.934)	0.032 (0.007, 0.057)
	>3	0.762 (0.704, 0.821)	0.836 (0.787, 0.885)	0.074 (0.030, 0.118)

Notes:

§ AUC is the area under the receiver operating characteristic curve measuring discrimination.

‡ These models include all the variables in the risk factors only model namely duration of HIV-infection, age at diagnosis (or at matched time point), CD4+ T-cell count, stimulant use (cocaine use, and/or upper use), tobacco smoking, alcohol consumption frequency, BMI, HCV status, and MACS center.

†The model comprising multiple markers concurrently, in addition to risk factors, included IL-6, IL-10, TNF-α, IP10, CXCL13.

Table 4.4: Optimism Correction of AUC Estimates for Models including Biomarker with Risk Factors

Biomarker	Prediagnosis Period (Years)	AUC				
		Apparent Performance§	Bootstrap‡	Test Performance†	Optimism CorrectedΨ	Cross ValidatedΨ
sCD27	0-1	0.900 (0.857, 0.943)	0.926 (0.884, 0.959)	0.875 (0.847, 0.891)	0.849 (0.819, 0.874)	0.849 (0.794, 0.904)
	1-3	0.878 (0.838, 0.917)	0.894 (0.857, 0.934)	0.847 (0.829, 0.864)	0.831 (0.810, 0.848)	0.837 (0.791, 0.884)
	>3	0.793 (0.740, 0.847)	0.823 (0.778, 0.877)	0.748 (0.721, 0.768)	0.718 (0.684, 0.738)	0.732 (0.672, 0.792)
CRP	0-1	0.900 (0.856, 0.943)	0.926 (0.883, 0.969)	0.874 (0.841, 0.888)	0.847 (0.814, 0.862)	0.847 (0.792, 0.903)
	1-3	0.867 (0.826, 0.909)	0.883 (0.843, 0.918)	0.842 (0.822, 0.853)	0.826 (0.804, 0.844)	0.826 (0.778, 0.874)
	>3	0.800 (0.748, 0.852)	0.831 (0.784, 0.870)	0.749 (0.717, 0.777)	0.719 (0.682, 0.760)	0.744 (0.686, 0.802)
CXCL13	0-1	0.897 (0.853, 0.940)	0.922 (0.881, 0.957)	0.871 (0.843, 0.884)	0.845 (0.816, 0.867)	0.846 (0.791, 0.901)
	1-3	0.881 (0.842, 0.920)	0.900 (0.860, 0.929)	0.848 (0.827, 0.871)	0.829 (0.809, 0.863)	0.842 (0.795, 0.888)
	>3	0.801 (0.749, 0.852)	0.835 (0.785, 0.880)	0.750 (0.721, 0.778)	0.716 (0.686, 0.750)	0.744 (0.686, 0.802)
GM-CSF	0-1	0.895 (0.848, 0.941)	0.925 (0.879, 0.960)	0.868 (0.841, 0.883)	0.838 (0.810, 0.864)	0.832 (0.772, 0.893)
	1-3	0.866 (0.823, 0.909)	0.886 (0.856, 0.921)	0.844 (0.825, 0.856)	0.825 (0.792, 0.844)	0.823 (0.772, 0.874)
	>3	0.772 (0.715, 0.829)	0.805 (0.741, 0.855)	0.734 (0.708, 0.752)	0.701 (0.683, 0.725)	0.701 (0.637, 0.765)
IFN-γ	0-1	0.901 (0.856, 0.946)	0.928 (0.876, 0.962)	0.871 (0.845, 0.884)	0.844 (0.826, 0.869)	0.840 (0.781, 0.899)
	1-3	0.867 (0.824, 0.910)	0.886 (0.839, 0.925)	0.845 (0.830, 0.858)	0.826 (0.814, 0.842)	0.825 (0.774, 0.876)
	>3	0.783 (0.727, 0.838)	0.807 (0.764, 0.860)	0.738 (0.710, 0.755)	0.713 (0.673, 0.733)	0.713 (0.650, 0.775)
IL10	0-1	0.898 (0.854, 0.943)	0.926 (0.890, 0.963)	0.872 (0.839, 0.885)	0.844 (0.802, 0.865)	0.837 (0.777, 0.896)
	1-3	0.877 (0.836, 0.918)	0.899 (0.857, 0.938)	0.848 (0.824, 0.867)	0.826 (0.804, 0.847)	0.834 (0.785, 0.883)

	>3	0.789 (0.734, 0.844)	0.816 (0.759, 0.866)	0.743 (0.720, 0.765)	0.717 (0.696, 0.743)	0.719 (0.657, 0.782)
IL12P70	0-1	0.886 (0.837, 0.935)	0.916 (0.872, 0.954)	0.863 (0.828, 0.880)	0.832 (0.793, 0.861)	0.823 (0.761, 0.886)
	1-3	0.870 (0.827, 0.913)	0.887 (0.852, 0.922)	0.844 (0.825, 0.860)	0.827 (0.802, 0.852)	0.825 (0.775, 0.876)
	>3	0.777 (0.721, 0.834)	0.804 (0.763, 0.858)	0.738 (0.712, 0.752)	0.711 (0.670, 0.729)	0.708 (0.644, 0.772)
IL1β	0-1	0.899 (0.854, 0.944)	0.921 (0.871, 0.962)	0.866 (0.838, 0.878)	0.844 (0.821, 0.861)	0.835 (0.775, 0.895)
	1-3	0.866 (0.822, 0.909)	0.885 (0.838, 0.924)	0.844 (0.825, 0.856)	0.824 (0.809, 0.841)	0.821 (0.769, 0.872)
	>3	0.772 (0.715, 0.829)	0.807 (0.756, 0.861)	0.732 (0.713, 0.751)	0.696 (0.672, 0.719)	0.697 (0.633, 0.761)
IL2	0-1	0.894 (0.847, 0.940)	0.925 (0.881, 0.962)	0.867 (0.844, 0.882)	0.836 (0.810, 0.859)	0.831 (0.770, 0.892)
	1-3	0.870 (0.827, 0.912)	0.894 (0.846, 0.934)	0.845 (0.825, 0.859)	0.821 (0.806, 0.837)	0.826 (0.776, 0.876)
	>3	0.776 (0.719, 0.832)	0.809 (0.747, 0.862)	0.738 (0.712, 0.754)	0.704 (0.683, 0.724)	0.705 (0.641, 0.768)
IL6	0-1	0.919 (0.880, 0.958)	0.943 (0.909, 0.973)	0.879 (0.842, 0.910)	0.855 (0.813, 0.895)	0.870 (0.818, 0.922)
	1-3	0.866 (0.823, 0.909)	0.887 (0.851, 0.913)	0.842 (0.825, 0.853)	0.821 (0.798, 0.849)	0.824 (0.774, 0.874)
	>3	0.814 (0.762, 0.866)	0.840 (0.795, 0.893)	0.752 (0.721, 0.789)	0.727 (0.690, 0.763)	0.755 (0.696, 0.814)
IL8	0-1	0.896 (0.850, 0.942)	0.922 (0.879, 0.965)	0.869 (0.841, 0.881)	0.843 (0.812, 0.858)	0.833 (0.772, 0.893)
	1-3	0.868 (0.825, 0.912)	0.891 (0.838, 0.924)	0.842 (0.823, 0.856)	0.821 (0.812, 0.844)	0.824 (0.773, 0.875)
	>3	0.771 (0.714, 0.828)	0.804 (0.759, 0.854)	0.738 (0.707, 0.752)	0.705 (0.662, 0.726)	0.702 (0.638, 0.766)
IP10	0-1	0.909 (0.865, 0.952)	0.939 (0.903, 0.975)	0.875 (0.852, 0.899)	0.845 (0.815, 0.877)	0.847 (0.789, 0.906)
	1-3	0.879 (0.837, 0.920)	0.894 (0.855, 0.935)	0.848 (0.830, 0.865)	0.832 (0.812, 0.850)	0.835 (0.785, 0.885)
	>3	0.786 (0.730, 0.842)	0.819 (0.751, 0.873)	0.748 (0.715, 0.776)	0.715 (0.694, 0.745)	0.723 (0.660, 0.786)
TNF-α	0-1	0.905 (0.863, 0.948)	0.930 (0.900, 0.966)	0.873 (0.848, 0.888)	0.849 (0.812, 0.870)	0.842 (0.785, 0.900)

	1-3	0.890 (0.851, 0.929)	0.912 (0.845, 0.946)	0.852 (0.828, 0.881)	0.831 (0.835, 0.864)	0.851 (0.805, 0.898)
	>3	0.785 (0.730, 0.841)	0.811 (0.755, 0.864)	0.743 (0.711, 0.760)	0.717 (0.685, 0.737)	0.717 (0.655, 0.780)
Multiple*	0-1	0.943 (0.910, 0.975)	0.969 (0.934, 1.000)	0.882 (0.846, 0.931)	0.858 (0.825, 0.909)	0.876 (0.823, 0.929)
	1-3	0.895 (0.856, 0.934)	0.919 (0.878, 0.953)	0.850 (0.825, 0.884)	0.828 (0.805, 0.866)	0.847 (0.798, 0.895)
	>3	0.836 (0.787, 0.885)	0.872 (0.822, 0.918)	0.765 (0.722, 0.819)	0.731 (0.688, 0.787)	0.764 (0.706, 0.823)

Notes:
§Apparent AUC statistics are calculated from the original data set, with 95% Wald confidence intervals
‡Bootstrap AUC statistics are calculated on bootstrap samples of the same size as the original data set, N=358.
†Test AUC statistics are calculated by scoring the original data set using the parameter estimates from fitting a logistic regression model on each bootstrap sample; and summarizing the values over the bootstrap replicates.
ΨOptimism corrected AUC is the Apparent AUC minus the difference between the Bootstrap AUC and Test AUC, while cross validated estimates are corrected for optimism using leave-one-out cross validation.
*The model comprising multiple markers concurrently, in addition to risk factors, included IL-6, IL-10, TNF-α, IP10, CXCL13.

Table 4.5: Goodness of Fit Tests Assessing Model Calibration

	Time-Point					
	0-1		1-3		>3	
	H-L GoF Test§	**Model Based GoF Test‡**	**H-L GoF Test§**	**Model Based GoF Test‡**	**H-L GoF Test§**	**Model Based GoF Test‡**
CRP	0.186	0.410	0.599	0.764	0.554	0.915
CXCL13	0.606	0.930	0.542	0.512	0.750	0.441
GM-CSF	0.425	0.214	0.538	0.175	0.966	0.659
IFN-γ	0.964	0.983	0.123	0.087	0.694	0.650
IL10	0.725	0.368	0.373	0.845	0.293	0.906
IL12P70	0.451	0.977	0.179	0.204	0.867	0.561
IL1β	0.891	0.550	0.140	0.179	0.473	0.720
IL2	0.205	0.729	0.321	0.167	0.481	0.952
IL6	0.695	0.961	0.373	0.236	0.231	0.485
IL8	0.927	0.968	0.143	0.267	0.280	0.116
IP10	0.907	0.900	0.193	0.943	0.054	0.216
TNF-α	0.725	0.783	0.286	0.839	0.196	0.852
sCD27	0.955	0.561	0.540	0.631	0.940	0.973
IL-6+IL-10+CXCL13+ IP10+TNF-α†	0.406	0.610	0.216	0.701	0.098	0.465

Notes:
§ The Hosmer-Lemeshow goodness of fit test was conducted yielding these p-values.
‡ A model based goodness of fit test was conducted in order to obtain these p-values.
†The model comprising multiple markers concurrently, in addition to risk factors, included IL-6, IL-10, TNF-α, IP10, CXCL13.

Figure 4.1: Calibration Plots for Models Including Biomarkers

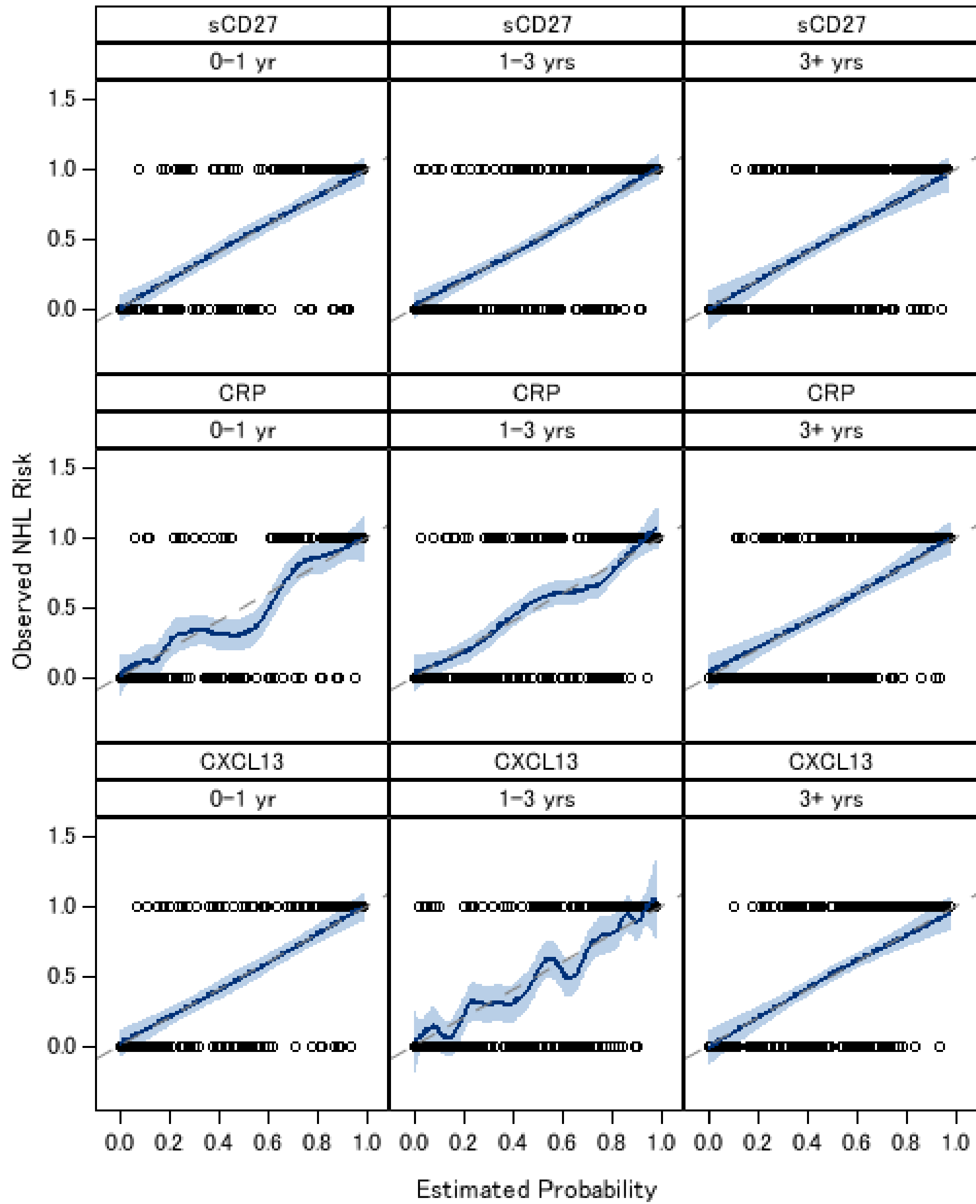


Figure 4.1 Legend:

Observed risks interpolated using LOWESS are plotted against the estimated probability from the logistic risk models. A 45 degree grey dashed line indicates perfect calibration.

Figure 4.1 continued: Calibration Plots for Models Including Biomarkers

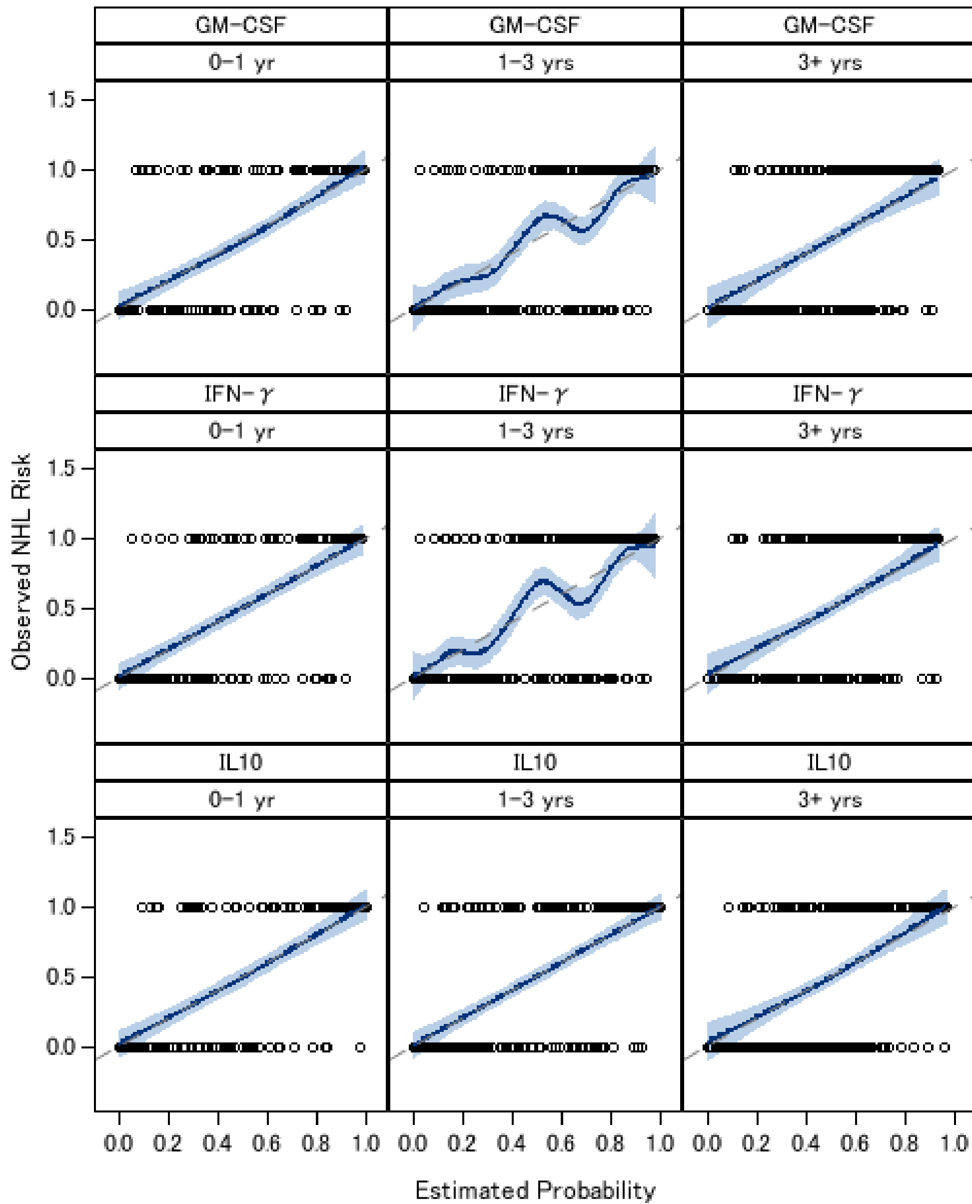


Figure 4.1 Legend:

Observed risks interpolated using LOWESS are plotted against the estimated probability from the logistic risk models. A 45 degree grey dashed line indicates perfect calibration.

Figure 4.1 continued: Calibration Plots for Models Including Biomarkers

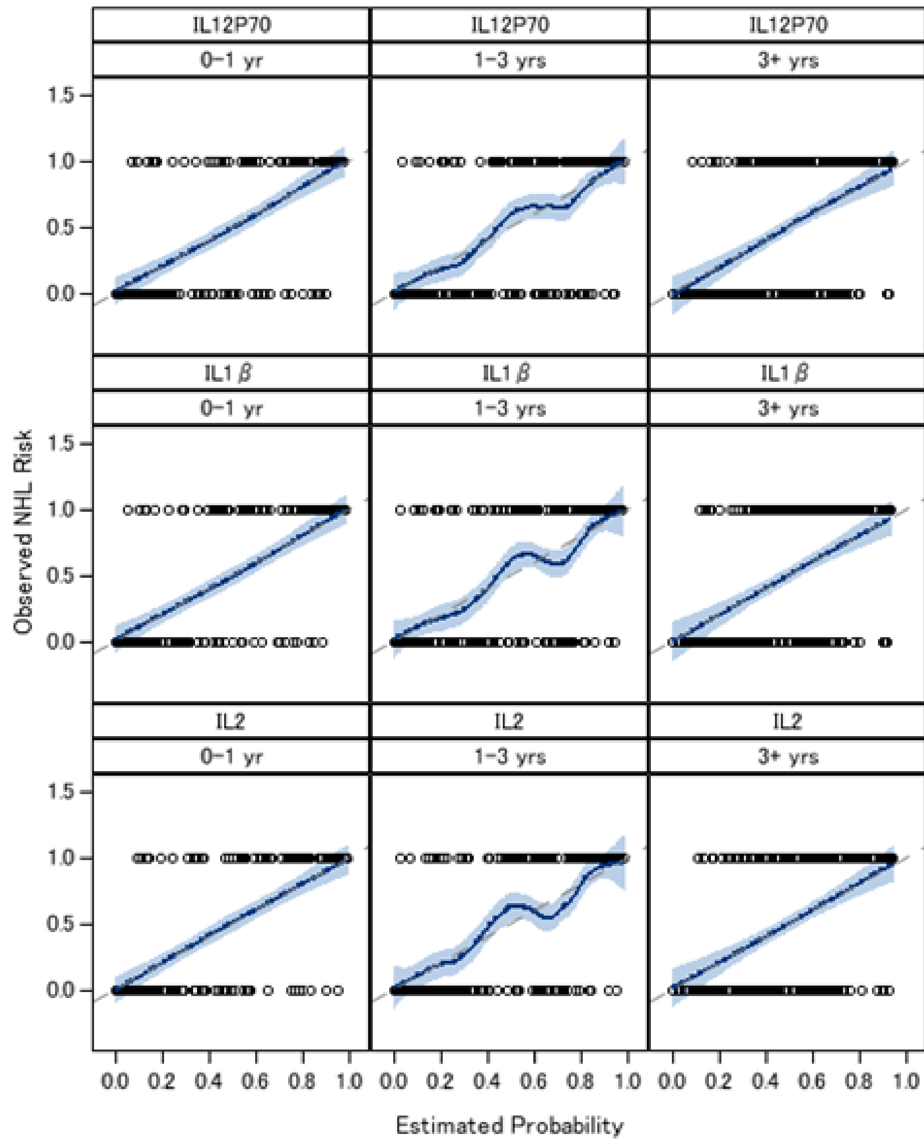


Figure 4.1 Legend:

Observed risks interpolated using LOWESS are plotted against the estimated probability from the logistic risk models. A 45 degree grey dashed line indicates perfect calibration.

Figure 4.1 continued: Calibration Plots for Models Including Biomarkers

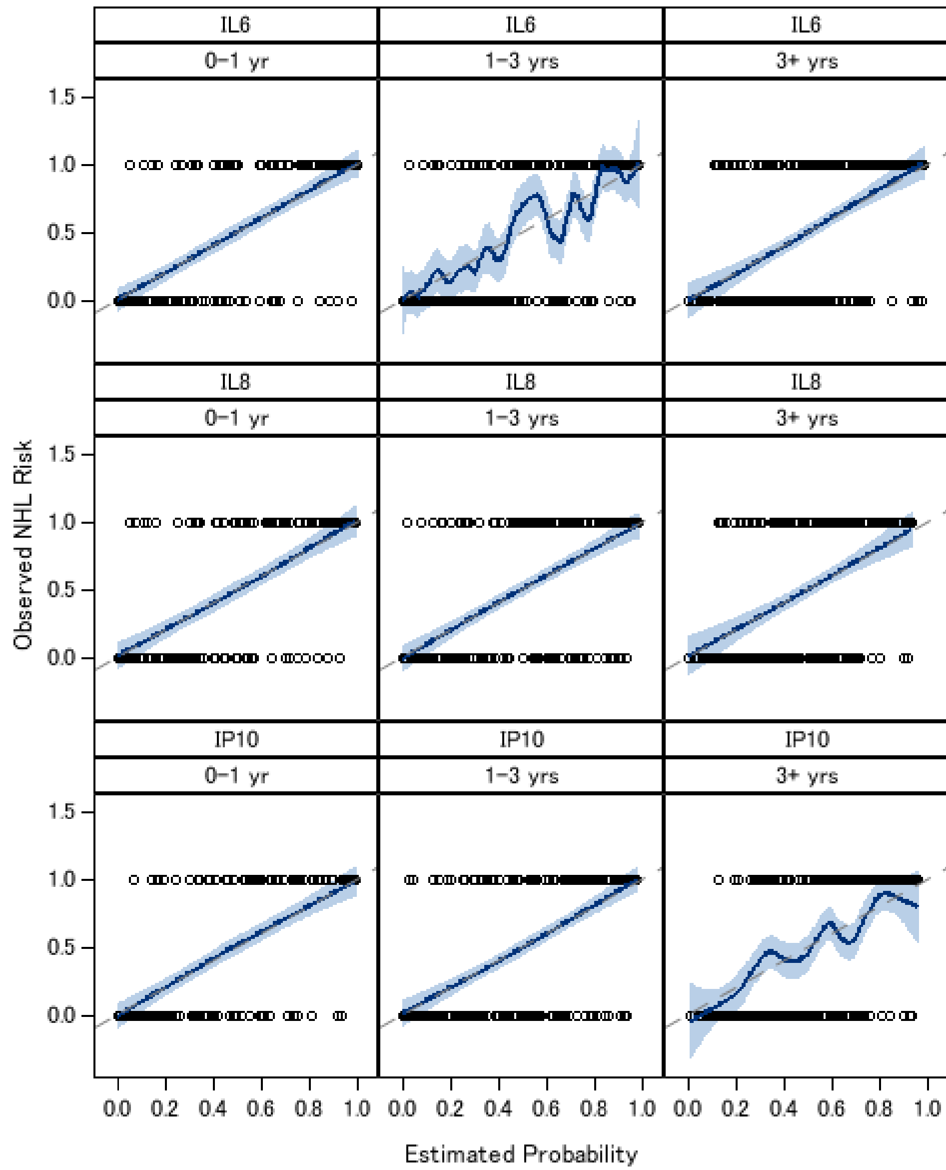


Figure 4.1 Legend:

Observed risks interpolated using LOWESS are plotted against the estimated probability from the logistic risk models. A 45 degree grey dashed line indicates perfect calibration.

Figure 4.1 continued: Calibration Plots for Models Including Biomarkers

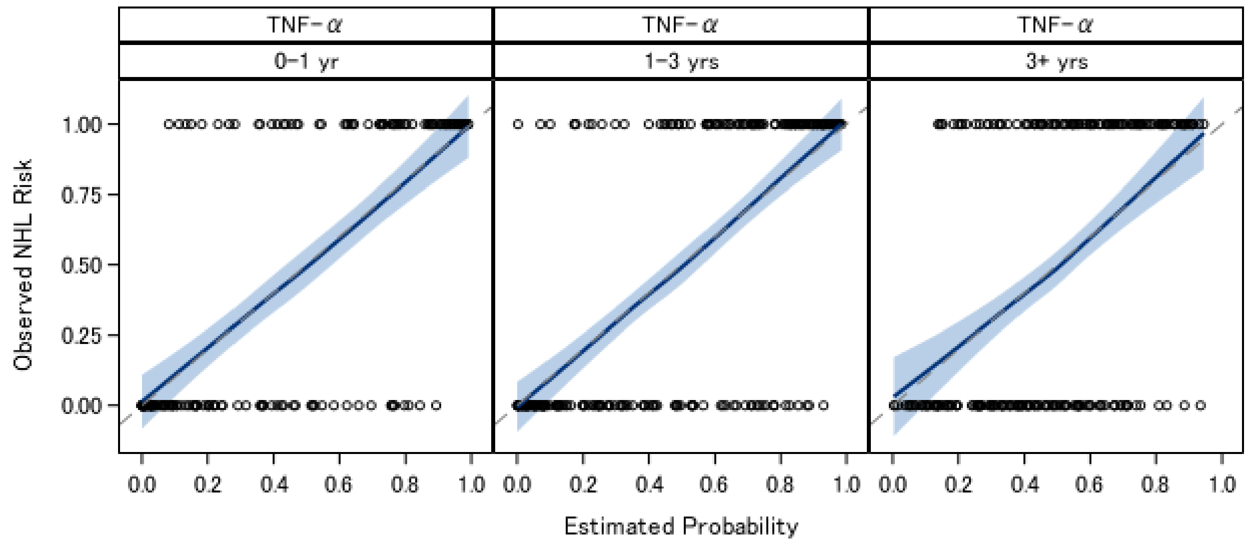


Figure 4.1 Legend:

Observed risks interpolated using LOWESS are plotted against the estimated probability from the logistic risk models. A 45 degree grey dashed line indicates perfect calibration.

Figure 4.1 continued: Calibration Plots for Models Including Biomarkers

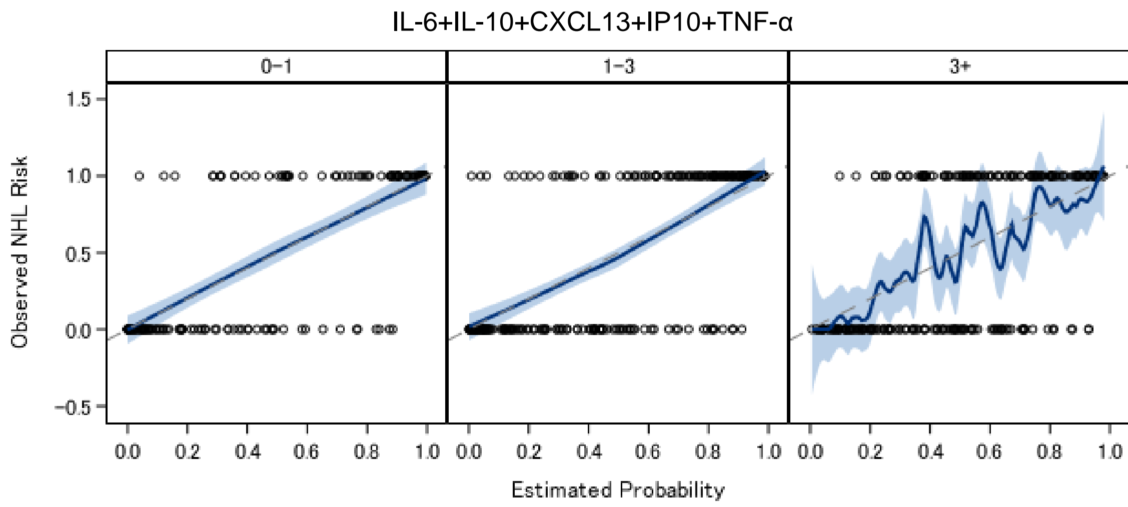


Figure 4.1 Legend:

Observed risks interpolated using LOWESS are plotted against the estimated probability from the logistic risk models. A 45 degree grey dashed line indicates perfect calibration.

5 Chapter V: Public Health Significance & Conclusions

AIDS-NHL is still the leading cause of mortality among HIV positive persons who progress to develop an AIDS defining illness. The exact causes of non-Hodgkin lymphoma (NHL) are still largely unknown, although severe immunodeficiency is among the most consistent factors associated with increased risk. We were able to examine differences in the signatures of several markers of immune activation and inflammation through comprehensive literature review and meta-analyses, and through prospective quantitation of several novel markers in the MACS. Our studies hence provide a significantly comprehensive body of evidence about circulating markers of immune activation and inflammation and their association with, and predictive ability for, AIDS-NHL.

In the three studies we present here, we find a consistent and coherent theme in observations: that elevated levels of several markers of immune activation, and inflammation are observed to precede AIDS-NHL diagnoses over extended periods of time prior to diagnosis. We find that the even in the era of widespread administration of highly active multi-agent antiretrovial therapy, alterations in these biomarkers from normal levels is associated with increased AIDS-NHL incidence, and can thus potentially be used in risk prediction modeling as illustrated by our in exploratory risk prediction models.

The sum-total of the findings from our studies provide further evidence for the general hypothesis that AIDS-NHL develops in as a result of chronic immune activation and inflammation. However, there still remains a need to further elucidate the precise etiologic roles of each of the biomarkers we included in our studies. Our studies contribute to the body of knowledge that will ultimately provide the impetus for improvements in clinical ability to predict,

screen and diagnose AIDS-NHL, to prognosticate and monitor HIV disease progression, and to foster improvements in treatment.

6 Chapter VI: References

1. Chan JK, Banks PM, Cleary ML, et al. A revised European-American classification of lymphoid neoplasms proposed by the International Lymphoma Study Group. A summary version. *Am J Clin Pathol*. 1995;103(5):543-560.
2. Armitage JO, Weisenburger DD. New approach to classifying non-Hodgkin's lymphomas: clinical features of the major histologic subtypes. Non-Hodgkin's Lymphoma Classification Project. *J Clin Oncol*. 1998;16(8):2780-2795.
3. Epeldegui M, Widney DP, Martínez-Maza O. Pathogenesis of AIDS lymphoma: role of oncogenic viruses and B cell activation-associated molecular lesions. *Curr Opin Oncol*. 2006;18(5):444-448.
4. Epeldegui M, Vendrame E, Martínez-Maza O. HIV-associated immune dysfunction and viral infection: role in the pathogenesis of AIDS-related lymphoma. *Immunol Res*. 2010;48(1-3):72-83.
5. Kirk O, Pedersen C, Cozzi-Lepri A, et al. Non-Hodgkin lymphoma in HIV-infected patients in the era of highly active antiretroviral therapy. *Blood*. 2001;98(12):3406-3412.
6. Gibson TM, Morton LM, Shiels MS, Clarke CA, Engels EA. Risk of non-Hodgkin lymphoma subtypes in HIV-infected people during the HAART era: a population-based study. *AIDS*. 2014;28(15):2313-2318.
7. Engels EA, Biggar RJ, Hall HI, et al. Cancer risk in people infected with human immunodeficiency virus in the United States. *Int J Cancer*. 2008;123(1):187-194.
8. Rios A. HIV-related hematological malignancies: a concise review. *Clin Lymphoma Myeloma Leuk*. 2014;14 Suppl:S96-103.
9. Vandenhende M-A, Roussillon C, Henard S, et al. Cancer-Related Causes of Death among HIV-Infected Patients in France in 2010: Evolution since 2000. *PLoS One*. 2015;10(6):e0129550.

10. Beral P. AIDS-associated non-Hodgkin lymphoma. *Lancet*. 1991;337(8745):805-809.

11. Seaberg EC, Wiley D, Martínez-Maza O, et al. Cancer incidence in the multicenter AIDS Cohort Study before and during the HAART era: 1984 to 2007. *Cancer*. 2010;116(23):5507-5516.

12. Hessol NA, Seaberg EC, Preston-Martin S, et al. Cancer risk among participants in the women's interagency HIV study. *J Acquir Immune Defic Syndr*. 2004;36(4):978-985.

13. Chao C, Xu L, Abrams D, et al. Survival of non-Hodgkin lymphoma patients with and without HIV infection in the era of combined antiretroviral therapy. *AIDS*. 2010;24(11):1765-1770.

14. Nolen BM, Breen EC, Bream JH, et al. Circulating mediators of inflammation and immune activation in AIDS-related non-hodgkin lymphoma. *PLoS One*. 2014;9(6):e99144.

15. Kirk O. Non-Hodgkin lymphoma in HIV-infected patients in the era of highly active antiretroviral therapy. *Blood*. 2001;98(12):3406-3412.

16. Carbone A. AIDS-related non-Hodgkin's lymphomas: from pathology and molecular pathogenesis to treatment. *Hum Pathol*. 2002;33(4):392-404.

17. Gaidano G, Carbone A, Dalla-Favera R. Pathogenesis of AIDS-related lymphomas: molecular and histogenetic heterogeneity. *Am J Pathol*. 1998;152(3):623-630.

18. Killebrew D, Shiramizu B. Pathogenesis of HIV-associated non-Hodgkin lymphoma. *Curr HIV Res*. 2004;2(3):215-221.

19. Brenchley JM, Price DA, Schacker TW, et al. Microbial translocation is a cause of systemic immune activation in chronic HIV infection. *Nat Med*. 2006;12(12):1365-1371.

20. Brenchley JM, Douek DC. The mucosal barrier and immune activation in HIV pathogenesis. *Curr Opin HIV AIDS*. 2008;3(3):356-361.

21. Kam LY, Targan SR. Cytokine-based therapies in inflammatory bowel disease. *Curr Opin Gastroenterol*. 1999;15(4):302-307.

22. Olsson J, Poles M, Spetz AL, et al. Human immunodeficiency virus type 1 infection is

associated with significant mucosal inflammation characterized by increased expression of CCR5, CXCR4, and beta-chemokines. *J Infect Dis*. 2000;182(6):1625-1635.

23. McGowan I, Elliott J, Fuerst M, et al. Increased HIV-1 mucosal replication is associated with generalized mucosal cytokine activation. *J Acquir Immune Defic Syndr*. 2004;37(2):1228-1236.
24. Sabin CA. HIV viremia and the development of AIDS-related lymphoma in patients treated with highly active antiretroviral therapy. *J Infect Dis*. 2009;200(1):8-10.
25. Conroy SM, Maskarinec G, Morimoto Y, et al. Non-hodgkin lymphoma and circulating markers of inflammation and adiposity in a nested case-control study: the multiethnic cohort. *Cancer Epidemiol Biomarkers Prev*. 2013;22(3):337-347.
26. Martí A, Marcos A, Martínez JA. Obesity and immune function relationships. *Obes Rev*. 2001;2(2):131-140.
27. Chao C, Page JH. Type 2 diabetes mellitus and risk of non-Hodgkin lymphoma: a systematic review and meta-analysis. *Am J Epidemiol*. 2008;168(5):471-480.
28. Szablewski L. Diabetes mellitus: influences on cancer risk. *Diabetes Metab Res Rev*. 2014;30(7):543-553.
29. Zhang J, Luo J, Liu F, et al. Diabetes mellitus potentiates diffuse large B -cell lymphoma via high levels of CCL5. *Mol Med Rep*. 2014;10(3):1231-1236.
30. Dal Maso L, Franceschi S. Epidemiology of non-Hodgkin lymphomas and other haemolymphopoietic neoplasms in people with AIDS. *Lancet Oncol*. 2003;4(2):110-119.
31. Achenbach CJ, Buchanan AL, Cole SR, et al. HIV viremia and incidence of non-Hodgkin lymphoma in patients successfully treated with antiretroviral therapy. *Clin Infect Dis*. 2014;58(11):1599-1606.
32. Chiu BC-H, Hou N. Epidemiology and etiology of non-hodgkin lymphoma. *Cancer Treat Res*. 2015;165:1-25.
33. Mounier N, Spina M, Gisselbrecht C. Modern management of non-Hodgkin lymphoma in

HIV-infected patients. *Br J Haematol*. 2007;136(5):685-698.

34. Guech-Ongey M, Simard EP, Anderson WF, et al. AIDS-related Burkitt lymphoma in the United States: what do age and CD4 lymphocyte patterns tell us about etiology and/or biology? *Blood*. 2010;116(25):5600-5604.

35. Chao C, Jacobson LP, Tashkin D, et al. Recreational drug use and T lymphocyte subpopulations in HIV-uninfected and HIV-infected men. *Drug Alcohol Depend*. 2008;94(1-3):165-171.

36. Armenian HK, Hoover DR, Rubb S, et al. Risk factors for non-Hodgkin's lymphomas in acquired immunodeficiency syndrome (AIDS). *Am J Epidemiol*. 1996;143(4):374-379.

37. Akhtar-Khaleel WZ, Cook RL, Shoptaw S, et al. Trends and Predictors of Cigarette Smoking Among HIV Seropositive and Seronegative Men: The Multicenter Aids Cohort Study. *AIDS Behav*. June 2015.

38. Nelson RA, Levine AM, Marks G, Bernstein L. Alcohol, tobacco and recreational drug use and the risk of non-Hodgkin's lymphoma. *Br J Cancer*. 1997;76(11):1532-1537.

39. Kaslow RA, Ostrow DG, Detels R, Phair JP, Polk BF, Rinaldo CR. The Multicenter AIDS Cohort Study: rationale, organization, and selected characteristics of the participants. *Am J Epidemiol*. 1987;126(2):310-318.

40. Detels R, Jacobson L, Margolick J, et al. The multicenter AIDS Cohort Study, 1983 to …. *Public Health*. 2012;126(3):196-198.

41. Grulich AE, Vajdic CM, Cozen W. Altered Immunity as a Risk Factor for Non-Hodgkin Lymphoma. *Cancer Epidemiol Prev Biomarkers*. 2007;16(3):405-408.

42. Gibson TM, Morton LM, Shiels MS, Clarke CA, Engels EA. Risk of non-Hodgkin lymphoma subtypes in HIV-infected people during the HAART era: a population-based study. *AIDS*. 2014;28(15):2313-2318.

43. Gibson TM, Engels EA, Clarke CA, Lynch CF, Weisenburger DD, Morton LM. Risk of diffuse large B-cell lymphoma after solid organ transplantation in the United States. *Am J*

Hematol. 2014;89(7):714-720.

44. Morscio J, Dierickx D, Tousseyn T. Molecular Pathogenesis of B-Cell Posttransplant Lymphoproliferative Disorder: What Do We Know So Far? *Clin Dev Immunol*. 2013;2013:1-13.
45. Morscio J, Dierickx D, Ferreiro JF, et al. Gene expression profiling reveals clear differences between EBV-positive and EBV-negative posttransplant lymphoproliferative disorders. *Am J Transplant*. 2013;13(5):1305-1316.
46. Bassig BA, Shu X-O, Koh W-P, et al. Soluble levels of CD27 and CD30 are associated with risk of non-Hodgkin lymphoma in three Chinese prospective cohorts. *Int J Cancer*. 2015;137(11):2688-2695.
47. Purdue MP, Hofmann JN, Kemp TJ, et al. A prospective study of 67 serum immune and inflammation markers and risk of non-Hodgkin lymphoma. *Blood*. 2013;122(6):951-957.
48. Saberi Hosnijeh F, Krop EJM, Scoccianti C, et al. Plasma cytokines and future risk of non-Hodgkin lymphoma (NHL): a case-control study nested in the Italian European Prospective Investigation into Cancer and Nutrition. *Cancer Epidemiol Biomarkers Prev*. 2010;19(6):1577-1584.
49. Vendrame E, Martínez-Maza O. Assessment of pre-diagnosis biomarkers of immune activation and inflammation: insights on the etiology of lymphoma. *J Proteome Res*. 2011;10(1):113-119.
50. Hosnijeh FS, Portengen L, Späth F, et al. Soluble B-cell activation marker of sCD27 and sCD30 and future risk of B-cell lymphomas: A nested case-control study and meta-analyses. *Int J cancer*. 2016;138(10):2357-2367.
51. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA*. 2000;283(15):2008-2012.
52. Hartung J, Knapp G, Sinha BK. *Statistical Meta-Analysis with Applications*. Vol 738. John

Wiley & Sons; 2011.

53. Epstein M, Rosner B, Breen EC, et al. Pre-diagnosis plasma immune markers and risk of non-Hodgkin lymphoma in two prospective cohort studies. *Unpublished*. 2017.

54. Breen EC, Hussain SK, Magpantay L, et al. B-Cell Stimulatory Cytokines and Markers of Immune Activation Are Elevated Several Years Prior to the Diagnosis of Systemic AIDS-Associated Non-Hodgkin B-Cell Lymphoma. *Cancer Epidemiol Biomarkers Prev*. 2011;20(7):1303-1314.

55. Rabkin CS, Engels EA, Landgren O, et al. Circulating cytokine levels, Epstein-Barr viremia, and risk of acquired immunodeficiency syndrome-related non-Hodgkin lymphoma. *Am J Hematol*. 2011;86(10):875-878.

56. De Roos AJ, Mirick DK, Edlefsen KL, et al. Markers of B-cell activation in relation to risk of non-Hodgkin lymphoma. *Cancer Res*. 2012;72(18):4733-4743.

57. Hussain SK, Hessol NA, Levine AM, et al. Serum biomarkers of immune activation and subsequent risk of non-hodgkin B-cell lymphoma among HIV-infected women. *Cancer Epidemiol Biomarkers Prev*. 2013;22(11):2084-2093.

58. Vendrame E, Hussain SK, Breen EC, et al. Serum Levels of Cytokines and Biomarkers for Inflammation and Immune Activation, and HIV-Associated Non-Hodgkin B-Cell Lymphoma Risk. *Cancer Epidemiol Biomarkers Prev*. 2013;23(2):343-349.

59. Edlefsen KL, Martínez-Maza O, Madeleine MM, et al. Cytokines in serum in relation to future non-Hodgkin lymphoma risk: Evidence for associations by histologic subtype. *Int J Cancer*. January 2014.

60. Purdue MP, Lan Q, Martinez-Maza O, et al. A prospective study of serum soluble CD30 concentration and risk of non-Hodgkin lymphoma. *Blood*. 2009;114(13):2730-2732.

61. Gu Y, Shore RE, Arslan AA, et al. Circulating cytokines and risk of B-cell non-Hodgkin lymphoma: a prospective study. *Cancer Causes Control*. 2010;21(8):1323-1333.

62. Purdue MP, Lan Q, Bagni R, et al. Prediagnostic serum levels of cytokines and other

immune markers and risk of non-hodgkin lymphoma. *Cancer Res*. 2011;71(14):4898-4907.

63. Vermeulen R, Hosnijeh FS, Portengen L, et al. Circulating soluble CD30 and future risk of lymphoma; evidence from two prospective studies in the general population. *Cancer Epidemiol Biomarkers Prev*. 2011;20(9):1925-1927.

64. Purdue MP, Lan Q, Kemp TJ, et al. Elevated serum sCD23 and sCD30 up to two decades prior to diagnosis associated with increased risk of non-Hodgkin lymphoma. *Leukemia*. 2015;29(6):1429-1431.

65. Orsini N, Li R, Wolk A, Khudyakov P, Spiegelman D. Meta-Analysis for Linear and Nonlinear Dose-Response Relations: Examples, an Evaluation of Approximations, and Software. *Am J Epidemiol*. 2012;175(1):66-73.

66. Hamling J, Lee P, Weitkunat R, Ambühl M. Facilitating meta-analyses by deriving relative effect and precision estimates for alternative comparisons from a set of estimates presented by exposure level or disease category. *Stat Med*. 2008;27(7):954-970.

67. Greenland S, Longnecker MP. Methods for trend estimation from summarized dose-response data, with applications to meta-analysis. *Am J Epidemiol*. 1992;135(11):1301-1309.

68. Hozo SP, Djulbegovic B, Hozo I. Estimating the mean and variance from the median, range, and the size of a sample. *BMC Med Res Methodol*. 2005;5(1):13.

69. Wan X, Wang W, Liu J, Tong T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. *BMC Med Res Methodol*. 2014;14(1):135.

70. Fritz A, Percy C, Jack A, et al. *International Classification of Diseases for Oncology*. Geneva: World Health Organization; 2000.

71. Hardy RJ, Thompson SG. A likelihood approach to meta-analysis with random effects. *Stat Med*. 1996;15(6):619-629.

72. Cochran WG. The Comparison of Percentages in Matched Samples. *Biometrika*. 1950;37(3/4):256.

73. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med*. 2002;21(11):1539-1558.

74. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327(7414):557-560.

75. Light RJ, Pillemer DB. *Summing up: The Science of Reviewing Research.* Harvard University Press; 1984.

76. Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ*. 1997;315(7109):629-634.

77. Duval S, Tweedie R. Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*. 2000;56(2):455-463.

78. Duval S, Tweedie R. A Nonparametric 'Trim and Fill" Method of Accounting for Publication Bias in Meta-Analysis. *J Am Stat Assoc*. 2000;95(449):89-98.

79. Greenland S, O'Rourke K. On the bias produced by quality scores in meta-analysis, and a hierarchical view of proposed solutions. *Biostatistics*. 2001;2(4):463-471.

80. R Core Team. R: A Language and Environment for Statistical Computing. 2016.

81. Schwarzer G. meta: An R package for meta-analysis. *R News*. 2007;7(3):40-45.

82. Viechtbauer W. Conducting Meta-Analyses in R with the metafor Package. *J Stat Softw*. 2010;36(3):1-48.

83. Von Hippel PT. The heterogeneity statistic I2 can be biased in small meta-analyses. *BMC Med Res Methodol*. 2015;15(1).

84. Kishimoto T, Akira S, Narazaki M, Taga T. Interleukin-6 family of cytokines and gp130. *Blood*. 1995;86(4).

85. Bertolini JN, Benson EM. The role of human interleukin-6 in B-cell isotype regulation and differentiation. *Cell Immunol*. 1990;125(1):197-209.

86. Annunziato F, Cosmi L, Liotta F, Maggi E, Romagnani S. The phenotype of human Th17 cells and their precursors, the cytokines that mediate their differentiation and the role of Th17 cells in inflammation. *Int Immunol*. 2008;20(11):1361-1368.

87. Romagnani S, Maggi E, Liotta F, Cosmi L, Annunziato F. Properties and origin of human Th17 cells. *Mol Immunol*. 2009;47(1):3-7.

88. Ekstrom Smedby K, Vajdic CM, Falster M, et al. Autoimmune disorders and risk of non-Hodgkin lymphoma subtypes: a pooled analysis within the InterLymph Consortium. *Blood*. 2008;111(8):4029-4038.

89. Biggar RJ, Chaturvedi AK, Goedert JJ, Engels EA. AIDS-Related Cancer and Severity of Immunosuppression in Persons With AIDS. *JNCI J Natl Cancer Inst*. 2007;99(12):962-972.

90. Yanik E, Katki H, Engels E. High Cancer Risk Among the HIV-Infected Elderly in the United States. CROI 2015, Feb 23-26, 2015, Seattle, WA.

91. Shiels MS, Engels EA, Linet MS, et al. The epidemic of non-Hodgkin lymphoma in the United States: disentangling the effect of HIV, 1992-2009. *Cancer Epidemiol Biomarkers Prev*. 2013;22(6):1069-1078.

92. Mocellin S, Marincola F, Rossi CR, Nitti D, Lise M. The multifaceted relationship between IL-10 and adaptive immunity: putting together the pieces of a puzzle. *Cytokine Growth Factor Rev*. 2004;15(1):61-76.

93. Rothman N, Skibola CF, Wang SS, et al. Genetic variation in TNF and IL10 and risk of non-Hodgkin lymphoma: a report from the InterLymph Consortium. *Lancet Oncol*. 2006;7(1):27-38.

94. Khatri VP, Caligiuri MA. A review of the association between interleukin-10 and human B-cell malignancies. *Cancer Immunol Immunother*. 1998;46(5):239-244.

95. Czarneski J, Lin YC, Chong S, et al. Studies in NZB IL-10 knockout mice of the requirement of IL-10 for progression of B-cell lymphoma. *Leukemia*. 2004;18(3):597-606.

96. Emilie D, Touitou R, Raphael M, et al. In vivo production of interleukin-10 by malignant cells in AIDS lymphomas. *Eur J Immunol*. 1992;22(11):2937-2942.

97. Voorzanger N, Touitou R, Garcia E, et al. Interleukin (IL)-10 and IL-6 are produced in vivo by non-Hodgkin's lymphoma cells and act as cooperative growth factors. *Cancer Res*. 1996;56(23):5499-5505.

98. Yu X, Chen B, Cheng J, Gao C, Zhang X, Bao W. The interleukin-10-1082A>G polymorphism and lymphoma risk: A meta-analysis. *Cancer Biomarkers*. 2014;14(5):381-388.

99. Li G, Li D. Relationship between IL-10 gene polymorphisms and the risk of non-Hodgkin lymphoma: A meta-analysis. *Hum Immunol*. 2016;77(5):418-425.

100. Dai Z-M, He A-L, Zhang W-G, et al. Association of the four common polymorphisms in interleukin-10 (rs1800890, rs1800896, rs1800871, and rs1800872) with non-Hodgkin's lymphoma risk: a meta-analysis. *Int J Clin Exp Med*. 2014;7(12):4720-4733.

101. Cao H-Y, Zou P, Zhou H. Genetic association of interleukin-10 promoter polymorphisms and susceptibility to diffuse large B-cell lymphoma: A meta-analysis. *Gene*. 2013;519(2).

102. Balkwill F, Mantovani A. Inflammation and cancer: back to Virchow? *Lancet*. 2001;357(9255):539-545.

103. Husson H, Lugli SM, Ghia P, et al. Functional Effects of TNF and Lymphotoxin α1β2 on FDC-like Cells. *Cell Immunol*. 2000;203(2):134-143.

104. Flier JS, Underhill LH, Bazzoni F, Beutler B. The Tumor Necrosis Factor Ligand and Receptor Families. *N Engl J Med*. 1996;334(26):1717-1725.

105. Körner H, Cretney E, Wilhelm P, et al. Tumor necrosis factor sustains the generalized lymphoproliferative disorder (gld) phenotype. *J Exp Med*. 2000;191(1):89-96.

106. Purdue MP, Lan Q, Kricker A, et al. Polymorphisms in immune function genes and risk of non-Hodgkin lymphoma: findings from the New South Wales non-Hodgkin Lymphoma Study. *Carcinogenesis*. 2007;28(3):704-712.

107. Chouchane L, Ahmed SB, Baccouche S, Remadi S. Polymorphism in the tumor necrosis factor-alpha promotor region and in the heat shock protein 70 genes associated with malignant tumors. *Cancer*. 1997;80(8):1489-1496.

108. Cyster JG, Ansel KM, Ngo VN, et al. A chemokine-driven positive feedback loop organizes lymphoid follicles. *Nature*. 2000;406(6793):309-314.

109. Widney DP, Gui D, Said JW, et al. Expression and Function of the Chemokine, CXCL13, and Its Receptor, CXCR5, in AIDS-Associated Non-Hodgkin's Lymphoma. *AIDS Res Treat*. 2009;4(Suppl 2):1-9.

110. Bürkle A, Niedermeier M, Schmitt-Gräff A, Wierda WG, Keating MJ, Burger JA. Overexpression of the CXCR5 chemokine receptor, and its ligand, CXCL13 in B-cell chronic lymphocytic leukemia. *Blood*. 2007;110(9).

111. Husson H, S. Freedman A, A. Cardoso A, et al. CXCL13 (BCA-1) is produced by follicular lymphoma cells: role in the accumulation of malignant B cells. *Br J Haematol*. 2002;119(2):492-495.

112. Gordon J. CD23 and B cell activation. *Clin Exp Allergy*. 1992;22(2):199-204.

113. Schroeder JR, Saah AJ, Hoover DR, et al. Serum soluble CD23 level correlates with subsequent development of AIDS-related non-Hodgkin's lymphoma. *Cancer Epidemiol Biomarkers Prev*. 8(11):979-984.

114. Gordon J, Millsum MJ, Flores-Romo L, Gillis S. Regulation of resting and cycling human B lymphocytes via surface IgM and the accessory molecules interleukin-4, CD23 and CD40. *Immunology*. 1989;68(4):526-531.

115. Herbelin A, Elhadadu S, Ouaazu F, de Grooteu D, Descamps-Latschaf B. Soluble CD23 potentiates interleukin-1-induced secretion of interleukin-6 and interleukin-1 receptor antagonist by human monocytes. *Eur J Immunol*. 1994;24(8):1869-1873.

116. Lens SM, Drillenburg P, den Drijver BF, et al. Aberrant expression and reverse signalling of CD70 on malignant B cells. *Br J Haematol*. 1999;106(2):491-503.

117. Nawrocki JF, Kirsten ES, Fisher RI. Biochemical and structural properties of a Hodgkin's disease-related membrane protein. *J Immunol*. 1988;141(2).

118. Goodwin RG, Alderson MR, Smith CA, et al. Molecular and biological characterization of a ligand for CD27 defines a new family of cytokines with homology to tumor necrosis factor. *Cell*. 1993;73(3):447-456.

119. Maurer D, Fischer GF, Fae I, et al. IgM and IgG but not cytokine secretion is restricted to the CD27+ B lymphocyte subset. *J Immunol*. 1992;148(12):3700-3705.

120. Horie R, Watanabe T. CD30: expression and function in health and disease. *Semin Immunol*. 1998;10(6):457-470.

121. Shanebeck KD, Maliszewski CR, Kennedy MK, et al. Regulation of murine B cell growth and differentiation by CD30 ligand. *Eur J Immunol*. 1995;25(8):2147-2153.

122. Croft M. Co-stimulatory members of the TNFR family: keys to effective T-cell immunity? *Nat Rev Immunol*. 2003;3(8):609-620.

123. Lens SMA, Tesselaar K, van Oers MHJ, van Lier RAW. Control of lymphocyte function through CD27–CD70 interactions. *Semin Immunol*. 1998;10(6):491-499.

124. Kennedy MK, Willis CR, Armitage RJ. Deciphering CD30 ligand biology and its role in humoral immunity. *Immunology*. 2006;118(2):143-152.

125. Widney D, Gundapp G, Said JW, et al. Aberrant Expression of CD27 and Soluble CD27 (sCD27) in HIV Infection and in AIDS-Associated Lymphoma. *Clin Immunol*. 1999;93(2):114-123.

126. Breen EC, Fatahi S, Epeldegui M, Boscardin WJ, Detels R, Martínez-Maza O. Elevated Serum Soluble CD30 Precedes the Development of AIDS-Associated Non-Hodgkin B Cell Lymphoma. *Tumor Biol*. 2006;27(4):187-194.

127. Widney DP, Breen EC, Boscardin WJ, et al. Serum Levels of the Homeostatic B Cell Chemokine, CXCL13, Are Elevated During HIV Infection. *J Interf Cytokine Res*. 2005;25(11):702-706.

128. Wada NI, Jacobson LP, Margolick JB, et al. The effect of HAART-induced HIV suppression on circulating markers of inflammation and immune activation. *AIDS*. 2015;29(4):463-471.

129. Lin L, Chu H. Quantifying publication bias in meta-analysis. *Biometrics*. 2018;74(3):785-794.

130. Breen EC, Reynolds SM, Cox C, et al. Multisite comparison of high-sensitivity multiplex cytokine assays. *Clin Vaccine Immunol*. 2011;18(8):1229-1242.

131. Noble JE, Wang L, Cerasoli E, et al. An international comparability study to determine the sources of uncertainty associated with a non-competitive sandwich fluorescent ELISA. *Clin Chem Lab Med*. 2008;46(7):1033-1045.

132. Hsu SM, Waldron JW, Hsu PL, Hough AJ. Cytokines in malignant lymphomas: review and prospective evaluation. *Hum Pathol*. 1993;24(10):1040-1057.

133. Carbone A, Gloghini A, Cabras A, Elia G. Differentiating germinal center-derived lymphomas through their cellular microenvironment. *Am J Hematol*. 2009;84(7):435-438.

134. Schmitter D, Koss M, Niederer E, Stahel RA, Pichert G. T-cell derived cytokines co-stimulate proliferation of CD40-activated germinal centre as well as follicular lymphoma cells. *Hematol Oncol*. 1997;15(4):197-207.

135. Holm S. A Simple Sequentially Rejective Multiple Test Procedure. *Scand J Stat*. 1979;6:65-70.

136. Benjamini Y, Yekutieli D. The control of false discovery rate in multiple testing under dependency. *Ann Stat*. 2001;29(4):1165-1188.

137. Vendrame E, Hussain SK, Breen EC, et al. Serum Levels of Cytokines and Biomarkers for Inflammation and Immune Activation, and {HIV}-Associated Non-Hodgkin B-Cell Lymphoma Risk. *Cancer Epidemiol Biomarkers {&} Prev*. 2013;23(2):343-349.

138. Rabkin CS, Engels EA, Landgren O, et al. Circulating cytokine levels, Epstein-Barr viremia, and risk of acquired immunodeficiency syndrome-related non-Hodgkin

lymphoma. *Am J Hematol*. 2011;86(10):875-878.

139. Tulpule A, Levine A. AIDS-related lymphoma. *Blood Rev*. 1999;13(3):147-150.

140. Tirelli U, Bernardi D. Impact of HAART on the clinical management of AIDS-related cancers. *Eur J Cancer*. 2001;37(10):1320-1324.

141. Vendrame E, Hussain SK, Breen EC. Serum Levels of Cytokines and Biomarkers for Inflammation and Immune Activation , and HIV-Associated Non-Hodgkin B-Cell Lymphoma Risk. 2014.

142. Breen EC, Boscardin WJ, Detels R, et al. Non-Hodgkin's B cell lymphoma in persons with acquired immunodeficiency syndrome is associated with increased serum levels of IL10, or the IL10 promoter −592 C/C genotype. *Clin Immunol*. 2003;109(2):119-129.

143. Schroeder JR, Saah AJ, Ambinder RF, et al. Serum sCD23 Level in Patients with AIDS-Related Non-Hodgkin's Lymphoma Is Associated with Absence of Epstein–Barr Virus in Tumor Tissue. *Clin Immunol*. 1999;93(3):239-244.

144. Breen EC, Epeldegui M, Boscardin WJ, Widney DP, Detels R, Martínez-Maza O. Elevated levels of soluble CD44 precede the development of AIDS-associated non-Hodgkin's B-cell lymphoma. *AIDS*. 2005;19(15):1711-1712.

145. Landgren O, Goedert JJ, Rabkin CS, et al. Circulating Serum Free Light Chains As Predictive Markers of AIDS-Related Lymphoma. *J Clin Oncol*. 2010;28(5):773-779.

146. Breen EC, Hussain SK, Magpantay L, et al. B-Cell Stimulatory Cytokines and Markers of Immune Activation Are Elevated Several Years Prior to the Diagnosis of Systemic {AIDS}-Associated Non-Hodgkin B-Cell Lymphoma. *Cancer Epidemiol Biomarkers {&} Prev*. 2011;20(7):1303-1314.

147. Hussain SK, Zhu W, Chang S-CS-C, et al. Serum levels of the chemokine CXCL13, genetic variation in CXCL13 and its receptor CXCR5, and HIV-associated non-hodgkin B-cell lymphoma risk. *Cancer Epidemiol Biomarkers Prev*. 2013;22(2):295-307.

148. Widney DP, Gui D, Popoviciu LM, et al. Expression and Function of the Chemokine,

CXCL13, and Its Receptor, CXCR5, in Aids-Associated Non-Hodgkin's Lymphoma. *AIDS Res Treat*. 2010;2010:1-9.

149. Association AD. *Standards of Medical Care in Diabetes--2010.* Vol 33 Suppl 1.; 2010.

150. Rice K. Equivalence Between Conditional and Random-Effects Likelihoods for Pair-Matched Case-Control Studies. *J Am Stat Assoc*. 2008;103(481):385-396.

151. Arnau J, Bono R, Blanca MJ, Bendayan R. Using the linear mixed model to analyze nonnormal data distributions in longitudinal designs. *Behav Res Methods*. 2012;44(4):1224-1238.

152. Arnau J, Bendayan R, Blanca MJ, Bono R. The effect of skewness and kurtosis on the robustness of linear mixed models. *Behav Res Methods*. 2013;45(3):873-879.

153. Yang S, Li J-Y, Xu W. Role of BAFF/BAFF-R axis in B-cell non-Hodgkin lymphoma. *Crit Rev Oncol Hematol*. 2014;91(2):113-122.

154. Novak AJ, Grote DM, Stenson M, et al. Expression of BLyS and its receptors in B-cell non-Hodgkin lymphoma: correlation with disease activity and patient outcome. *Blood*. 2004;104(8):2247-2253.

155. Murakami S. Soluble interleukin-2 receptor in cancer. *Front Biosci*. 2004;9:3085-3090.

156. Makgoeng SB, Bolanos R, Jeon CY, et al. Markers of Immune Activation and Inflammation, and Non-Hodgkin Lymphoma: A Meta-Analysis of Prospective Studies. *Unpublished*. 2018.

157. Yasuda N, Lai PK, Ip SH, et al. Soluble interleukin 2 receptors in sera of Japanese patients with adult T cell leukemia mark activity of disease. *Blood*. 1988;71(4):1021-1026.

158. Yoshida N, Oda M, Kuroda Y, et al. Clinical significance of sIL-2R levels in B-cell lymphomas. *PLoS One*. 2013;8(11):e78730.

159. De Paiva CS, Yoon K-C, Pangelinan SB, et al. Cleavage of functional IL-2 receptor alpha chain (CD25) from murine corneal and conjunctival epithelia by MMP-9. *J Inflamm (Lond)*. 2009;6:31.

160. El Houda Agueznay N, Badoual C, Hans S, et al. Soluble interleukin-2 receptor and metalloproteinase-9 expression in head and neck cancer: prognostic value and analysis of their relationships. *Clin Exp Immunol*. 2007;150(1):114-123.

161. Ennishi D, Yokoyama M, Terui Y, et al. Soluble interleukin-2 receptor retains prognostic value in patients with diffuse large B-cell lymphoma receiving rituximab plus CHOP (RCHOP) therapy. *Ann Oncol*. 2009;20(3):526-533.

162. Goto N, Tsurumi H, Goto H, et al. Serum soluble interleukin-2 receptor (sIL-2R) level is associated with the outcome of patients with diffuse large B cell lymphoma treated with R-CHOP regimens. *Ann Hematol*. 2012;91(5):705-714.

163. Goto H, Tsurumi H, Takemura M, et al. Serum-soluble interleukin-2 receptor (sIL-2R) level determines clinical outcome in patients with aggressive non-Hodgkin's lymphoma: in combination with the International Prognostic Index. *J Cancer Res Clin Oncol*. 2005;131(2):73-79.

164. Kovacs E, Kuehn JJ. Measurements of IL-6, soluble IL-6 receptor and soluble gp130 in sera of B-cell lymphoma patients. Does viscum album treatment affect these parameters? *Biomed Pharmacother*. 2002;56(3):152-158.

165. Jost PJ, Ruland J. Aberrant NF-kappaB signaling in lymphoma: mechanisms, consequences, and therapeutic implications. *Blood*. 2007;109(7):2700-2707.

166. Smedby KE, Foo JN, Skibola CF, et al. GWAS of follicular lymphoma reveals allelic heterogeneity at 6p21.32 and suggests shared genetic susceptibility with diffuse large B-cell lymphoma. *PLoS Genet*. 2011;7(4):e1001378.

167. Lane BR, King SR, Bock PJ, Strieter RM, Coffey MJ, Markovitz DM. The C-X-C chemokine IP-10 stimulates HIV-1 replication. *Virology*. 2003;307(1):122-134.

168. Charbonneau B, Maurer MJ, Ansell SM, et al. Pretreatment circulating serum cytokines associated with follicular and diffuse large B-cell lymphoma: A clinic-based case-control study. *Cytokine*. 2012;60(3):882-889.

169. Ouedraogo DE, Makinson A, Kuster N, et al. Increased T-Cell Activation and Th1 Cytokine Concentrations Prior to the Diagnosis of B-Cell Lymphoma in {HIV} Infected Patients. *J Clin Immunol*. 2012;33(1):22-29.

170. Chen Q, Zheng T, Lan Q, et al. Single-nucleotide polymorphisms in genes encoding for CC chemokines were not associated with the risk of non-Hodgkin lymphoma. *Cancer Epidemiol Biomarkers Prev*. 2013;22(7):1332-1335.

171. Cagigi A, Mowafi F, Phuong Dang L V, et al. Altered expression of the receptor-ligand pair CXCR5/CXCL13 in B cells during chronic HIV-1 infection. *Blood*. 2008;112(12):4401-4410.

172. Epeldegui M, Martínez-Maza O. Immune Activation: Contribution to AIDS-Associated Non-Hodgkin Lymphoma. *For Immunopathol Dis Therap*. 2015;6(1-2):79-90.

173. Pepe MS, Janes H, Longton G, Leisenring W, Newcomb P. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *Am J Epidemiol*. 2004;159(9):882-890.

174. Pepe MS. Evaluating technologies for classification and prediction in medicine. *Stat Med*. 2005;24(24):3687-3696.

175. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J*. 2014;35(29):1925-1931.

176. Steyerberg EW, Eijkemans MJ, Harrell FE, Habbema JD. Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. *Stat Med*. 2000;19(8):1059-1079.

177. Steyerberg EW, Borsboom GJJM, van Houwelingen HC, Eijkemans MJC, Habbema JDF. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. *Stat Med*. 2004;23(16):2567-2586.

178. Pepe M, Janes H. Methods for Evaluating Prediction Performance of Biomarkers and Tests. In: Springer New York; 2013:107-142.

179. Larsen JT, Kumar SK, Dispenzieri A, Kyle RA, Katzmann JA, Rajkumar S V. Serum free light chain ratio as a biomarker for high-risk smoldering multiple myeloma. *Leukemia*. 2013;27(4):941-946.

180. Kyle RA, Rajkumar S V. Criteria for diagnosis, staging, risk stratification and response assessment of multiple myeloma. *Leukemia*. 2009;23(1):3-9.

181. Dispenzieri A, Kyle RA, Katzmann JA, et al. Immunoglobulin free light chain ratio is an independent risk factor for progression of smoldering (asymptomatic) multiple myeloma. *Blood*. 2008;111(2).

182. Prugger C, Luc G, Haas B, et al. Multiple biomarkers for the prediction of ischemic stroke: the PRIME study. *Arterioscler Thromb Vasc Biol*. 2013;33(3):659-666.

183. Gold KA, Kim ES, Liu DD, et al. Prediction of survival in resected non-small cell lung cancer using a protein expression-based risk model: implications for personalized chemoprevention and therapy. *Clin Cancer Res*. 2014;20(7):1946-1954.

184. Pencina MJ, D'Agostino RB, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med*. 2008;27(2):157-72; discussion 207-12.

185. Pepe M, Janes H. Methods for Evaluating Prediction Performance of Biomarkers and Tests. In: Springer New York; 2013:107-142.

186. D'Agostino RB, Griffith JL, Schmid CH, Terrin N. Measures for evaluating model performance. In: *Proceedings of the American Statistical Association - Biometrics Section*. ; 1997:253-258.

187. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation*. 2007;115(7):928-935.

188. Zhang Y, Dai Y, Zheng T, Ma S. Risk Factors of Non-Hodgkin Lymphoma. *Expert Opin Med Diagn*. 2011;5(6):539-550.

189. Steyerberg EW, Eijkemans MJ, Harrell FE, Habbema JD. Prognostic modelling with

logistic regression analysis: a comparison of selection and estimation methods in small data sets. *Stat Med*. 2000;19(8):1059-1079.

190. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*. 2010;21(1):128-138.

191. Wang B. Variable selection in ROC regression. *Comput Math Methods Med*. 2013;2013:436493.

192. SAS Institute Inc. 54866 - Logistic model selection using area under curve (AUC) or R-square selection criteria. SAS Support.

193. Hosmer DW, Hosmer T, Le Cessie S, Lemeshow S. A comparison of goodness-of-fit tests for the logistic regression model. *Stat Med*. 1997;16(9):965-980.

194. Crowson CS, Atkinson EJ, Therneau TM. Assessing Calibration of Prognostic Risk Scores. *Stat Methods Med Res*. 2016;25(4):1692-1706.

195. Austin PC, Steyerberg EW. Graphical assessment of internal and external calibration of logistic regression models by using loess smoothers. *Stat Med*. 2014;33(3):517-535.

196. Steyerberg EW. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating.* Springer; 2009.

197. Kumar A, Abbas W, Herbein G. TNF and TNF receptor superfamily members in HIV infection: new cellular targets for therapy? *Mediators Inflamm*. 2013;2013:484378.

198. Ganna A, Reilly M, de Faire U, Pedersen N, Magnusson P, Ingelsson E. Risk prediction measures for case-cohort and nested case-control designs: an application to cardiovascular disease. *Am J Epidemiol*. 2012;175(7):715-724.