STUDY OF THE CLINICAL PRESENTATION AND DIAGNOSTIC TESTS OF INFECTIOUS MONONUCLEOSIS

by

XINYAN CAI

(Under the Direction of Mark Ebell)

ABSTRACT

Background: Infectious mononucleosis (IM) is a common disease most commonly caused by the Epstein-Barr virus (EBV), and is most often seen in adolescents and children. Ordering a serological test for infectious mononucleosis (IM) in all patients with a sore throat is costly and impractical. Clinical prediction rules (CPRs) for the diagnosis of IM that combines symptoms, signs, and hematologic parameters may improve the diagnosis of IM and help clinicians prioritize diagnostic testing. **Methods:** We performed a systematic review and meta-analysis of the accuracy of the clinical signs, symptoms, hematologic parameters, and serological tests in patients with suspected IM. The test threshold was estimated using a convenience sample of US primary care physicians. We then used the structured data extracted from the electronic health records of a university health center between 2015 and 2019 to develop and validate the CPRs. The CPRs for the diagnosis of IM were developed using four statistical methods: traditional logistic regressions, fast and frugal trees (FFTs), classification and regression trees (CARTs), and artificial neural networks (ANNs). The CPRs were developed based on the clinical symptoms and signs with (IM-Lab) and without hematologic parameters (IM-Nolab) and were internally validated. **Results:** Based on our systematic review, the most helpful hematologic parameters for ruling in IM include lymphocytes greater than

4×10^9/L, lymphocytes greater than 40% to 50%, or atypical lymphocytes greater than 40%. A combination of lymphocytes greater than 50% and atypical lymphocytes greater than 10% was also found to be helpful to rule in disease. Most of the individual clinical findings have limited diagnostic value in ruling out the disease when absent. We used data from clinical vignettes to the estimate a test threshold for IM of 9.5% (95% CI: 8.2% to 10.9%), we identified the probability of IM in the low- and high-risk groups as 8.8% and 31.2% for IM-Nolab logistic regression model (AUC=0.76); 4% and 79.4% for IM-Lab logistic regression model (AUC-0.94); 7.3% and 32.2% for IM-Nolab CART model (AUC=0.69); 5.9% and 61.8% for IM-Lab CART model (AUC=0.93); 8.2% and 33.5% for IM-Nolab FFT model (AUC=0.71); 5% and 68.2% for IM-Lab FFT model (AUC=0.94); 8.8% and 50.4% for IM-Nolab ANN model (AUC=0.70); and 4.4% and 69.3% for IM-Lab ANN model (AUC=0.97). The discrimination plots showed good discriminations for the IM-Nolab models and excellent discriminations for the IM-Lab models. The Calibration plots in the validation groups showed fair agreement between our predicted outcome and the observed test results for both IM-Nolab and IM-Lab models using all statistical methods. **Conclusion:** The derived IM-Lab and IM-Nolab models provided useful tools to help clinicians make a rapid diagnosis of IM. The IM-Nolab score has potential utility in telehealth visits, but the IM-Lab score provides a more accurate result. When externally validated, such risk scores would be useful for improving the diagnosis of IM and helping clinicians prioritize diagnostic testing.

INDEX WORDS: Infectious mononucleosis, Epstein-Barr virus, diagnostic testing, sore throat, clinical prediction rules, clinical presentation, college health

STUDY OF THE CLINICAL PRESENTATION AND DIAGNOSTIC TESTS OF INFECTIOUS MONONUCLEOSIS

by

XINYAN CAI

BA, East Tennessee State University, 2013

MSPH, University of South Carolina, 2017

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2021

STUDY OF THE CLINICAL PRESENTATION AND DIAGNOSTIC TESTS OF INFECTIOUS MONONUCLEOSIS

by

XINYAN CAI

Major Professor: Mark Ebell

Committee: Jose Cordero
Kevin Dobbin
Garth Russo

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
August 2021

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude and dedication to my major dissertation advisor and my mentor in my doctoral program, Dr. Mark Ebell. Thank you for your guidance and leadership throughout my doctoral student experience. Your advices and suggestions greatly enlightened me during my research, I cannot make it without your patient and insight support throughout my candidacy.

I have also greatly benefited from many scholars in my department, especially Dr. Cordero and Dr. Dobbin. Their inspirational comments and advices have taken my research to the next level. I am also thankful to Dr. Russo for his flexibility and willingness to provide me assistance at every stage of my research project.

My thanks extend to my fellow graduate students at the Department of Biostatistics and Epidemiology, we worked together in the Miller Hall for countless days and nights, and thanks for giving me someone to moan at when work wasn't progressing smoothly. I would also like to thank Lauren for her contribution to the projects, and her outstanding research work provides me a great help.

Finally, I would like to thank my beloved family. To my mom and dad, you instilled in me the appreciation of knowledge and consistently supported me in the ten years of overseas study from undergraduate to PhD. You always show your love and encouragement no matter when I needed. With your cultivation, I become a man with self-discipline and determination. I love you all. I am particularly grateful to my wife, Kailun, for her unconditional love accompanying me in the past and for years to come. I would not have been able to accomplish my degree without her making every day a joyful one to live. Thank you for raising me to believe in myself and being my soulmate in my life. I love you, immensely.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

Statement of the Problem

Infectious mononucleosis (IM) is a clinical syndrome that is primarily caused by Epstein-Barr virus (EBV). The EBV is a gamma herpes virus that can replicate in B-lymphocytes, as well as in a variety of epithelial cells.[1] EBV has been found in over 90% of humans and will persist lifelong after infection. The infection is primarily spread by saliva.[2] The heterophile antibody that agglutinates cells of animal species, as well as antibodies against viral capsid antigen IgG (VCA IgG), IgM (VCA IgM) and Epstein-Barr nuclear antigen (EBNA), are produced during acute EBV infection.[3] People infected with EBV in childhood normally experience asymptomatic or mild pharyngeal disease. For people infected with EBV after childhood, 30% to 45% are diagnosed with IM.[1]

The incidence of IM varies by age with the highest incidence among persons aged 10 to 19 years (6 to 8 per 1000 person-years). For a person younger than 10 years old or older than 30 years old, the incidence rate is less than one case per 1000 person-years.[4] IM is most common among young adults, especially among military personnel or college students, with the incidence rate ranging from 11 to 48 cases per 1000 persons per year.[5,6] A previous study[7] showed that the most common symptoms of IM, which include sore throat, lymph node enlargement, fever, and tonsillar enlargement, are shared by at least 98% of IM confirmed adolescents. The other common physical signs include pharyngeal inflammation (85%), transient palatal petechiae (50%), jaundice (10%), and rash (3%). Most symptoms were resolved within one month after the symptom onset; fatigue and hypersomnia were the most common symptoms that persisted over a 6 month period.[8]

Limited studies have been conducted to determine the accuracy of the clinical examination for the diagnosis of IM. A previous study using the heterophile antibody test as the reference standard found that the presence of splenomegaly, posterior auricular adenopathy, axillary adenopathy, palatine petechiae, and pharyngeal exudate are more useful to increase the likelihood of IM, while the absence of pharyngeal erythema and fatigue decreased the likelihood of IM.[9] However, this finding is not consistent with other studies. Although the presence of individual symptoms and signs was associated with IM illness, in isolation they were not accurate. The white blood cell count (WBC) and the differential WBC count (the absolute number and the percentage of neutrophils, lymphocytes, monocytes, eosinophils, and basophils present in the blood) are useful hematologic parameters to confirm the diagnosis for people with suspected IM. An increase in the absolute number of lymphocytes, percent of lymphocytes, and atypical lymphocytes all increase the likelihood of IM.[7]

Several serologic tests have been used for the diagnosis of IM. The Paul-Bunnell test, which is the original serologic test, is used to detect the heterophile antibodies by agglutination of sheep or horse cells.[10] Davidson's test added guinea pig kidney absorption to increase the specificity of the test.[11] The heterophile tests are in the form of a solid-phase immunoassay or latex agglutination. The heterophile test is less sensitive in the first week of illness. Using the microagglutination slide test, only 70% of adolescents became positive during the first week of illness, and 95% became positive after three weeks of illness.[7] Children below twelve years of age are positive in less than 50% of infections after the first week of disease onset, compared with detecting more than 70% of infections in older patients.[12]

Tests for the detection of VCA-IgG and VCA-IgM were developed after the heterophile test. The VCA tests (sensitivity: 97%, specificity: 94%, positive likelihood ratio (LR+): 64, negative likelihood ratio (LR-): 0.03) are overall more accurate than the heterophile antibody test (sensitivity: 83% to 87%, specificity: 91% to 97%, LR+: 52 to 75, LR-: 0.14 to 0.18).[13,14] The VCA-IgM appears in the early phase of EBV infection, and since it can be detected earlier than

VCA-IgG, the isolated VCA IgM is always treated as the indicator of early stage of acute infection. However, since the VCA-IgM may disappear over a short time or present with low titers, it may not be detected with traditional tests.[3] The test to detect EBNA is also highly sensitive but may only be detectable after six weeks; thus, it might not be an appropriate test to detect acute infection of IM, but combined with VCA-IgG and VCA-IgM test, it can help to distinguish the patients with previous infection and those with acute infection.[14] However, the VCA-IgG and VCA-IgG tests are expensive.[3]

The recommended diagnostic strategy is to obtain a white blood cell count with the differential hematologic pattern and to use the heterophile test among patients aged 10 to 30 years who are suspected of having IM. Patients with a positive heterophile test or with more than 50% of lymphocytes and at least 10% atypical lymphocytes are highly likely to have IM, and no further test is suggested. If the test is negative, a further confirmation test should be considered using a second heterophile test at a later time or a VCA-IgM test.[14]

While IM is a common condition among adolescents and young adult, there are still limited well-designed prospective studies that address the natural history of IM, as well as examine the factors that predict the duration of the illness. There have been no studies to develop and validate a clinical prediction rule that combines symptoms, signs, and hematologic parameters to improve diagnosis of IM in patients with suspected IM. Also, the decision thresholds for IM testing have not been studied or determined. The test threshold for IM is the “tipping point” where physicians make the decision to either rule out IM or order additional diagnostic tests. Once the test threshold for IM has been decided, a clinical prediction rule can be designed to classify patients as low-risk or high-risk for IM in order to help clinicians to make decisions on diagnostic testing.

Further research to understand the clinical presentation and the diagnostic tests of IM among adolescents and young adults is important for patients to avoid inappropriate antibiotic use. It would also help physicians to order appropriate diagnostic test and give better advice on

the restriction of physical activities.[14] College students live in a more communal living condition, and they are more likely to become infected than non-students with the same age.[5] The University Health Center (UHC) provides comprehensive healthcare for the students of the University of Georgia (UGA). UHC has electronic health records that maintain the details of the presence or absence of symptoms and signs, and the results of laboratory testing at every visit. In this thesis, I will access anonymized data of all UGA students with a serological test for mononucleosis in the past five years, along with their age, sex, vital signs, signs, symptoms, and blood test results, as well as any mono test results.

Aims of the Dissertation

1. To evaluate the accuracy of signs, symptoms, hematologic parameters, and the serological diagnostic tests for the diagnosis of IM through systematic review and meta-analysis.

2. To determine the decision thresholds for ordering diagnostic tests for IM based on the cross-sectional survey of physicians.

3. To identify the clinical signs and symptoms that predict IM using the data collected from the UHC. This information will be used to develop and internally validate prediction models for the diagnosis of IM by applying logistic regression analysis, classification and regression tree (CART), fast and frugal tree (FFT) and artificial neutral network (ANN) analysis. We will also assess whether adding hematologic parameters to the models can improve the diagnosis of IM in these models.

To accomplish the first aim, we will perform two systematic review and meta-analyses. First, we will review the literature to identify prospective studies or case series evaluating the accuracy of signs, symptoms, and hematologic parameters in patients with suspected with IM. Second, we will systematically review to identify the prospective or retrospective cohort studies

evaluating the accuracy of the EBV immunoassays and the heterophile antibody tests among patients with suspected EBV infections to determine the most accurate test to detect primary EBV infection. We will perform data abstraction, evaluate study quality, and calculate the summary estimations by meta-analysis. To accomplish the second aim, we will send online surveys to a convenience sample of US primary care clinicians and ask their medical specialty, years in practice, and practice site. We will present them with seven clinical scenarios that have different risks of IM. For each scenario, we will ask them if they will order a test for this patient, and we will use this information to determine their decision threshold for testing. To accomplish the third aims, we will obtain the de-identified data from electronic health records at the UGA UHC for patients who are clinically suspected of having IM in the past five years, based on the fact that a diagnostic test for IM was ordered. We will abstract data regarding demographics, signs, symptoms, and laboratory results, including results of the hematologic parameters and any test for IM. The UHC staff will be responsible for the data collection, linkage of the clinical and laboratory data, and removing the identifier's information. We will create the clinical prediction rules using logistic regression, as well as apply modern machine learning techniques that include classification and regression tree, fast and frugal tree, and artificial neural network analysis to build prediction models with and without laboratory parameters.

We chose university students at UGA UHC for the following reasons. First, IM is very common among young adults, especially among university students. Second, limited studies have examined the natural history of IM among college students and addressed the students' risk factors associated with prolonged duration of IM symptoms.[6] Third, UHC has an excellent electronic health record (EHR), which records patients' symptoms, clinical signs, medical history, and laboratory results. This sophisticated system provided a capable source for the development and validation of the clinical prediction rules.

Objectives of Aims

Each aim has its own objectives and methods, which are briefly described in Table 1.1.

Dissertation Outline

Chapter one has provided an abstract of our study, which gives us a brief overview of IM, study objectives, a summary of the study methods and the data collection. In Chapter 2, we will expand the background of our study in detail and will highlight the gaps in the literature. In Chapter 3, we will describe the methods of each specific aim, including the description of study design and settings, study population, data collection, and statistical analysis. Chapters 4 through 8 will be written manuscripts to be submitted for publication. Each manuscript will be traditionally formatted, with an introduction, methods, results, and discussion. Therefore, the manuscripts may contain duplicated information from earlier chapters. Chapter 9 will summarize the findings of this dissertation and will provide suggestions for future studies.

Chapter 1 Tables

Table 1.1 Brief Description of the objective, method, and data sources in each aim

Aim	Objectives	Data Sources	Methods
1	Objective 1: To evaluate the diagnostic accuracy of the signs, symptoms, and hematologic parameter of IM through systematic review and meta-analysis. Objective 2: To determine the accuracy of EBV immunoassays and the heterophile antibody tests among patients with suspected EBV infection that used immunofluorescent test (IFA) as the reference standard.	-Medline database using the PubMed	1. Define the inclusion criteria and the search strategy, and search literature from the database by three investigators using a shared Google Document and Dropbox folder. 2. Abstract data that describe study characteristics, study quality, and predictive accuracy 3. Use the QUADAS-2 tool to assess the study quality. 4. Similar individual predictors and predictor with similar cutoffs will be grouped together 5. Calculate summary measures of accuracy, which includes sensitivity, specificity, likelihood ratios, the diagnostic odds ratios, and the area under the receiver operating characteristic curve (AUC). We will also create summary receiver operating characteristic (ROC) curves for each predictor.
2	To determine the decision thresholds for ordering diagnostic tests for IM	-A cross-sectional survey, including years in practice, practice sites and specialty of physicians; the survey also includes the assessment of the preferred management option (order test for IM or not) in 7 scenarios for patients with a sore throat.	1. Send online survey to inform the potential participants about the instruction of the study. 2. Collect the surveys from participants after CME courses. 3. Obtain the frequency of clinical settings, specialty and years in practice. 4. Perform the mixed effect logistic regression model to determine the physicians' clinical decision.

			5. Perform the stratified logistic regression analysis by years of practice, specialty, and practice site to determine the physician decision with respect to different categories.
3	Objective 1: To identify the clinical signs and symptoms that predict IM with and without hematologic parameters using the data collected from the UHC. This information will be used to develop and internally validate prediction models for the diagnosis of IM by applying logistic regression models. Objective 2: To develop and internally validate decision support tools using three innovate statistical methods (CART, FFT, ANN) for assessing the likelihood of infectious mononucleosis (IM) among university students using symptoms, clinical signs, and hematologic parameters.	-Demographic information, symptoms, signs, laboratory data, hematologic parameters, results of the heterophile antibody test from UHC electronic health record	1. Univariate analysis to select variables. 2. Perform logistic regression using derivation cohort with signs and symptoms as independent variables to build the parsimonious model that has good predictive accuracy for IM 3. Perform logistic regression using derivation cohort with clinical signs, symptoms, and adding hematologic parameters as independent variables to build the parsimonious model. 4. Identify patients with low, and high-risk of IM for both models. 5. Create a simplified point score to weight each predictor. 6. Validate the score using the validation cohort by assessment of AUC, the calibration plots, and predictive accuracy based on percentage with IM in each risk category for both models. 7. Using CART model, FFT model, and ANN model to predict IM using derivation set with symptoms and signs as predictors, then included the hematologic parameters as the independent variables. 8. Assess the performance of each of above models using the validation cohort by the assessment of AUC, the ROC and calibration plots.

CHAPTER 2

LITERATURE REVIEW

Pathophysiology of Epstein-Barr virus

IM was first described in the 1920s and is colloquially known as “the kissing disease”. It is a common viral infection that causes sore throat, fatigue and lymphadenopathy.[15] IM can be caused by numbers of viruses, including Epstein-Barr virus (EBV), human cytomegalovirus (CMV) and toxoplasma gondii.[16] EBV was discovered in 1964 by electron microscopy of tumor cells from Burkitt’s lymphoma by Epstein, Achong, and Barr. In 1968, EBV was confirmed to be the etiologic agent of heterophile-positive IM. It infects more than 90 percent of humans and persists for the lifetime of the infected person.

EBV is a gamma herpes virus. The EBV genome is encased within a nucleocapsid, and it consists of a linear DNA molecule that encodes almost 100 viral proteins. During viral replication, these proteins are essential for regulating the expression of viral genes, replicating viral DNA, forming structural parts of the virion, and modulating the host immune response.[2] The replication of EBV has a well-known tropism for B-lymphocytes. The B-lymphocytes are infected after the primary envelope glycoprotein, gp350, connects on the surface of the B cell to the viral receptor, the CD21 molecule. The major histocompatibility-complex (MHC) class II molecule also acts as a cofactor for the infection of B cells. B cell infection also causes a latent infection.[17] The infection is characterized by an outpouring of polyclonal IgM immunoglobulins, which include antibodies against other tissues. Among healthy adults, 1 to 50 B cells per million in the circulation are infected with EBV. Once infected by EBV, the individual will carry the virus in latent form for life.[2] Clinically identifiable reactivation of latent EBV infection is rare. EBV nuclear antigen (EBNA) and latent membrane proteins (LMP) are found in infected cells during latent

infection.[3] The EBV replicates not only in B-lymphocytes but also actively replicates in epithelial cells from the pharynx and parotid duct along with the epidermal skin in immunocompromised hosts.[18]

Infection with EBV results in both humoral and cellular immunity. The humoral immunity includes the antibodies against antigens. The antibodies against the virus capsid antigen IgG (VCA IgG) and IgM (VCA IgM) are important in preventing the spread of acute infection. The VCA-IgG and VCA-IgM both appear at the time of clinical onset of the symptoms; the VCA-IgG will remain positive for life, whereas VCA-IgM will disappear after several weeks.[3] The cellular immune response is important for the control of EBV infection, although the findings of antibodies directed against LMP and the EBNA are also potentially useful for diagnosing the infection. During the primary infection, natural killer cells and CD4+ and CD8+ cytotoxic T cells are activated and help control the growth of EBV-infected B cells.[17] The ability of EBV to persist indicates that the EBV has evolved strategies to elude the immune system.[2]

EBV is primarily spread by contacting an infected person's saliva. The virus replicates in cells in the oropharynx, and all seropositive persons actively shed virus in the saliva. The B cells are subsequently infected after contact with epithelial cells in the oropharynx. The incubation period of the infection is between four and eight weeks.[19] Besides deep kissing, the primary EBV infection is also transmitted through blood.[20] Based on the National Health and Nutrient Examination Survey (NHANES) cycles 2009 to 2010, the prevalence of EBV antibody among the US population aged between 6 and 19 years old is between 50% and 85%, with the highest prevalence of VCA IgG antibodies among adolescents aged 18 to 19 years (85%).[21] Over 90% of young adults worldwide are infected with EBV based on seroepidemiologic surveys.[22] Because of improved economic and sanitary conditions, early childhood infection became uncommon over the past decades. Currently, most EBV infections occur during adolescence and young adulthood.[23] It is unclear how young children contract EBV; a reasonable supposition

is that they are infected by their parents or siblings who are infected with EBV. Parents may intermittently shed it in their oral secretions.[24]

EBV can cause a variety of clinical manifestations. Apart from IM and atypical primary EBV infection, such as meningoencephalitis or hepatitis, EBV has also been implicated in a range of malignancies, which include Burkitt's lymphoma, nasopharyngeal carcinoma, gastric carcinoma and malignant lymphoepithelial.[18] EBV causes approximately 90% of IM cases. This dissertation will focus on IM illness due to the EBV infection.

Epidemiology of IM

Incidence of IM

Studies have determined that 75% of young adults between the ages of 18 and 22 develop typical IM after primary EBV infection.[21] It remains unclear why primary EBV infection leads to IM in adolescents but not younger children, the possible explanation is that IM in adolescents may reflect the response of cross-reactive memory CD8 T cells. For example, adolescents are more likely to have high numbers of influenza-specific CD8+ T cells, which could react more strongly with EBV.[25] However, it is still debatable whether pre-existing CD8+ T cell immunity to EBV would increase or decrease IM.

One previous study[4] shows that the incidence of IM is highest in persons 15 to 19 years of age (the annual incidence rate for people across different age ranges from 995 to 1846 cases per 100,000 persons per year). The incidence of IM for persons younger than 5 years and older than 30 years is less than 100 cases per 100,000 persons per year (49 cases per 100,000 persons per year in persons aged 0-4 years, 64 cases per 100,000 persons per year in persons aged 30-34 years, 7 cases per 100,000 persons per year in persons aged above 35 years). Mild infection among younger children is often undiagnosed.[14] The age-specific incidence of IM is much greater for high school and college-age compared to elementary school and post-college age persons.[4]

The incidence rate of IM reported from studies in different locations varies, with the estimated incidence rate ranging from 11 to 121 per 10,000 population per year (Table 2.1).[4] The overall incidence of IM in the United States is approximately 500 cases per 100,000 person-years with the highest incidence between the ages of 15 and 24 years. The infection is common in communal living situations, especially among active-duty military and college students, with annual incidence rates ranging from 11 to 48 cases per 1,000 population per year based on three studies of student and military populations.[26-28] A prospective cohort study[29] on 241 EBV seronegative students enrolled at Edinburgh University found that after 3 years of follow up, 110 (46%) students were infected with EBV, and 27 (25%) of them have diagnosed with IM. A prospective study[30] on 143 students without EBV infection at baseline from the University of Minnesota found that after 4 years of observation, 46% subjects experienced EBV infection. Of these, 77% had infectious mononucleosis. So far, no seasonal pattern or annual cycles have been found with regards to the incidence of IM.[31] There is no evidence that having IM is different between males and females.[17]

Patients with IM have an increase in the number of lymphocytes and their immature form, atypical lymphocytes. The atypical lymphocytes are a type of WBC that are involved in the immune response and are responding to EBV-infected B cells. Most symptoms caused by IM are due to the proliferation and activation of T cells and B cells in response to the infections.[2] Activation of B cells by EBV leads to the production of polyclonal antibodies and causes elevated titers of EBV-specific antibodies.[2]

Typical Presentation

The clinical presentations of IM are summarized in Table 2.2. Based on the results from different studies, the most common symptoms and signs related to IM include sore throat, fever, pharyngitis, and cervical adenopathy. Other symptoms and signs of people with IM include rash, headache, nausea, jaundice, fatigue, axillary or inguinal adenopathy, muscle or joint pain, splenomegaly, and hepatomegaly. The most recent study[32] of 150 confirmed cases concluded

that the leading symptoms and signs of IM are cervical adenopathy (77%), fatigue (77%), sore throat (74%), and pharyngitis (73%). The most uncommon complaints are axillary adenopathy (5%), inguinal adenopathy (5%), or hepatomegaly (7%). However, a study from 1951[33] of 210 IM cases reported that the complaints of axillary adenopathy (46%), inguinal adenopathy (38%) and splenomegaly (43%) are commonly seen, which is inconsistent with the previously described study. This may be due to the reference standard test selection or different inclusion criteria of the study population among studies.

The signs and symptoms of IM vary by age. Children less than 4 years old with IM are more likely to have rashes, abdominal pain, and neurologic problems than adolescents and older adults.[24] Among adolescents, the most common symptoms of IM include headache, sore throat and fatigue, and they were less likely to have diarrhea, rashes and jaundice.[34] The most frequent symptoms for adults include fatigue and sore throat. They were less likely to have myalgia or arthralgia.[35] Older people are more likely to have fatigue and body pain, and rarely have sore throat.[36] Acute infection of EBV for young children usually is either asymptomatic or expresses with mild upper respiratory tract symptom, while for those older patients (aged greater than 40 years), pharyngitis and cervical lymphadenopathy are less frequent; instead myalgias are prominent.[37,38]

Natural history of IM

The acute phase of IM lasts between 5 and 15 days, followed by a period of recovery. Most patients will completely recover between 6 and 8 weeks after the onset of illness.[39] One prospective study[32] examined the natural history of IM by following 150 patients for 6 months. According to this study, sore throat and fatigue were the most common symptoms and occurred in about three-quarters of the patients. Most symptoms, such as sore throat, fever, headache, rash, nausea, resolved within one month, although 28% still had fatigue and 16% had myalgia. There is no significant improvement for the recovery at the second-month visit compared with the first-month visit. At the second-month visit, the common complaints include fatigue (21%),

followed by sleeping too much (18%), and headache (15%). At six months, most symptoms associated with IM were reported by less than 10% of patients, except for fatigue (13%) and sore throat (11%). The study found that it took up to six months to achieve a stable level of functional recovery. Recurrences of IM and chronic mononucleosis are relatively rare.[39] Dalrymple[40] conducted a cohort study on 131 patients with IM and found that 90% of patients recovered completely within six weeks, and even fatigue had resolved. This study found that bed rest didn't shorten the duration of the symptoms.

Diagnosis

Clinical Features of IM

There are limited published, prospective studies determining the value of clinical examination among adults with IM.[14] One prospective study[34] described the clinical features of IM among 590 clinical suspected patients, 330 of whom were confirmed to have IM based on a positive EBV IgM serology and presence of heterophile antibodies. The diagnostic accuracy of symptoms and signs related to IM are summarized in table 2.3. The presence of splenomegaly, hepatomegaly, rashes, and lymphadenopathy are most helpful to increase the likelihood of IM, while the absence of headache and sore throat are most helpful to reduce the likelihood. Most patients with IM will not have diarrhea (sensitivity: 15%) or jaundice (sensitivity: 17%). IM should be suspected and evaluated in patients with sore throat, headache, and lymphadenopathy because these are the most sensitive findings, especially if they are adolescents or young adults.

One study[9] included more than 700 patients aged above 16 years with a sore throat, 15 of whom were confirmed to have IM based on a heterophile antibody test. The accuracy of symptoms and signs in this study are summarized in table 2.4. The presence of pharyngeal erythema (positive LR: 16.7, negative LR: 0.05), loss of energy (positive LR: 4.0, negative LR: 0.09), anterior cervical adenopathy (positive LR: 1.63, negative LR: 0.53), and swollen glands

(positive LR: 1.5, negative LR: 0.31) are more useful to increase the likelihood of IM, and the absence of these symptoms and signs are more likely to reduce the likelihood of IM.

The diagnostic accuracy of symptoms and signs of IM is inconsistent between studies. A possible explanation for these inconsistencies is that the reference standard for the confirmation of EBV infection varies between studies. Also, the proportions of different age groups in the sample population are different between studies.[34] So far, there is limited value of a diagnosis of IM solely based on an individual signs or symptoms.[35] To date, no published studies have attempted to develop a clinical prediction rule for the diagnosis of IM using a combination of symptoms and signs.

Accuracy of hematological parameters

The diagnosis of IM is based on clinical, serologic, and hematological findings.[41] The complete blood count (CBC) and leukocyte differential count information are some of the most easily obtained hematologic parameters. These parameters help diagnose IM for clinically suspected patients.[35]

CBC and leukocyte differential counts are two standard clinical laboratory tests in medical practice.[42] CBC uses a peripheral blood sample and measures white blood cell (WBC), red blood cell (RBC) counts, mean cellular volume, mean cellular hemoglobin concentration, and hematocrit. A platelet count is not included in the CBC information. The leukocyte differential count is measured by an automated test, which gives the percentages of leukocytes that are mature neutrophils, lymphocytes, monocytes, eosinophils, basophils, or other cell types.[43] The WBC gives the absolute number of lymphocytes and confirms the suspected infection, while the leukocyte differential count may help identify the type of pathogen, for example bacterial or viral. Increases in the absolute and relative lymphocytes, as well as the presence of atypical lymphocytes, all increase the likelihood of IM.[7]

Hoagland developed a standard criteria for diagnosing IM, which includes at least 50% of leukocytes being lymphocytes, combined with greater than 10% of atypical lymphocytes and

confirmation of a heterophile antibody test.[41] One prospective cohort study[44] compared two automated hematology analyzers, the Coulter analyzer and Sysmex analyzer, and used a positive heterophile antibody test as the reference standard to identify the sensitivity and specificity of different spectrums of atypical lymphocytes in outpatient populations. The conclusion from the study also verified the overall accuracy of Hoagland's morphologic criteria. The result of the diagnostic accuracy across the different spectrums of atypical lymphocytes and lymphocytes are shown in table 2.4. The different test results showed that sensitivity would decrease and specificity would increase as a higher percentage is used to define an abnormal number of atypical lymphocytes.[35] This study also reported that a high percentage of lymphocytes combined with atypical lymphocytes increased the likelihood of mononucleosis more than each individually (LR+ is 12 for patients with ≥50% lymphocytes and ≥10% atypical lymphocytes versus LR+ is 4.1 for patients with ≥50% lymphocytes, or LR+ is 9.4 for patients with ≥10% atypical lymphocytes).

Another prospective cohort study considered the accuracy of the percentage of lymphocytes and atypical lymphocytes among 590 young adults with clinically suspected IM.[34] The reference standard for this study is a positive EBV IgM serology confirmed by a heterophile antibody test. The results show that the sensitivity and the specificity of using a cutoff of greater than 50% lymphocytes as the criteria are 38% and 82%, respectively. The sensitivity and specificity of the test using atypical lymphocytes greater than 10% as the criteria is 65.6% and 71%, respectively. This study also evaluated the accuracy of WBC counts greater than 10,000/ml (sensitivity: 46.2%, specificity: 77.2%).

Monocytosis is also one of the biomarkers for the diagnosis of IM. Monocytes are WBC produced in macrophages and dendritic cells in the immune system. Monocytosis is clinically defined as approximately greater than $1x10^9$/L monocytes circulating in the blood.[45] An earlier study compared the accuracy of using lymphocytosis with using monocytosis for the diagnosis of IM among adults with suspected IM, with the heterophile antibody test as the reference

standard. Based on the result, both monocytosis and lymphocytosis help diagnose IM in adults or children. However, lymphocytosis defined as lymphocyte counts greater than 4 x 10^9/L (sensitivity: 84%, specificity: 94%, positive likelihood ratio: 14, negative likelihood ratio: 0.17) is more accurate to detect IM, than monocytosis defined as monocyte count greater than 1 x 10^9/L (sensitivity: 72%, specificity: 89%, positive likelihood ratio: 6.5, negative likelihood ratio: 0.31).[46]

A systematic review[35] evaluated the accuracy of the individual value of the symptoms and signs, as well as WBC information and differentials to diagnose IM. This study suggests that the presence of sore throat, splenomegaly, posterior cervical, inguinal or axillary adenopathy, palatine petechiae, and atypical lymphocytes are better able to rule in IM. So far, there have been clinical prediction rules developed for strep throat, pneumonia, influenza, and other infections to aid in diagnosis.[47-50] Although IM is a common condition, there are limited well-designed prospective studies to address the diagnosis of IM, and there has been no published study attempting to develop and validate a clinical prediction rule that combines symptoms and signs with hematological parameters to help clinicians make an early diagnosis during the course of the disease.[35]

Heterophile antibody test

In addition to the hematological criteria, other laboratory tests have been used to diagnose IM. A reliable, accurate and rapid serologic test is critical to differentiate between IM and other infectious or non-infectious diseases, such as hematological malignancies, that could also cause pharyngitis or lymphocytosis.[51] During the acute phase of IM, the heterophile antibodies are produced that cross-react in the form of sheep cell agglutinins during active stages of the infection, which lead to direct immunologic response to the IM.[11] The heterophile antibody tests relies on the agglutination of animal red cells by heterophile antibodies in human serum. The Paul-Bunnell test is the heterophile test used to detect heterophile antibodies through the characteristic properties of antibodies in IM by agglutination of sheep RBC with

titers of 1:14 or more in unabsorbed sera.[17] The Paul-Bunnell test is available in a convenient latex agglutination form.[52]

The Paul-Bunnell test was then modified by adding guinea pig kidney absorption of serum. The interference of Forssmann antibodies can be easily detected and prevented by adding guinea pig kidney emulsion in the test.[53] This test is available in solid-phase immunoassay form.[11] Approximately 76% of the adolescents with IM had a positive heterophile antibody test during the first week of symptoms, followed by 93.5% positive for the second week, and 90% after the third week.[7]

One prospective study[12] compared nine commercially available heterophile antibody detection kits for diagnosis of IM. Five of the kits (i.m. absorption, Monospot, Monosticon, Mono-Plus, and Monolatex) are in the form of slide agglutination assays; the rest of the kits (Cards Mono, Cards OS Mono, Preview Mono, and Monolert) are solid-phase immunoassays. The study collected 108 blood samples from 103 patients aged from 2 to 60 years. The reference standard test was the EBV specific immunofluorescence assay for the determination of antibodies to VCA and enzyme immunoassay for the determination of Epstein-Barr nuclear antigen (EBNA). The sensitivity, specificity, and likelihood ratios of latex agglutination and solid-phase immunoassay tests for diagnosing IM between different age groups are summarized in table 2.6. Based on the result, the heterophile antibody test is more sensitive in patients older than 13 years, detecting 79% to 91% of infections in latex agglutination and 71% to 79% in solid-phase immunoassays, compared to patients younger than 12 years, detecting only 38% to 50% of infections in latex agglutination and 25% to 50% in solid-phase immunoassays. The overall sensitivities ranged from 71% to 84% for slide agglutination kits and ranged from 63% to 71% for solid-phase immunoassay kits. The specificities for slide agglutination kits ranged 84% to 98% compared to the reference method and ranged 95% to 100% for solid-phase immunoassays.[12]

Another prospective cohort study[54] evaluated six commercially available kits from 100 blood samples of the participants. The reference standard was the VCA-IgM test with the EBNA test. Three commercial rapid IM kits (Monolatex, Mono-Lex, and Mono-Latex) were latex agglutination slides and the others (Mono-Plus, IM-Check, and Clearview IM) were solid-phase assays. Compared with the reference standard test, the sensitivity of latex agglutination form ranged from 85% to 91%, and the specificity ranged from 95% to 100%; the sensitivity of solid-phase assays ranged from 70% to 92%, and specificity ranged from 98% to 100%. Therefore, compared with solid-phase assays, latex agglutination slides are more sensitive, and therefore are better able to rule out disease when test results are negative. There is no significant difference in the ability of ruling-in disease between the two methods when the test is positive.[54]

Heterophile antibody test is rapid, cost-effective and simple compared to other serologic tests. The heterophile antibodies can be detected within one week after onset of IM. However, these tests are less sensitive than VCA tests, especially in children that are EBV infected.[24] Therefore, other tests should be considered in order to confirm children with EBV infection or for adults with severe infection.

Anti-EBV immunoassays

The VCA-IgM, VCA-IgG, and EBNA tests were developed for the detection of EBV-specific antibody responses. They are commonly used for routine screening by indirect immunofluorescence assay (IFA) or enzyme-linked immunosorbent assay (ELISA).[3] At present, only a few rapid test kits can detect an EBV-specific antibody response.

One case-control study[55] compared 12 commercial kits of the EBV-specific antibody tests and heterophile antibody tests among five different subgroups. These groups include EBV infected patients, immunocompromised patients with recent CMV infection, healthy persons, and EBV seronegative patients.[55] The total sample size consisted of 248 sera for EBV-specific antibody tests evaluation and 241 sera for heterophile antibody tests evaluation. The EBV specific antibody tests include the tests detecting VCA and EBNA antibodies, detecting VCA

and early antigen (EA) antibodies, and detecting the combination of VCA, EA, and EBNA antibodies. The commercial kits of the heterophile antibody tests include Paul-Bunnel-Davidsohn, Monosticon Dri-Dot, Avitex-IM, Clearview IM, and Cards+/-OS Mono. The results of the study are shown in Table 2.7.[55] The EBV specific antibody tests were more likely to reduce the likelihood IM compared with heterophile antibody tests when the result was negative (negative likelihood ratio, 0.03 versus 0.12). However, when the results were positive, the heterophile antibody tests had a better ability to rule-in disease compared to EBV specific antibody tests (positive likelihood ratio, 88 versus 16). The previous evaluation showed the test to detect the combination of VCA IgM and VCA IgG has both high sensitivity and specificity, but the EBNA IgG test had a low sensitivity.[56] This study also proposed that the combination of VCA-IgM, VCA-IgG, and EBNA IgG tests were the most optimal test kit to diagnose IM. Since this study was a case-control study, where the cases and controls were pre-selected during the phase of enrolling patients, thus, selection bias might exist due to the selection criteria of the control group (e.g. selected controls from patients with previous EBV infection or with cytomegalovirus infection).

The VCA-IgM tests are used to detect the early stage of acute infection since the antibody of VCA-IgM can be detected earlier than VCA-IgG.[57] However, the VCA-IgM antibody may not be produced at all or only for a short time; therefore, the results of the tests for this antibody may be false positive. Such cases may require further diagnosis, including immunoblotting for VCA-IgG. This method is commonly used to isolate VCA-IgG in order to interpret the result, and can distinguish the acute infection from the past infection.[3] EBNA antibodies usually cannot be detected until six to eight weeks after the onset of clinical symptoms and is used to distinguish between acute and previous infection. The presence of VCA-IgG and VCA-IgM with the absence of EBNA indicates acute infection; and the presence of VCA-IgG and EBNA with the absence of VCA-IgM indicates previous infection.[3] The VCA-IgG

and VCA-IgM tests are more useful among IM suspected patients whose heterophile test results are negative.[14]

Previous diagnose strategy

So far, there are no evidence-based guidelines for the diagnosis of IM among patients with suspected IM. The previous case-control study[13] offered the following recommendation based on the findings: 1. Initially perform a heterophile test, followed by a WBC combined with the differential hematologic test. 2. If the situation is not urgent, repeat the heterophile test and hematologic test in the second week; if the situation is urgent, the EBV-IgG and EBV-IgM antibody tests should be obtained in the first week. 3. If none of IgG and IgM antibodies are detected from the first testing, then repeat the EBV-specific antibody test in the second week.

A systematic review[35] suggested that patients aged 10 to 30 years with a sore throat, fever, and the clinical signs of anterior cervical adenopathy, posterior cervical adenopathy, inguinal adenopathy, palatal petechiae, or splenomegaly are at high risk of IM. Those patients should be confirmed based on a WBC combined with the differential hematologic test. The heterophile antibody test and rapid test are also recommended for those patients with suspected IM. If the hematologic test shows that symptomatic patients have greater than 50% lymphocytosis and at least 10% atypical lymphocytosis, then IM is highly suspected and there is no need for an additional antibody test. A positive heterophile antibody test result among symptomatic patient is also considered strong evidence of having of IM; however, patients with CMV or toxoplasmosis may produce false negative results. The review also recommended the VCA-IgM test for an accurate diagnosis, since a negative result of that test is strong evidence against IM.[35]

Clinical prediction rules

Overview

Clinicians need to make a diagnosis and to make choices on treatment, as well as to advise patients regarding prognosis. These tasks involve judgment on the likelihood of diseases, the likelihood of treatment effectiveness, and the likelihood that the patient will experience the clinical outcome. To accomplish these tasks, clinicians are increasingly relying on evidence-based practices, including clinical prediction rules (CPR).[58]

CPRs can be defined as a guideline that using the best combination of the patient's medical signs, symptoms, and other clinical findings to predict the probability of a disease.[58] The CPRs may provide evidence of the treatment and further diagnostic testing based on the probability of the disease, and can be of great value to assist clinical decision making. The CPRs aid in selecting appropriate interventions and diagnosing certain diseases. If the probability of a disease is above the treatment threshold, treatment may be initiated; if the probability of a disease is below the test threshold, the disease may be ruling out; if the probability of a disease is between the test and treatment threshold, no treatment is indicated but more information is needed. In public health, various CPRs have been developed to predict the future occurrence of diseases among asymptomatic individuals.[58]

The benefit of a CPR is that when limited resources are available, a CPR can separate patients with a high risk from the low-risk group, which enables the clinician to target the patients with the highest expected benefit. Another benefit is that CPR can guide use of diagnostic tests. The clinicians can use CPR to distinguish low-risk patients who may not need a diagnostic test from the higher-risk patients who do. CPR can also increase the accuracy of the diagnostic or prognostic assessment, especially in the situation when the test or treatment is risky, or to reduce unnecessary cost while maintaining the quality of patient care.[59]

Derivation

CPRs are developed with the following steps: 1. identify the potential clinical predictors, 2. assess the presence or absence of clinical predictors and determine the outcome of interest (diagnosis, prognosis, treatment success) among the patient population, 3. conduct the statistical analysis.[59] The development of CPRs starts with the identification of potential clinical predictors. The consideration of potential clinical predictors relies on an extensive literature review and etiological knowledge of the disease. The potential clinical predictors include medical history, physical examination, and basic laboratory tests. The susceptible patient population is then defined. Each patient is then assessed for the presence and absence of the clinical predictors and whether or not they have the outcome of interest. This patient population is also referred to as the derivation or training population, because it will be used to train the prediction model and to predict the clinical outcome. Statistical analysis reveals which predictors are statistically associated with the clinical outcome and which predictors can be excluded from the rule without loss of predictive strength. In most of the cases, the statistical techniques used in this process are based on logistic regression analysis. Other techniques include fast and frugal tree analysis, discriminant analysis, and neural networks.[59]

Validation

CPRs should not be considered for clinical application if they are not validated. There are two reasons for this. First, the association between predictors and outcomes in a CPR may be due to chance, and the association may change in other populations. Second, the predictors may be idiosyncratic to the clinicians creating the rule, or only applicable to the design of individual studies, and the rule may fail in other settings.[59]

During the validation process, we aim for at least internal validation, in which the researchers determine the internal validity of the predictive model, and the predictions are valid for subjects from the underlying population where the development data originates from. The techniques to assess internal validity are showed in Table 2.8.[60] Each technique has its

advantages and drawbacks. The split-sample method has comparatively small samples for validation, which leads to an unstable model performance assessment. Another problem is that with internal validation only, the validation and derivation samples came from the same source of population, and the findings might not be generalizable to other population with different settings. Cross-validation is an extension of split-sample validation and can use a larger part of the sample for model development. However, the cross-validation procedure may need to be repeated ten or more times to obtain truly stable results. Bootstrap validation draws samples with replacement from the original sample. The performance of bootstrap validation is rather stable since the same sample is used to develop the model as well as test the model. The disadvantage is that only automated modeling strategies can be used during the bootstrap validation process. For example, fitting a full model without selection, or an automated model selection procedure. It may be difficult to repeat the intermediate steps, such as visually assessing linearity, collapsing categories of variables, or studying univariate and multivariate p values.[60] The validity of a CPR is assessed via classification accuracy, discrimination, and calibration.

The test threshold is the “tipping point” at which the clinician decides to rule-out the disease or gather additional data, and the treatment threshold is the “tipping point” at which clinicians decide to gather additional data or to start treatment. To assess the classification accuracy, we need to make sure that the patients with a low-risk of disease is below the test threshold; the patients with a moderate-risk of disease is falling between the test and treatment threshold; and the patients with a high-risk of disease is above the treatment threshold.[61]

Discrimination statistics tell us how accurately predictions discriminate between patients with or without the outcome of interest (i.e. with or without the disease, with or without mortality). Several methods are used to measure how well the CDR classifies patients in a binary prediction problem. The concordance (*c*) statistic is the output metric most widely used to denote the discriminative ability of generalized linear regression models. The *c* statistic is a rank

order statistic for prediction against true results, similar to Somer's D statistic. For binary outcomes, *c* is equivalent to the area under the receiver operating characteristic (ROC) curve, which plots the sensitivity (true positive rate) against 1-specificity (false positive rate) for consecutive cutoffs for the likelihood of an outcome.[58] We can also calculate the area under the ROC curve (AUROCC). The AUROCC can be interpreted as the area is proportional to ability of the test to discriminate diseased people from non-diseased people. It is generally between 0.5 to 1, and a perfect test has AUROCC closer to 1.[60]

Calibration is another method for model validation. Calibration refers to the agreement between observed outcomes and predictions. The calibration plot has been used to assess model calibration. The calibration plot has predictions on the *x*-axis, and the outcome on the *y*-axis. The calibration plot with perfect predictions should be the observed outcome closer to the diagonal line. The Hosmer-Lemeshow goodness-of-fit test is commonly used for the data with binary outcomes. Hosmer-Lemeshow goodness-of-fit can be tested with an X^2 statistic to assess the ability of a model to fit a given set of data.[60]

Gaps in the Literature

Although IM is common among adolescents and young adults, there is no prospective research that addresses the diagnosis of IM by combining clinical symptoms and signs with the WBC count and differential cell count. Further research should focus on the development and validation of a clinical prediction rule (CPR) to help physicians make an earlier diagnosis of IM among young persons with suspected symptoms and signs.

Previous studies have determined the value of individual symptom or physical sign among adults with diagnosis IM. However, the individual symptom or sign is not sufficient for diagnosing IM. In addition, the location of the test threshold of IM has not been studied in clinical practice. By identifying the test threshold of IM, physicians can easily classify patients between the low- and high-risk groups, and to make decisions about whether ordering additional diagnostic tests of IM is necessary. Although previous studies have considered using

hematologic parameters as criteria to IM, the reference standard that confirmed the IM illness are varied, as well as the inclusion criteria for the patients. There have been no studies that consider the common symptoms and signs along with the hematologic information as the CPR to improve diagnosis of IM. The EBV specific antibody tests are currently the most accurate test to diagnose IM; however, these tests are costly, and the results may not be available for four weeks.

We hope to develop and validate CPRs among college students, since college students live in a more communal living condition and they are more vulnerable to being infected. Every year, 10% of university students developed IM.[62] By early diagnosis of IM, patients can avoid inappropriate antibiotic use, and physicians can better advise their patients regarding the restriction of physical activities. There is also a need for studies to understand the natural history of IM and to identify the risk factors associated with the longer duration of illness among college students. The frequency of clinical visits among college students for sore throat or other suspected symptoms related to IM enables us to obtain clinical information and gives us a good opportunity to learn the risk factors that are associated with IM illness.

We will address these gaps in the literature through the three aims of this dissertation. First, we will determine the test threshold for ordering a diagnostic test for IM. We will then divide the data from the UHC to the derivation cohort and validation cohort and use the derivation cohort to identify the signs and symptoms that predict IM, and to develop the CPRs with this information. We will then use the validation cohort to internally validate rules for the diagnosis of IM, and to stratify the patients into low-, and high-risk groups based on the thresholds identified. We will also assess the diagnostic accuracy of hematologic parameters and add this information to the CPR to see if these parameters improve the diagnosis of IM. We will use four differential methods, including logistic regression, classification and regression tree (CART), fast and frugal tree (FFT), and artificial neural network (ANN) methods to develop the CPR.

Chapter 2 Tables and Figures

Table 2.1 Summary of IM incidence in different populations

Site of Study	Criteria for case ascertainment	Period	Estimated rate (per 10,000 population per year)
Atlanta, GA, US[63]	Heterophile-positive cases in laboratories	1968-1969	45
North East Scotland[64]	Heterophile-positive cases in hospital	1960-1969	11-44
Wisconsin, US[65]	Heterophile-positive cases in laboratories	1960-1961	68
Oxford, England[66]	Heterophile-positive cases in hospital	1955	56
Stockholm, sweden[67]	Clinical IM, hospitalization in Stockholm	1940-1957	5-43
Malmo, Sweden[68]	Clinical IM, hospitalization in Malmo	1954-1960	30-40
Rochester, Minn, US[4]	Heterophile-positive cases in hospital	1950-1969	121

Table 2.2 Summary of the clinical presentations of IM patients from previous studies

Study:	Hoagland, 1960[41]	Stevens, 1951[33]	Sumaya & Ench, 1985[24]	Rea, 2001[32]
Number of Patients with IM:	200	210	113	150
age range	6 to 25 years	under 68 years	under 16 years	greater than 16 years
Reference Standard:	Heterophile antibody test	Heterophile antibody test	IgM or IgG to EBV-VCA	IgM to EBV-VCA
Symptoms and Signs:				
Sore throat	100%	52%	65%	74%
Lymph node enlargement	100%			57%
Fever	98%	34%	90%	45%
Pharyngeal lymphatic hyperplasia	98%			
Pharyngitis	85%	52%	45%	73%
Transient palatal petechiae	50%	52%		
Jaundice	10%	7%		
Rash	3%	11%	20%	15%
Cervical adenopathy		76%	88%	77%
Axillary adenopathy		46%	32%	5%
Inguinal adenopathy		38%		5%
Splenomegaly		43%	80%	8%
Hepatomegaly		15%	64%	7%
Headache		35%		50%
muscle or joints pain		25%		28%
Chills		17%		
Abdominal pain		15%	20%	
Nausea		14%		27%
Fatigue		13%		77%

Table 2.3 Diagnosis of IM: Accuracy of Signs and Symptoms with presence of both heterophile antibodies and EBV IgM positive serology as reference standard[34]

Signs and Symptoms	Sensitivity	Specificity	Positive LR	Negative LR
Splenomegaly	53%	71%	1.83	0.66
Hepatomegaly	37%	73%	1.37	0.86
Rash	17%	86%	1.21	0.97
Lymphadenopathy	89%	25%	1.19	0.44
Headache	83%	27%	1.14	0.63
Sore throat	83%	27%	1.14	0.63
Loss of appetite	74%	24%	0.97	1.09
Malaise	86%	8%	0.93	1.75
Temperature >38C	79%	13%	0.91	1.62
Nausea/vomiting	41%	54%	0.89	1.09
Abdominal pain	38%	56%	0.85	1.12
Jaundice	17%	79%	0.81	1.05
Myalgia	32%	59%	0.79	1.15
Diarrhea	15%	79%	0.71	1.08

Note: LR=Likelihood ratio, and sensitivity is the percentage of patients with IM who have each of these findings.

Table 2.4 Diagnosis of IM: Accuracy of Signs and Symptoms with presence of heterophile antibodies as reference standard[9]

Signs and Symptoms	Sensitivity	Specificity	Positive LR	Negative LR
Pharyngeal erythema	100%	94%	16.67	0.005
Loss of energy	93%	77%	4.04	0.09
Anterior cervical adenopathy	70%	57%	1.63	0.53
Swollen glands	87%	42%	1.50	0.31
Rhinorrhea	47%	63%	1.27	0.84
Nasal congestion	47%	63%	1.27	0.84
Cough	40%	64%	1.11	0.94
Headache	60%	45%	1.09	0.89
Nasal discharge	40%	49%	0.78	1.22
Anorexia	47%	36%	0.73	1.47
Inguinal adenopathy	53%	18%	0.65	2.61
Posterior cervical adenopathy	40%	13%	0.46	4.62
Tonsillar exudate	33%	16%	0.39	4.19
Posterior auricular adenopathy	33%	3%	0.34	22.33
Temperature >37.8C	27%	16%	0.32	4.56
Axillary adenopathy	27%	9%	0.30	8.11
Palatine petechiae	27%	5%	0.28	14.60
Marked axillary adenopathy	20%	1%	0.20	80.00
Pharyngeal exudate	13%	7%	0.14	12.43
Splenomegaly	7%	1%	0.07	93.00
Hepatomegaly	0%	1%	0.01	100.00

Note: LR=Likelihood ratio, the sensitivity is the percentage of patients with IM who have each of these findings. If the sensitivity is 100, 99.5 percent was used to calculate the likelihood ratio to avoid dividing by zero.

Table 2.5 Diagnostic test for IM using different spectrums of atypical lymphocytes and lymphocytes, with positive heterophile antibody as reference standard[44]

Hematologic Criteria	Sensitivity	Specificity	Positive LR	Negative LR
≥10% atypical lymphocytes	75%	92%	9.4	0.27
≥20% atypical lymphocytes	56%	98%	28	0.44
≥40% atypical lymphocytes	25%	100%	50	0.75
≥50% lymphocytes	66%	84%	4.1	0.4
≥50% lymphocytes and >=10% atypical lymphocytes	61%	95%	12	0.41

Table 2.6 Sensitivity and specificity with 95% confidence intervals and likelihood ratios of latex agglutination kits and solid-phase immunoassay kits for diagnosis of IM. The reference standard is the rapid test to detect antibodies to VCA and EBNA.

Patients with suspected IM; Reference standard is EBV-specific antibody tests[12]	**Sensitivity**		**specificity**		**LR+**		**LR-**	
	0-12 yrs	**>=13 yrs**	**0-12 yrs**	**>=13 yrs**	**0-12 yrs**	**>=13 yrs**	**0-12 yrs**	**>=13 yrs**
latex agglutination	44% (38% to 50%)	85% (79% to 91%)	96% (92% to 100%)	90% (82% to 98%)	11	8.5	0.58	0.16
solid-phase immunoassay	37.5% (25% to 50%)	75% (71% to 79%)	96% (92% to 100%)	97% (94% to 100%)	9.4	25	0.65	0.26

Note: Sensitivity and specificity represent the midpoint of the range from nine different tests in the study. If the specificity is 100, 99.5 percent was used to calculate the likelihood ratio to avoid dividing by zero.

Table 2.7 Sensitivity, specificity, and likelihood ratios between heterophile antibody test and EBV specific antibody test for diagnosing IM[55]

Test	Sensitivity	Specificity	Positive LR	Negative LR
Heterophile antibody tests	88% (81% to 95%)	99% (98%, 100%)	88	0.12
Anti-VCA, EBNA tests	97% (95% to 99%)	94% (89% to 99%)	16	0.03

Note: Sensitivity and specificity represent the midpoint of the range from different tests in the study. If the specificity is 100, 99.5 percent was used to calculate the likelihood ratio to avoid dividing by zero.

Table 2.8 Overview of characteristic techniques for internal validation[60]

Method	Development	Validation
Split-sample	1/2 to 2/3 of Original	Independent 1/2 to 1/3
Cross-validation	10 times 90% of Original	Independent 10 times 10%
Bootstrap	Bootstrap sample of size N	Original 100%

CHAPTER 3

METHODS

Aim 1 method overview

The accuracy of individual symptoms, signs, several easily obtainable hematologic parameters, and serologic tests for the diagnosis of infectious mononucleosis (IM) still needs to be confirmed. This aim will perform a systematic review of the literature to identify a comprehensive list of publications that used an individual symptoms, clinical signs, hematologic parameters, and serologic tests for the diagnosis of IM. We will then perform a meta-analysis on the accuracy of the symptoms, clinical signs, hematologic parameters and the serologic tests for the diagnosis of IM. This study will follow the preferred reporting items for systematic reviews and meta-analyses (PRISMA) protocol for guidance.

Sub-Aim 1: Accuracy of the clinical signs, symptoms, and hematologic parameters

Search strategy and selection criteria

We will search the PubMed database to include articles with no restrictions on language. We will use the following search strategy to identify the initial list of publications in the PubMed:

("clinical diagnosis"[tiab] OR "history and physical"[tiab] OR "medical history"[tiab] OR "physical examination"[tiab] OR "signs and symptoms"[tiab] OR "physician's overall estimate"[tiab] OR "overall impression"[tiab] OR "medical history taking"[tiab] OR "physical examination"[tiab] OR "biomarker"[tiab] OR "PCR"[tiab] OR "polymerase chain reaction"[tiab] OR “Epstein-Barr nuclear antigen”[tiab] OR "EBNA"[tiab] OR "IgG"[tiab] OR "IgM"[tiab] OR "VCA"[tiab] OR "WBC"[tiab] OR "white blood cell"[tiab] OR "ELISA"[tiab] OR "leukocyte count"[tiab] OR "complete blood count"[tiab] OR "CBC"[tiab] OR “lymphocyte”[tiab] OR “lymphocytes”[tiab] OR “monocyte” [tiab] OR

“monocytes”[tiab] OR “atypical lymphocytosis”[tiab] OR “lymphocytosis”[tiab] OR “leukocytosis”[tiab] OR “heterophile antibody” [tiab] OR “Paul Bunnell”[tiab] OR “infectious mononucleosis”[tiab] OR “Epstein-Barr Virus”[tiab] OR “hematologic”[tiab]) AND ("infectious mononucleosis"[MeSH Terms] OR “Epstein-Barr Virus”[MeSH Terms] OR “EBV”[tiab] OR “Herpesvirus 4”[MeSH Terms] OR “Mononucleosis”[tiab] OR “Glandular Fever”[tiab] OR “sore throat”[tiab]) AND ("sensitivity"[tiab] OR "specificity"[tiab] OR "likelihood ratio"[tiab] OR "predictive value"[tiab] OR "receiver operating characteristic"[tiab] OR "AUC"[tiab] OR "AUROCC"[tiab] OR "ROC"[tiab] OR "diagnosis"[tiab] OR “diagnostic”[tiab] OR "accuracy"[tiab] OR “predictor”[ti]) NOT (“carcinoma”[tiab] OR “cancer”[tiab] OR “transplantation”[tiab] OR “HIV”[tiab] OR “Hodgkin lymphoma”[tiab] OR “case report” [tiab] OR “cytomegalovirus”[tiab] OR “influenza”[tiab] OR “transplant”[tiab] OR “hepatitis”[tiab])

The filters “has abstract” and “human” will be applied to the search. We will also manually search for relevant articles from the reference lists. Only published and peer-reviewed studies will be included in our study.

All titles, abstracts and the full-text papers will be screened in parallel by two independent reviewers. A third reviewer (ME) will be responsible for resolving discrepancies among the reviewers through a consensus discussion at each stage.

Inclusion and exclusion criteria

We will include studies that perform original data collection and provide sufficient information to construct a 2 x 2 table for the diagnostic accuracy of at least one symptom, clinical sign, or hematologic parameter (e.g., WBC, lymphocytes, atypical lymphocytes, monocytes) for IM against a reference standard test. In this study, we will limit the reference standard tests to commercial heterophile antibody tests (e.g., Monospot test) and viral capsid antigen (VCA)-specific tests (e.g., the indirect immunofluorescence test (IFA), or an enzyme-linked immunosorbent assay (EIA)) for the detection of EBV infection.

We will only include studies with patients that are suspected to have IM or EBV infection and have sufficient information to calculate both the sensitivity and specificity of the test, or case series that have sufficient data to calculate the sensitivity. We will include prospective cohort studies or case series for the study of symptoms and signs, while we will also include retrospective cohort studies for the study of hematologic parameters. Data will collect the data from patients in any age group. We will include studies with both inpatient and outpatient settings with no restrictions on languages, country, or year of publication.

We will exclude review articles, individual case reports, commentaries, editorials, case-control studies, retrospective cohort studies for the study of symptoms or signs, modeling studies, and studies with sample sizes less than five. We will also exclude studies that focus on immunodeficient subjects or subjects with a special condition (i.e., with another acute illness consistent with EBV-related IM, transplant recipients, or HIV-positive subjects), or studies of IM that primarily caused by toxoplasmosis or cytomegalovirus infection.

Data abstraction and analysis

Two investigators will extract the aggregate study data. The following data will be extracted from studies: study characteristics (country, year of recruitment, setting), study populations (mean or median age, gender, sample size, inclusion and exclusion criteria), the index tests being studied, the reference standard test being used, and the results from each study to construct 2 x 2 tables.

We will assess the quality of the selected studies using the quality assessment of diagnostic accuracy studies 2 (QUADAS-2) tool. The four key domains of the QUADAS-2 tool include patient selection, evaluation of the index test performance, evaluation of the reference test performance, and flow of patients through the study with corresponding timing.[69] We will define low, unclear, and high risk of bias for each domain according to the study quality. A third reviewer will resolve any disagreements during the quality assessment. The corresponding definition questions based on the QUADAS-2 instruments are shown in Appendix A.

For data abstraction, we will extract the data needed for the construction of 2 x 2 contingency tables that comparing the diagnostic performance of each clinical findings with the reference standard test. Then, we will calculate the pooled measures of diagnostic performance for each element, such as sensitivity, specificity, positive (LR+) and negative likelihood ratios (LR-), and diagnostic odds ratios (DORs), with their corresponding 95% confidence intervals. The DOR will be calculated by LR+ divided by LR-, where a higher DOR indicates a better diagnostic accuracy for the underlying test to detect IM in our study.[70] If there is only data from a single study, the point estimate and a 95% confidence intervals will be presented; if there are data from two studies, ranges will be presented; if there are data from three or more studies, the summary estimates will be calculated based on a bivariate meta-analysis, and the area under the receiver operating characteristics curve (AUC) will be determined. We will combine the similar cutoffs for the same index test if clinically reasonable; for example, white blood cell counts greater than $9.5x10^9$/L, $10x10^9$/L, and $10.5x10^9$/L will be combined into a single cutoff of "greater than 9.5 to 10.5 x 10^9/L".

We will import the data into R and perform the bivariate analysis using the mada package .[71,72] We will also calculate the AUCs and DORs with 95% confidence intervals (CIs) using this package. If there is only one study describe the accuracy of a test, we will use the diagti procedure in Stata 15.1 (Stat-Corp) to estimate the overall accuracy with 95% CIs.

We will also for the presence of threshold effects during the analysis to check whether the sensitivity decrease and the specificity increase with increasing diagnostic cutoffs or thresholds.[73] The threshold effect will be determined by visually inspecting the receiver operating characteristics curves (ROCs) stratified by the cutoffs of the index test. We will present the summary estimate of diagnostic accuracy of the test stratified by the cutoff value instead of report one summary statistic if threshold effect exists. We will also perform subgroup analyses on selected symptoms, signs, and hematologic parameters by the reference standard

test chosen (heterophile antibody test versus VCA-specific test) and the study settings (inpatient versus outpatient).

Sub-Aim 2: Accuracy of the serological tests

Search strategy and selection criteria

We will search the PubMed database to include articles with no restrictions on language. The search strategy and selection criteria for the initial list of publications will be the same as sub-aim 1.

Inclusion and exclusion criteria

We will include published studies that performed original data collection. The inclusion criteria are as follows: prospective or retrospective cohort studies; provide sufficient information to estimate both sensitivity and specificity for the accuracy of at least one serological test to detect primary EBV infection; and sera samples were drawn from patients with suspected EBV infection or IM within any age group. We will limit our index tests to two categories: heterophile antibody tests to detect the presence of heterophile antibodies, and serological immunoassays to detect VCA IgM, VCA IgG with or without EBNA IgG.

We will limit our reference standard test to immunofluorescence assay (IFA) for the detection of VCA IgM, VCA IgG and EBNA IgG. The VCA IgM, VCA IgG and EBNA IgG are EBV-serological markers that are used to determine the serological profiles characterizing the primary infection. The IFA test is used to detect anti-VCA IgM and IgG, and the anticomplement immunofluorescence assay (ACIF) is used to detect anti-EBNA IgG. A primary EBV infection is indicated if there is a positive result for the VCA IgM and/or VCA IgG with no or a weak antibody response to EBNA IgG.[74] The same reference standard test should be performed on all included sera to avoid the verification bias. No stipulation as to language, country, or year of publication for included studies will be made.

We will exclude case reports, comments, editorials, case-control studies, modeling studies, or studies with fewer than 5 sera samples. We will also exclude studies if a majority of the subjects are immunodeficient or present with another special condition (i.e., with another acute illness consistent with EBV-related IM, transplants recipients, or HIV-positive subjects). Studies of IM caused by toxoplasmosis or cytomegalovirus infections will be excluded as well.

Assessment of methodological quality

Two reviewers will independently assess the risk of bias using a quality assessment of diagnostic accuracy studies 2 (QUADAS-2) checklist tailored to this review.[69] If there is any disagreement, a third reviewer will be consulted. The four key domains of the QUADAS-2 tool include patient selection, evaluation of the index test performance, evaluation of the reference test performance, and flow of patients through the study with corresponding timing. We will define the quality of each domain as low, moderate, or high risk of bias. The QUADAS-2 instruments adapted for this study and the corresponding questions are similar as the previous study, except that we will limit the reference standard test in this study to IFA.

Data abstraction and analysis

Two investigators will independently extract the aggregate data. Any discrepant results will be resolved through discussion between the investigators. The following data will be extracted from each study: characteristics (country, year of recruitment, setting), populations (mean or median age, gender, sample size, inclusion and exclusion criteria), the index test used (commercial test name, antibodies detected), the reference standard test used, and the results from each study to construct 2 by 2 tables. If a study used more than one commercial test, we will consider each particular commercial test as an individual study.

We will perform random effects meta-analyses of the prevalence of EBV-infected patients or patients with IM stratified by age group using the metaprop procedure in R (version 3.5.2). We will first evaluate the diagnostic accuracy for different categories of serologic tests for the detection of primary EBV infection. To evaluate the accuracy of serological immunoassays,

all patients testing with a negative result for EBNA IgG and either with positive VCA IgM only or with positive VCA IgM and VCA IgG will be considered as primary EBV-infection. We will also assess the diagnostic accuracy for the detection of the anti-VCA IgM or anti-VCA IgG alone for the immunoassays being studied using the IFA results as the reference standard. The diagnostic accuracy for the detection of anti-VCA IgM helps to distinguish acutely infected patients from those having had past infections and from seronegative patients. The diagnostic accuracy for the detection of anti-VCA IgG helps to distinguish patients having had past infections from primarily infected and seronegative patients.

The pooled measures of the diagnostic performance of each element will include sensitivity, specificity, and LR+ and LR- with 95% CI, and the DOR.[70] If there is only one study for an index test, the point estimate and a 95% confidence interval will be calculated using the binomial exact method; if there are two studies, the ranges will be presented; if there are more than two studies, the summary estimate will be calculated based on bivariate analysis.[75]

To describe heterogeneity among studies, we will calculate the AUC for each index test. To assess the study population and design characteristics as potential determinants of diagnostic accuracy, we will visually inspect the ROC curves to explore the sources of heterogeneity.

Aim 2 Method Overview

Ordering a serological test for infectious mononucleosis (IM) in all patients with sore throat is costly and impractical. The test threshold to determine when to order a diagnostic test for IM based on the patient's symptoms has not been previously studied. The objective of this aim is to determine the test threshold for IM in the management of patients with a sore throat. To achieve this objective, we will send online surveys to a convenience sample of US primary care clinicians regarding their decision-making about whether or not to order a test for IM in a patient with a sore throat. The clinical vignettes are created in the survey for seven patients with different combinations of symptoms and signs. The probability of IM for each vignette will be estimated by the investigator based on a plausible range. We will then ask the clinicians to decide whether to test or not test for IM, and mixed-effect logistic regressions will be used to determine the test threshold for IM.

Participants

We will recruit a convenience sample of primary care physicians and email a participation invitation nationwide from spring to fall 2020. Attached to this email will be a link to an online questionnaire about an IM test threshold. We will administer the survey using Qualtrics, provide through the University of Georgia (UGA), which gives a secure and automated method for data collection. Respondents can access the online survey either on a mobile phone or on a personal computer. We will include all participants who submit the Qualtrics survey in our study. All surveys will be completed anonymously, and each IP address can only be used once to avoid duplicate responses.

Study Design

This is a cross-sectional study of clinicians regarding their clinical decision-making regarding the diagnosis of IM in an outpatient setting with the availability of the office-based examination for IM. After giving the informed consent, each clinician will be asked about their medical specialty (family medicine, internal medicine, physician assistant, or nurse practitioner),

years in practice, practice site (primary care, urgent care, or emergency department), and whether they are working at a university health center. Then, each clinician will be presented with seven separate clinical scenarios of patients with a sore throat and different combinations of signs and symptoms, corresponding to an estimated likelihood of IM ranging from 1% to 30%. For each scenario, the likelihood of IM will be given to the clinicians according to expert opinion and a previously published meta-analysis by the co-investigator on the accuracy of the symptoms and signs for the diagnosis of IM.[35] We will inform the clinicians that the heterophile antibody test, the "Monospot" test, is the only confirmatory test available to the physicians for diagnosing IM. The "Monospot" test is considered to be 80% sensitive in the first 7 days of infection and 95% sensitive after 7 days. We will then ask the clinicians to select from one of the clinical decisions below for each of 7 clinical viignettes:

- You feel that IM is unlikely, and you will not order any tests for IM.
- You feel that more information is needed, and you will order a "Monospot" test.

In the threshold model, these two options are consistent with being below the test threshold and being above the test threshold, respectively. Examples of the online questionnaire are shown in Appendix C.

Data Analysis

For the descriptive analysis, we will summarize the characteristics of each participating clinician descriptively. We will also summarize the frequencies and the percentage of clinicians that ruling out IM for each vignette.

The test threshold will be determined by adopting the method described in a previous study.[61] This method is based on a logistic regression analysis of the physician's decision regarding the disease probability. The following logistic regression equation will be used to determine the test threshold:

$$ln\left[\frac{p}{1-p}\right] = a + bx$$

Equation 2.1

where *p* is the probability that a clinician decided not to rule out the disease and to order a diagnostic test when the test threshold is being estimated; the value *x* is the probability of IM predetermined by the investigator for each vignette; while *a* and *b* are the model coefficients.

The test threshold of disease is defined as the probability of disease where 50% of the clinicians would decide to rule out disease and the other 50% decide to order diagnostic test or initiate treatment. Thus, we will define the probability of ruling out IM (*p*) as 0.5. At this probability, 50% of the clinicians will rule out IM without ordering the additional test and another 50% of clinicians will decide to order the "Monospot" test to confirm IM. The following equation is obtained after inverting Equation 1 with respect to *x* and replacing *p* with $\bar{x}$,

$$X_{test}=-a/b$$

Equation 2.2

where $\hat{a}$ and $\hat{b}$ are the coefficients estimated from Equation 1. The resulting X_{test} is deemed as a test threshold. Since each clinician needed to evaluate seven scenarios, mixed-effect logistic regression models will be used, and we will apply a random intercept term to Equation 1 in order to adjust for inter-physician variability. The confidence intervals of the test threshold for IM will be determined by using the covariance matrix for the estimated coefficients in the model.[76]

To achieve the subgroup comparison, we will also stratify the test threshold model by years of medical practice (<=10 years vs. >10 years), practice sites (primary care vs. non-primary care), clinician specialties (family physician vs. non-family physician), and whether they work at a student health center.

We will perform all statistical analyses using R software version with version 3.0.2.[72] The mixed-effect logistic regression will be implemented by glmer() function from lme4 package.

Aim 3 Methods Overview

Individual symptoms or signs of infectious mononucleosis (IM) are of limited value in determining the presence of the disease. Clinicians would benefit from a clinical prediction rule (CPR) in the diagnosis of IM using a combination of clinical symptoms, signs, and hematologic parameters. The objective of this aim is to develop and to internally validate simple risk scores based on clinical symptoms and signs both with and without hematologic parameters in order to diagnose IM among college students using the traditional logistic regression method.

To achieve this goal, we will extract the data from the electronic health records of a university health center from 2015 to 2019, and the data will be divided into derivation and validation cohorts. We will then use the multivariate logistic regression models to develop two prediction models in the derivation cohort: one with only symptoms and signs (IM-Nolab) and one adding hematologic parameters to the model (IM-Lab). We will create the point scores based on the regression coefficients and group the patients into low- and high-risk groups. We will then validate these scores using the validation cohort. We will also use three innovate statistical methods (classification and regression tree (CART), fast and frugal tree (FFT), and artificial neural network (ANN)) to develop the IM-Nolab and IM-Lab models, and internally validate each model in the validation cohort.

Study Design and Data Collection

This is a cross-sectional study of a previously collected dataset from the university health center (UHC) at the University of Georgia (UGA). We will assemble the standardized dataset of the patients' symptoms, signs, and laboratory parameters using extant literature. We include any patients cared for between September 1st, 2015 and January 1st, 2019 at the UGA UHC who had a diagnostic test for IM. The UHC at UGA provides primary care, specialty health care, education and prevention-focused services to approximately 35,000 students enrolled at the university each year. The UHC has four primary care clinics with approximately 20 primary care clinicians available during regular business hours. The UHC is accredited by the Joint

Commission, the nation's most prestigious accrediting board, as well as by the Commission on Colleges of the Southern Association of Colleges and Schools (SACS).[77] Thus, this health care facility is a suitable and ideal place to study IM illness in a college-health population.

The UHC uses an electronic health record system (EHR) to record and maintain the patients' symptoms, signs, and laboratory test results for each clinical visit. The UGA health center staff, who are not study team members, are responsible for linking the clinical and laboratory data and for removing any identifier/personal information, including name, age in years, birthdate, address, contact information, and student ID number. Each patient was assigned a random ID number only known by the staff in order to maintain confidentiality. The de-identified data was securely transferred from the UHC to the study investigators for analysis. We will merge the dataset for clinical presentation and laboratory parameters by each patient's ID number, as created by health center staff, and the date of visit. We then exclude patients with no record available regarding signs, symptoms, hematological parameters or diagnostic test results.

Independent Predictors and Outcome Measures

The independent predictors in this study include demographic data, symptoms, signs, as well as hematologic parameters. The clinical symptoms and signs are recorded as the presence of one or more of the following conditions: fever, diarrhea, vomiting, fatigue, headache, joint pain, myalgia, nausea, rash, sore throat, swollen lymph nodes, cough, anterior cervical lymphadenopathy, posterior cervical lymphadenopathy, pharyngeal erythema, tonsillar erythema, exudative pharyngitis, tonsillar enlargement, and/or tonsillar exudate. The clinical symptoms were reported by patients in the portal and the clinical signs were evaluated by physicians during the physical examination. The fields in the EHR are optional, where patients and physicians only need to check the symptoms and signs the patient had. We will assume that all the blank inputs in the symptoms or signs for each patient as negative. The laboratory findings include a lymphocyte counts and its percentage, a neutrophil count and its percentage,

a monocyte counts and its percentage, a white blood cell count, as well as the percentage of the atypical lymphocytes.

The primary outcome is the positive result of IM as diagnosed by the Monogen test. The Monogen test is the heterophile antibody test in a convenient latex agglutination form.[78] The sensitivity and specificity of Monogen are 94.2% (95% CI: 87.9% to 97.9%) and 91.3% (95% CI: 84.7% to 95.8%), when using a hemagglutination test as the reference standard, and are 99% (95% CI: 89.5% to 99.5%) and 93.3% (95% CI: 85.7% to 96.4%), when using an EBV-specific test to resolve the non-discrepant results.[78] The heterophile antibody test is rapid and cost-effective; therefore, it is widely used for IM diagnosis in the health clinics.[24]

Data Analysis

We will use a temporal split sample approach to derive the derivation and validation cohorts. The data collected from consecutive patients visiting the UGA UHC from January 1st, 2017 to January 31st, 2019 will be used as the derivation cohort, and we will use the derivation cohort to build the model and develop the point scores (N=1498, 64%). We will then use the data collected from September 1st, 2015 to December 31st, 2016 as a validation cohort (N=844, 36%) to evaluate the accuracy of the point scores.

Variable selection

For a baseline assessment, we will present the continuous variables in medians with interquartile ranges (IQRs) and we will also summarize the dichotomous variables by the frequencies of occurrence with their corresponding proportions. We will stratify the characteristics of the independent predictors by derivation and validation datasets. We will then compare the values variables with and without a positive test for IM in the derivation set. The student's t-test will be used to compare the continuous variables, and the Pearson χ^2 statistics or Fisher's exact test will be used to compare the proportion of the categorical variables between patients with and without a positive test for IM, as appropriate. We will consider the association between predictors and outcome is statistically significant if the two-tailed p-values

are less than 0.05. We will then select variables that are significantly associated with our outcome at $p<0.1$ for the inclusion in multivariable analysis.

Logistic regression models

To develop the point scores using the derivation set, we will fist use the patient's symptoms and signs as independent predictors (IM-Nolab). Then, we will add the hematologic parameters as independent predictors (IM-Lab) to the model. If two predictors are statistically correlated, we will select the one that has a stronger association with the outcome in the univariate analysis in order to avoid multicollinearity. The hematological parameters will be converted to binary predictors to simplify calculations for the final risk score. The cutoffs will be decided based on the inspection of histograms, as well as the biological plausibility from previous systematic review.[35] To build a more parsimonious model, we will apply a forward model selection guided by the Akaike Information Criterion (AIC).[14,15] The beta coefficients will be determined from the final multivariate models. To improve the model prediction, we will apply a uniform shrinkage factor to the regression coefficients to move them toward zero. We will then assign a point score to each predictor in the model by transforming its corresponding beta coefficient. Each beta coefficient will be divided by the smallest beta value and then rounded to the nearest integer. We will then categorize the study population into low-risk and high-risk groups based on the testing threshold calculate from previous aim.

To assess the performance of the model, we will validate the model internally by using 10-fold cross-validation. The area under the receiving operating characteristic curve (AUC) and the associated 95% confidence intervals (CIs) will be calculated from the cross-validation samples. We will then validate the point scores using the validation cohort. We will evaluate the overall ability of by a receiving operating characteristic curve (ROC) and AUC statistics. The calibration of the scores will be evaluated using the Hosmer-Lemeshow (H-L) test and a calibration plot, which measures how well the predicted outcome matched the observed

outcome. We will also evaluate the accuracy of the scores based on the probability of patients being classified into low- and high-risk groups in the validation cohort.

Classification and regression tree analysis

The CART approach is a non-parametric statistical method for multivariate analysis. CART is a form of binary recursive partitioning in which each parent node can be divided into two child nodes by identifying the predictors that best differentiate the population into groups with or those without the outcome of interest, in this case laboratory-confirmed IM cases. The child nodes may themselves be divided into additional children in the same manner.[79] CART analysis consists of four basic steps: building the tree, stopping the tree-building process, "pruning" the tree, and identifying the optimal tree.[79] During the tree-building process, a CART approach will check all possible splitting variables and will assign a predicted class to each resulting node that maximizes the sensitivity and specificity of the classification. This process stops when prespecified minimum number of cases is reached at each terminal node or when all observations within each child node have an identical distribution of predictor variables.

The Gini index method will be used to break the parent nodes into child nodes with the default split size set to allow the tree to expand. We will use the logworth statistic to rank each candidate predictor in order to identify the optimal split for each node. The logworth statistic is the negative log of adjusted p-values for the chi-square statistic. Splits will then be chosen based on the significant values for the candidate predictors.[80] For hematologic parameters, the cutoff values are decided by maximizing the sum of mean squares for the differences between groups. In order to minimize the number of branches without significantly affecting the goodness-of-fit, the tree is pruned by omitting either predictors that have lower sensitivities or variables that are not classify a substantial number of patients into a risk group.[79] In our analysis, we will specify a minimum number of cases at each terminal node of 10 to avoid having unstable estimates with a wide confidence interval. We will then categorize the nodes into low-risk and high-risk groups based on the testing threshold calculate from previous aim.

Fast and frugal tree analysis

We will then apply the FFT analysis for multivariate analysis. The FFTs usually make decisions faster than a traditional regression analysis and take less time to implement.[81] In the FFT, each level of the tree has one binary classification (cue) and a cue-based question (node). The cues are ranked, and the solution to each cue generates an exit or a further node. The procedure to construct a FFT includes selecting cues, setting the decision threshold for each cue, and determining the sequence of the cues and the exit for each cue. A *fan* algorithm is used to address these tasks by selecting the tree in the exit structure, restricting the size of the tree and removing the unnecessary nodes.[81]

In our study, we will create and visualize the FFTs and select the trees using the *fan* algorithm on the derivation cohort. To simplify the models for later ease of use, we will limit the maximum number of nodes to three. We will first plot and visualize all cues and rank the cues by weighted accuracy (WACC) for IM among all cues for each model. The WACC is defined as the weighted average of sensitivity and specificity dictated by a weight parameter. The weight parameter shows how sensitivity relates to the specificity. We will then select the top three cues with the highest WACC as the final nodes in the FFTs models. Finally, we will construct and visualize the FFT models and will calculate the probability of IM as classified at each node. The probability of IM for each node will be calculated and we will classify the nodes into low-risk and high-risk groups based on the testing threshold calculate from previous aim.

Artificial neural network

ANNs have the ability to learn mathematical relationships from a corresponding output variable among a set of input variables.[82] A typical ANN contains a series of nodes within three layers (input, hidden, output). The input nodes consist of all possible predictors, and the output node represents the outcome variable.[82] The nodes within the hidden layer are intermediate values that are calculated by the network. The hidden nodes allow the network to model complex, nonlinear relationships between the input variables and the output variable. Similar to

the beta coefficient in logistic regression models, each input node is connected to the hidden node by a connection weight, and the connection weight contains the information acquired by trained ANNs.[82] Since there was no existing theory to predetermine the optimal number of hidden layers, we will select the number of hidden layers based on the accuracy of the prediction network.

A training algorithm for an ANN will be used to minimize the difference between the value of the predicted output and the actual outcome variable, thereby minimizing the mean square error (MSE) of the network, by estimating the connection weights.[82] The back-propagation training algorithm is one of the most commonly used algorithms to determine optimal connection weights. The learning rate refers to how much each weight changed after the training, with higher learning rates suggesting greater weight changes. The momentum allows the magnitude of each weight change to be proportional.[82]

We will develop ANN models in the derivation cohort using a standard feed-forward, back-propagation ANN. The numbers of hidden-layer neurons will be determined to optimize the sensitivity and the specificity of the models. To adjust for back-propagation, we will set the learning rate as 0.2 and the momentum as 0.9, respectively.[83] The output neuron is created to have a value of 1 for patients with IM and 0 for patients without IM. The predicted IM cases will be calculated through an iterative process. We will then compare the predicted outputs to the actual output values, and the connection weights will be adjusted according to any errors in the network. This process will be completed until the minimum MSE is lower than 0.01. We will then calculate the likelihood of IM for each patient by ANN.[83]

Model validation

We will evaluate the performance of each model using the validation cohort. We will first plot the ROC curves and calculate the AUC to evaluate the discrimination of each model. The calibration plot will be visualized for each method. We will also compare the accuracy of IM probability in the low- and high-risk group in the derivation and validation cohorts.

CHAPTER 4

ACCURACY OF SIGNS, SYMPTOMS, AND HEMATOLOGIC PARAMETERS FOR THE DIAGNOSIS OF INFECTIOUS MONONUCLEOSIS[1]

[1] Cai, X., Ebell, M. H., & Haines, L. To be submitted to *the Journal of the American Board of Family Medicine.*

Abstract

Background: The accuracy of individual symptoms, signs, and several easily obtainable hematologic parameters for the diagnosis of infectious mononucleosis (IM) still needs to be confirmed. Improving diagnosis of IM based on the clinical findings could prompt physicians to better identify which patients need a diagnostic test for IM.

Method: This was a systematic review to determine the accuracy of symptoms, signs, and hematologic parameters in patients with suspected IM that used heterophile antibody test or viral capsid antigen tests as the reference standard. Two reviewers reviewed all studies in parallel and assessed the quality of the selected studies using the QUADAS-2 criteria. The pooled measures of diagnostic performance were calculated by bivariate meta-analysis for each clinical finding, which included sensitivity, specificity, likelihood ratios, the diagnostic odds ratios, and the area under the receiver operating characteristic curve.

Results: Seventeen studies were included in our final analysis. The prevalence of IM ranged from 2.1% to 80% among prospective cohort studies. The presence of splenomegaly (LR+: 2.39, 95% CI: 1.11 to 5.51), palatal petechiae (LR+: 1.32 to 11.40), posterior cervical lymphadenopathy (LR+: 3.16, 95% CI: 1.45 to 5.20), and axillary or inguinal cervical lymphadenopathy (LR+: 3.05, 95 CI: 1.85 to 4.70) were moderately useful for ruling in IM. The most helpful hematologic parameters for ruling in IM include lymphocytes greater than 4×10^9/L and greater than 40% to 50%, or atypical lymphocytes greater than 40%. A combination of lymphocytes greater than 50% and atypical lymphocytes greater than 10% (LR+: 50.40, 95% CI: 8.43 to 162) was also found to be helpful to rule in disease. Most of the clinical findings have limited diagnostic value in ruling out the disease when absent.

Conclusion: Although most symptoms and signs were unhelpful, the likelihood of IM is appreciably increased by several examination findings, and hematologic parameters were more accurate than symptoms and signs. Since most clinical findings have limited diagnostic value in

ruling out the disease, physicians should not rely on the absence of any individual symptom or clinical sign for ruling out IM.

Background

Infectious mononucleosis (IM) is a disease caused by the Epstein-Barr virus (EBV). IM is most common among young adults, especially among person aged 15 to 24 years, with the incidence rate ranging from 6 to 8 cases per 1000 person years.[4] A review of the literature demonstrates that adolescents with a sore throat accompanied by the presence of fever, myalgias, fatigue, and chills are suspected of having IM.[35] The most common symptoms and signs among IM patients include sore throat, lymph node enlargement, fever, tonsillar enlargement, pharyngeal inflammation, fatigue, and rashes.[32,41] However, the diagnostic accuracy of the symptoms and signs of IM are inconsistent among studies.[9,34]

A recent systematic review[35] of the diagnostic accuracy for individual symptoms and signs found that the absence of sore throat (sensitivity: 0.81; negative likelihood ratio (LR-): 0.51-0.62), headache (sensitivity: 0.66; LR-: 0.63-0.73), or any lymphadenopathy (sensitivity: 0.91; LR-: 0.23-0.44) reduces the likelihood of IM, and the presence of posterior cervical adenopathy (specificity: 0.87; positive likelihood ratio (LR+):1.6-5.9) or splenomegaly (specificity: 0.71-0.99; LR+: 1.9-6.6) increases the likelihood of IM. From this review, only two studies[9,34] reported on both the sensitivity and specificity of the symptoms and clinical signs, and other studies used patients with known IM, which can only determine the sensitivity. Therefore, the accuracy of the clinical findings still needs to be confirmed.

Previous studies[34,41,45] also found that several easily obtainable hematologic parameters, including complete blood counts (CBC) and leukocyte differential counts, are helpful in diagnosing IM among clinically suspected patients. The presence of atypical lymphocytosis has been shown to significantly increase the likelihood of IM, and a high percentage of lymphocytes also helps to increase the likelihood of IM when combined with atypical lymphocytosis.[34,44] Patients presenting with monocytosis were also found to have a higher likelihood of IM.[44,46] Thus far, there is only one systematic review[35] focused on the accuracy of clinical presentations or hematologic parameters for the diagnosis of IM. Most of the included studies from the

previous review were old, and some prospective cohort studies or case series were not included in that analysis.

In the present analysis, we performed an updated meta-analysis of the accuracy of the symptoms, clinical signs, and hematologic parameters for the diagnosis of IM. We also did a subgroup analysis of the commonly reported risk factors according to reference standard tests and in various patient settings to assess potential sources of heterogeneity.

Method

Our study followed the preferred reporting items for systematic reviews and meta-analyses (PRISMA) protocol for guidance.[84]

Search strategy and selection criteria

This systematic review was conducted by searching the PubMed database for articles published through 30 September 2020, with no restrictions on language. The following search concepts were used involving all possible elements: “signs, symptoms, and diagnostic tests”; “infectious mononucleosis or Epstein-Barr virus”; and “accuracy of diagnosis” linked by the Boolean operator “AND”. The filters “has abstract” and “human” were applied to the search. The complete search strategy for the PubMed database is shown in Appendix A. We also manually searched for relevant articles from the reference lists of the included studies. This meta-analysis included only published and peer-reviewed studies.

All titles and abstracts were screened in parallel by two independent reviewers (XC, LH). If there was any uncertainty regarding eligibility, a third reviewer (ME) was consulted. The full-text paper for each abstract was then independently reviewed by two reviewers (XC, LH), and the third reviewer (ME) was responsible for resolving discrepancies among the reviewers through a consensus discussion at this stage.

Inclusion criteria

This systematic review and meta-analysis includes studies that performed original data collection and provided sufficient information to construct a 2 x 2 table for the diagnostic

accuracy of at least one symptom, clinical sign, or hematologic parameter (e.g., WBC, lymphocytes, atypical lymphocytes, monocytes) for IM against a reference standard test. In this meta-analysis, we limited the reference standard tests to commercial heterophile antibody tests (e.g., Monospot test), viral capsid antigen (VCA)-specific tests (e.g., the indirect immunofluorescence test (IFA), or an enzyme-linked immunosorbent assay (EIA)) for the detection of EBV infection.

Eligible studies included cohort studies with patients that were suspected to have IM or EBV infection and had sufficient information to calculate both the sensitivity and specificity of the test, or case series that had sufficient data to calculate the sensitivity. Only prospective cohort studies or case series were included for the study of symptoms and signs, while retrospective cohort studies were also included for the study of hematologic parameters. Data were collected from patients in any age group. The same reference standard test had to be performed for all patients. We included studies with both inpatient and outpatient settings with no restrictions on languages, country, or year of publication.

We excluded review articles, individual case reports, commentaries, editorials, case-control studies, retrospective cohort studies for the study of symptoms or signs, modeling studies, and studies with sample sizes less than five. A study was also excluded if it focused on immunodeficient subjects or subjects with a special condition (i.e., with another acute illness consistent with EBV-related IM, transplant recipients, or HIV-positive subjects); studies of IM primarily caused by toxoplasmosis or cytomegalovirus infection were eliminated from consideration as well. A list of excluded studies and a log of the respective reasons can be provided by the authors upon request.

Data abstraction

Two investigators (XC, LH) were responsible for extracting the aggregate study data. The following data were extracted from studies: study characteristics (country, year of recruitment, setting), study populations (mean or median age, gender, sample size, inclusion

and exclusion criteria), the index tests being studied, the reference standard test being used, and the results from each study to construct 2 x 2 tables. If the absolute numbers were not reported, they were estimated based on the total number of patients, the sensitivity, and the specificity.

Quality Assessment

Two reviewers (XC, LH) assessed the quality of the selected studies using the quality assessment of diagnostic accuracy studies 2 (QUADAS-2) tool. The four key domains of the QUADAS-2 tool include patient selection, evaluation of the index test performance, evaluation of the reference test performance, and flow of patients through the study with corresponding timing.[69] We defined low, unclear, and high risk of bias for each domain according to the study quality. A third reviewer (ME) was responsible for resolving any disagreements during the quality assessment. We then displayed the proportion of studies by differing degree of risk of bias.[69] The full QUADAS-2 instruments adapted for our study and their corresponding definition questions are shown in Appendix A.

Analytic Strategy

In our primary analysis, we calculated summary estimates of the diagnostic accuracy of the symptoms, clinical signs, and hematologic parameters of IM. We first extracted the data needed for the construction of 2 x 2 contingency tables that compared the diagnostic performance of each clinical findings with the reference standard test. Then, we calculated the pooled measures of diagnostic performance for each element, such as sensitivity, specificity, positive (LR+) and negative likelihood ratios (LR-), and diagnostic odds ratios (DORs), with their corresponding 95% confidence intervals. The DOR is calculated by LR+ divided by LR-, where a higher DOR indicates a better diagnostic accuracy for the underlying test to detect IM in our study.[70] If there were only data from a single study, the point estimate and a 95% confidence intervals are presented; if there were data from two studies, ranges were presented; if there were data from three or more studies, the summary estimates were calculated based on a

bivariate meta-analysis, and the area under the receiver operating characteristics curve (AUC) was determined. Similar cutoffs for the same index test were combined if clinically reasonable; for example, white blood cell counts greater than 9.5x109/L, 10x109/L, and 10.5x109/L were combined into a single cutoff of “greater than 9.5 to 10.5 x 109/L”.

We imported the data into R and performed a bivariate analysis on three or more studies for the same index test with same or similar cutoffs using the mada package .[71,72] The AUCs and DORs with 95% confidence intervals (CIs) were also calculated using this package. If only one study described the accuracy of a test, we used the diagti procedure in Stata 15.1 (Stat-Corp) to estimate the overall accuracy with 95% CIs.

We also looked for the presence of threshold effects during the analysis to check whether the sensitivity decreased and the specificity increased with increasing diagnostic cutoffs or thresholds.[73] The threshold effect was determined by visually inspecting the receiver operating characteristics curves (ROCs) stratified by the cutoffs of the index test. If a significant threshold effect existed, it was generally not feasible to calculate or report one summary statistic for the diagnostic performance of each test, and we instead presented the summary estimate of diagnostic accuracy of the test stratified by the cutoff value.

Finally, subgroup analyses were performed on selected symptoms, signs, and hematologic parameters by the reference standard test chosen (heterophile antibody test versus VCA-specific test) and the study settings (inpatient versus outpatient).

Patient and public involvement

Patients did not engage in developing the research, measuring the research outcomes, conducting the research, nor in preparing the manuscript.

Result

Study characteristics

There were 1173 abstracts identified by our initial search, as well as 14 from the review of the reference lists. Thus, the authors reviewed a total of 1187 abstracts, of which 191 were reviewed in full, and 17 studies met our inclusion and exclusion criteria and were included in our quantitative analysis. The search process is summarized in Figure 4.1.

The characteristics of each included study are summarized in Table 4.1. The number of patients studied ranged from 25 to 1,000. The average age of the participants was between 4 and 32 years, with 37% to 72% being female. There were eight studies set in the United States,[9,13,32,85-89] one each in Canada,[44] Israel,[34] and Saudi Arabia,[90] and the remainder were from Europe. We identified 12 studies with information on the accuracy of clinical symptoms and signs, of which 8 were prospective cohort studies[9,85,88,90-94] and 4 were case series;[32,86,89,95] only the sensitivity of the symptoms and signs could be calculated from these case series. We also identified 10 studies for the study of hematologic parameters,[9,13,44,46,87,88,90,91,93,96] of which 3 were retrospective studies.[46,87,96] Seven studies used the heterophile antibody test as the reference standard test for IM[9,44,46,87,89,94,96] and 4 studies used the VCA-specific test as the reference standard;[13,85,88,93] 6 studies used both the heterophile antibody test and VCA-specific test as the confirmatory test for IM.[32,86,90-92,95] The prevalence of IM ranged from 2.1%[9] to 80%[85] in the prospective cohort studies. Six prospective cohort studies[9,13,46,87,92,93] had a prevalence of IM below 50% and two[9,46] were below 10%.

Quality assessment

The study quality was assessed using the QUADAS-2 framework. Of the 17 studies, 3 were judged to have a high risk of bias, [85,89,94] 5 a moderate risk of bias,[13,86-88,90] and the remainders were judged to have a low risk of bias. The detailed description of the quality assessment using the QUADAS tool is described in Table 4.2.

Accuracy of the clinical findings

The accuracy of the clinical symptoms, clinical signs, and hematologic parameters is summarized in Table 4.3. The clinical symptoms significantly associated with IM based on likelihood ratios were headache (LR+: 1.19, 95% CI: 1.01, 1.45; LR-: 0.72, 95% CI: 0.5, 0.98) and sore throat (LR+: 1.12, 95% CI: 1.01, 1.25; LR-: 0.67, 95% CI: 0.41, 0.99). No other symptoms had likelihood ratios significantly associated with the diagnosis of IM.

Regarding signs, splenomegaly (LR+: 2.39, 95% CI: 1.11, 5.51; LR-: 0.66, 95% CI: 0.5, 0.84), palatal petechiae (LR+: 1.32-11.40; LR-: 0.57-0.94), and any lymphadenopathy (LR+: 1.26, 95% CI: 1.05, 1.65; LR-: 0.37, 95% CI: 0.2, 0.67) significantly increased the likelihood of IM when present and reduced the likelihood of IM when absent. The absence of lymphadenopathy was the most accurate to rule out IM, and it had the highest AUC at 0.81. The presence of any lymphadenopathy (LR+: 1.26, 95% CI: 1.05, 1.65; LR-: 0.37, 95% CI: 0.20, 0.67), posterior cervical lymphadenopathy and axillary (LR+: 3.16, 95% CI: 1.45, 5.2; LR-: 0.68, 95% CI: 0.41, 0.93) or inguinal lymphadenopathy (LR+: 3.05, 95% CI: 1.85, 4.7; LR-: 0.67, 95% CI: 0.36, 0.91) were more helpful for ruling in than for ruling out IM. However, the likelihood ratios for posterior cervical lymphadenopathy and axillary or inguinal lymphadenopathy were calculated solely based on a single study with 709 IM susceptible patients. Pharyngitis had the highest sensitivity and was reported in two case series (sensitivity: 0.94; 95% CI: 068, 0.99). No data was available on the specificity of pharyngitis as it was usually a criterion for study entry; therefore, the likelihood ratios for pharyngitis could not be determined. Fever (measured temperature ≥ 37 to 38.0°C) was commonly reported but has little discriminatory value; this might be due to the fact that having a fever was usually required as an entrance criterion for the studies.

Hematologic parameters were more accurate in diagnosing IM than symptoms and signs (Table 4.3). Based on the analysis, absolute lymphocyte counts greater than 4×10^9/L increased the likelihood of IM (LR+: 10.20, 95% CI: 4.79, 16.00). The likelihood of IM was also increased

by an increasing percentage of atypical lymphocytes, from atypical lymphocytes greater than 10% (LR+: 8.97, 95% CI: 3.39, 19.50) to atypical lymphocytes greater than 40% (LR+: 50.26, 95% CI: 38.60, 64.10). The ability to rule out IM with different percentages of atypical lymphocytes was similar: the negative likelihood ratio ranged from 0.45 (95% CI: 0.38, 0.52) to 0.75 (95% CI: 0.68, 0.82). Our analysis also found that a higher cutoff values for the percentage of lymphocytes had a stronger ability to rule in IM and a weaker ability to rule out disease. Our analysis also found that patients with a combination of lymphocytes greater than 50% and atypical lymphocytes greater than 10% would be more accurately ruling in IM (LR+: 50.4, 95% CI: 8.43, 162) compared to their individual effect. Other hematologic parameters associated with IM included monocytes greater than 1×10^9/L (LR+: 1.46-7.89, LR-: 0.11-0.98), leukocytes greater than 5×10^9/L (LR+: 1.05-1.36, LR-: 0.15-0.7) or 10×10^9/L (LR+: 2.55, 95% CI: 1.58, 3.99; LR-: 0.68, 95% CI: 0.56, 0.82).

The diagnostic odds ratio (DOR) estimates the overall diagnostic accuracy of each index test. Based on the results, the highest DORs for the diagnosis of IM were atypical lymphocytes greater than 40% (DOR: 355, 95% CI: 7.43, 622), followed by a combination of atypical lymphocytes greater than 10% and lymphocytes greater than 50% (DOR: 81.20, 95% CI: 19.10, 216.00), and then absolute lymphocyte counts greater than 4×10^9/L (DOR: 30.6, 95% CI: 6.53, 99.40). Posterior cervical lymphadenopathy (DOR: 5.18, 95% CI: 1.55, 12.60) had the best DOR compared to other individual symptoms and signs.

Summary ROC curves for the percentage of lymphocytes and for atypical lymphocytes by cutoff values are shown in Figure 4.2. The differences in accuracy are related to the differences in the cutoffs, which indicate the threshold effects. Therefore, we reported an estimate of these hematologic parameters separately by cutoff values. The summary ROC curves by the use of different reference standard tests for the lymphocytes greater than 50% and the atypical lymphocytes greater than 10% are shown in Figure 4.3. For the subgroup analysis, we found that the symptom of sore throat was relatively more specific in the outpatient

setting compared to in hospitalized patients, and the sensitivity of lymphocytes greater than 50% was higher in the outpatient setting compared to the inpatient setting. However, these findings were limited by the number of studies. No clear pattern for other clinical presentations or hematologic parameters was found regarding the reference standard test chosen (Figure 4.4) or study settings (Figure 4.5) based on visualizing the ROC curves.

Discussion

Summary of the findings

This study is an updated meta-analysis on the individual symptoms, clinical signs, or hematologic parameters for predicting IM. Eight new studies with data for several symptoms and signs and three new articles about hematologic parameters that were not included in previous review[35] were included in our analysis, supporting the need for an updated review.

There is a wide variation in the prevalence of IM among our included studies, with the range between 2.1% and 80% for patients with IM. The varying prevalence of IM is due to the different study settings and inclusion criteria. The study[9] with a prevalence of 2.1% had very broad inclusion criteria, which comprised all patients aged between 16 to 73 years with sore throat or strep throat. The study[90] having the highest prevalence (80%) had more restrictive inclusion criteria: clinical symptoms suggestive of IM, lymphocytes greater than 50%, and atypical lymphocytes greater than 10%.

The clinical history and physical examination are critical components of the evaluation of patients with suspected IM, and knowledge of the clinical signs and symptoms associated with IM would help physicians to decide whether diagnostic testing for IM is necessary. Based on our findings regarding the measurement of DORs for the overall discrimination, the presence of splenomegaly (LR+: 2.39, 95% CI: 1.11 to 5.51), palatal petechiae (LR+: 1.32 to 11.40), posterior cervical lymphadenopathy (LR+: 3.16, 95% CI: 1.45 to 5.20), and axillary or inguinal cervical lymphadenopathy (LR+: 3.05, 95 CI: 1.85 to 4.70) were moderately useful for ruling in IM. Therefore, patients presenting with one of these clinical signs could prompt physicians to

order a diagnostic test for IM. Most of the clinical findings have limited diagnostic value in ruling out the disease; only the absence of any lymphadenopathy (LR-: 0.37, 95% CI: 0.20 to 0.67) was moderately helpful to rule out IM. Therefore, physicians should not rely on the absence of any individual symptom or clinical sign for ruling out IM.

The hematologic parameters based on the laboratory test were found to be more accurate in diagnosing IM compared to the patients' symptoms and signs. Lymphocytes greater than 4×10^9/L, monocytes greater than 1×10^9/L, leukocytes greater than 5×10^9/L or 10×10^9/L, and higher percentages of lymphocytes and atypical lymphocytes all significantly increase the likelihood of IM. We also found that the combination of lymphocytes greater than 50% and atypical lymphocytes greater than 10% (LR+: 50.40, 95% CI: 8.43 to 162.00) are useful to rule in disease.

Clinical implications

Compared to the previous review,[35] we found additional studies that addressed lymphadenopathy, sore throat, headache, and splenomegaly, allowing us to make a more precise summary estimate of the accuracy for these symptoms and signs.

A previous study[24] found that children less than 4 years with IM are more likely to have rashes, abdominal pain, and neurologic problems than adolescents and older adults. For adolescents, the most common symptoms of IM include headache, sore throat, and fatigue, and they were less likely to have diarrhea, rashes, and jaundice compared to other age groups.[34] The adults with IM were more likely to have fatigue and sore throat and less likely to have myalgia or arthralgia.[35] Older people are more likely to have fatigue and body pain and rarely have a sore throat.[36] Acute infection of EBV for young children is usually either asymptomatic or expressed by mild upper respiratory tract symptoms, while for those older patients (aged greater than 40 years), pharyngitis and cervical lymphadenopathy are less frequent, and myalgia is prominent.[37,38] However, there were not enough prospective cohort studies for the diagnosis of

IM in older adults; therefore, we did not conduct a stratified meta-analysis on the symptoms and signs based on the different age groups.

The pre-test probability of IM is approximately 8% for adolescents aged 16 to 20 years with sore throat according to an Australian study.[97] Based on our findings, the presence of splenomegaly (LR+: 2.39) would increase the probability of IM from 8% to 17%, and posterior cervical lymphadenopathy (LR+: 3.16) would increase the probability to 21%. For hematologic parameters, a lymphocyte greater than 40% is an important diagnostic finding (LR+: 5.31), and would increase the probability of IM to 32%. An atypical lymphocyte greater than 40% (LR+: 50.26), and the combination of lymphocytes greater than 50% and atypical lymphocytes greater than 10% (LR+: 50.40) would increase the probability of IM to 81%. We found only three studies[9,13,44] that assessed the accuracy of a combinations of the lymphocyte and atypical lymphocyte percentages; no other studies were found that investigate the diagnostic accuracy of a combination or combinations of signs, symptoms and hematologic parameters.

Strength and Limitation

The strength of our study is that we included a comprehensive literature search and used a contemporary bivariate meta-analysis in the study. We found additional studies that addressed diagnostic accuracy compared to a previous review and were able to provide a more precise summary estimation.

However, there are several limitations in this review. First, most studies did not specify the duration of symptoms at the time patients were enrolled. Since the clinical findings varied by the progression of disease, it would be useful if the accuracy of the clinical findings were stratified by the duration from the onset of symptoms and signs. Also, the severity of the symptoms and clinical signs at the point patients were enrolled could influence the clinicians' interpretation, and most studies did not clearly describe the inclusion criteria and the course of the illness at the time of enrollment, which might have biased the results. Second, the inclusion criteria for patients' enrollment are varied among studies, and there was significant

heterogeneity in the population, study settings, and the choice of reference standard tests across studies, which led to relatively large confidence intervals for the estimates of accuracy. Furthermore, previous studies have suggested that the diagnostic accuracy of the symptoms and clinical signs varied by age and gender, but there were not enough studies to perform a stratified analysis by demographic information. Third, some of the typical symptoms of IM, such as sore throat or pharyngitis, were considered as the inclusion criteria for the study population, and this might have biased the estimation of the individual value of these symptoms and signs for the diagnosis of IM due to incorporation bias.

Conclusion

In conclusion, while the symptoms, signs, and hematologic parameters are important for the clinical diagnosis, only a few key signs and symptoms are significantly associated with IM. The likelihood of IM is appreciably increased by the presence of splenomegaly, palatal petechiae, posterior adenopathy, and axillary or inguinal adenopathy, while it is significantly decreased by any lymphadenopathy. Hematologic parameters are more useful in diagnosing IM, and an absolute lymphocyte count over 4×10^9/L, monocytes over 1×10^9/L, leukocytes over 5×10^9/L, as well as a higher percentage of lymphocytes and/or atypical lymphocytes are helpful for diagnosing IM. Well-designed prospective studies are needed to investigate the role of the clinical findings and office-based tests among patients within a week of disease onset, as well as to evaluate the natural history of IM to help clinicians better understand the course of the disease.

Chapter 4 Tables and Figures

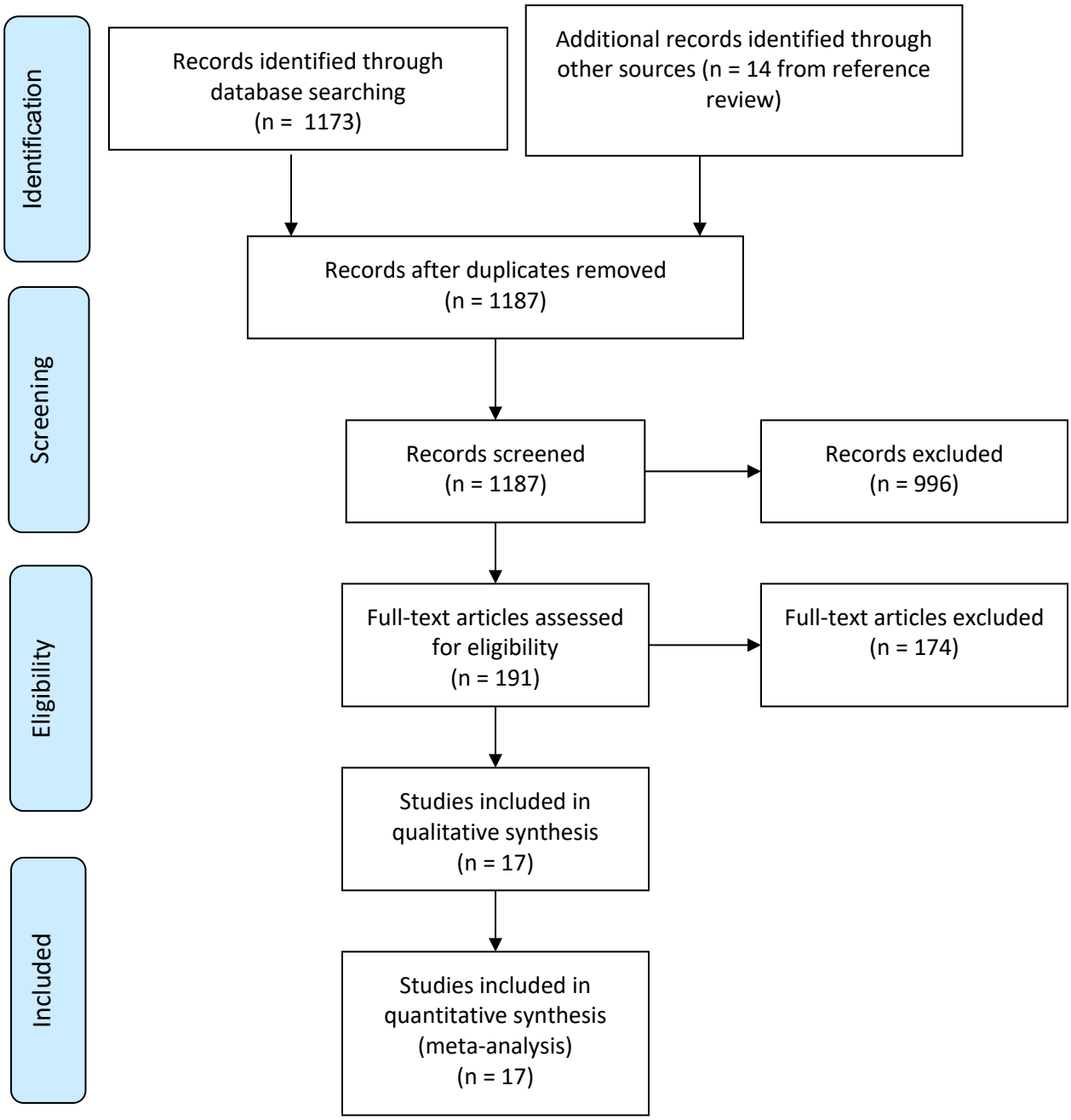


Figure 4.1 PRISMA flow diagram describing the search process

Table 4.1 Characteristics of included studies

Author, Year	Design	Number	Inclusion Criteria	Age	Gender	Reference Standard Test (Prevalence of IM or EBV infection)	Country	Year(s) patients recruited
Balfour, 2005[85]	Prospective cohort	25	College students who aged >=18 years who had a clinical diagnosis of IM and presented within 10 days after the onset of symptoms were included. Students were excluded if they had acute infection consistent with IM or if they were immunocompromised.	Mean 21 years	72% female	EBV profile was tested by EIA (80%)	US	2002-2004
Biggs, 2013[46]	Retrospective cohort	726	Retrospective patients presenting with sore throat, fever, and lymphadenopathy and undergoing Monospot test at university health service were included in the study. Patients without full blood count results were excluded from the study.	Positive group: Mean 21 years Negative group: Mean 30 years	NA	Monospot test (Heterophile antibody test) (6.9%)	UK	2011-2012
Grotto, 2003[91]	Prospective cohort	590	Young adults who were clinically suspected with IM and reported diagnosis of clinical IM for physical and lab examinations at military lab were included. Patients with intermediate EBV or CMV IgM were excluded from the study.	NA	40% female	Mono-Latex (heterophile antibody test) and detection of IgM by ELISA kit (64.4%)	Israel	1988-1991

Hossain, 1989[90]	Prospective cohort	38	Patients aged 5 to 32 years presenting with symptoms suggestive of IM were studied. The minimal criteria included temperature greater than 37C, greater than 50% lymphocytes and monocytes, and at least 10% atypical lymphocytes.	NA	NA	IM Quick test to detect heterophile antibodies and EBV profile was tested by indirect/anticomplement immunofluorescence assay (IFA) (76.3%)	Saudi Arabia	NA
Llor, 2012[92]	Prospective cohort	144	Consecutive adults aged over 14 years with sore throat and four Centor criteria — tonsillar exudate, fever, lymph glands tenderness and absence of cough — and negative pharyngeal testing for group Aβ were recruited.	Mean 24 years	53% female	OSOM MonoTest (Genzyme) as immunochromatographic technology to detect EBV profile, then Paul-Bunnell test (10.9%)	Spain	2006-2009
Lennon, 2010[96]	Retrospective cohort	1000	1000 patients with tonsillitis both in an outpatient and inpatient setting were analyzed to compare the L/WCC ratio in 500 positive and 500 negative Monospot test results.	NA	68% female	Monospot test (Heterophile antibody test) (50%)	Ireland	NA
Sumaya, 1985[86]	Case series	113	Pediatric patients aged 16 years and younger who had clinical manifestations consistent with IM (fever, tonsillopharyngitis, cervical adenopathy, hepatomegaly, splenomegaly) were included in the study. Patients with WBC>=50% or >=5000 lymphocytes and at least 10% atypical lymphocytes were also included.	Median 4 years	NA	EBV profiles were tested by indirect/anticomplement immunofluorescence tests (IFA), and rapid slide tests to detect heterophile antibodies (100%)	US	1976-1982

Ventura, 2004[87]	Retrospective cohort	147	Consecutive patients with clinical features that were suspected with IM and had heterophile antibody tests ordered were included in the study.	Mean 20 years	58% female	Mono-Latex slide (Heterophile antibody test) (46.3%)	US	NA
Rea, 2001[32]	Case series	150	Patients aged 16 years or older with EBV infection, had a positive heterophile antibody test, and reported the onset of symptoms within 2 weeks of the test ordered were included in the study. Those suffering chronic, disabling medical condition, or having been treated with steroid were excluded.	Mean 22 years	48% female	All patients had positive heterophile antibody test; IFA was used to detect VCA-IgM and VCA-IgG (100%)	US	NA
Brigden, 1999[44]	Prospective cohort	181	Sera was obtained from patients with clinical diagnoses of mononucleosis who subsequently tested positive for heterophile antibody using Monosticon test and from 181 patients with clinical suspected IM who tested negative for heterophile antibody test.	Mean 21 years	59% female	Monosticon Dri-Dot test (heterophile antibody test) (50%)	Canada	NA
Ginsburg, 1977[88]	Prospective cohort	43	Children aged 1 to 13 years were selected based on the symptoms and signs compatible with IM. All of them had positive Monospot reactions and showed atypical lymphocytes, and their serum was collected during the acute phase at pediatric outpatient clinic.	Mean 7.7 years	37% female	EBV profile were tested by indirect/anticomplement immunofluorescence test (69.8%)	US	1974-1975

Fleisher, 1983[13]	Prospective cohort	500	Sera from consecutive patients seeking treatment at university health service with illness suggestive of IM were included for EBV-specific serological test. WBC and differential counts were performed uniformly during weekdays and sporadically at other times.	NA	NA	EBV profile were tested by indirect/anticomplement immunofluorescence test (28.3%)	US	1980-1981
Krabbe, 1981[93]	Prospective cohort	43	Consecutive hospitalized children aged between 6 months and 7 years were selected if displaying one of the following: nonbacterial pharyngitis or tonsillitis, lymphadenopathy, hepatosplenomegaly, rash or unknown etiology, a blood count with more than 10% atypical lymphocytes, or thrombocytopenia.	Below 7 years	NA	VCA-IgM and IgG were tested by indirect immunofluorescence test (18.6%)	Denmark	1981
Chretien, 1977[89]	Case series	150	Patients aged between 17 and 29 years diagnosed with IM from university health service were included in the study. The diagnoses were based on usual clinical criteria for the presence of atypical lymphocytes on peripheral blood smears and a positive heterophile antibody tests.	NA	36.7% female	Monospot test (heterophile antibody test) (100%)	US	NA

Aronson, 1982[9]	Prospective cohort	709	Consecutive ambulatory adult patients aged between 16 and 73 years presenting sore throat or strep throat from four primary care settings were included in the study. The extensive clinical data and heterophile antibody tests were obtained.	Mean 32 years	60% female	Monospot test (heterophile antibody test) (2.1%)	US	1976-1977
Gartzonika, 2012[95]	Case series	118	Sera from patients aged between 1 and 47 years with a clinical suspicion of IM or acute EBV infection were included in the study.	Median 21 years	51% female	EBV profile tested using ELISA test and Cellognost-Mononucleosis test to detect heterophile antibodies (100%)	Greece	NA
Ho-Yen, 1981[94]	Prospective cohort	61	Sera from patients with suspected IM and having positive PBD test results were included in the study.	Mean 17 years	50% female	Paul-Bunnel Davidsohn (PBD) test (heterophile antibody test) (61%)	UK	NA

Table 4.2 Quality assessment of included studies

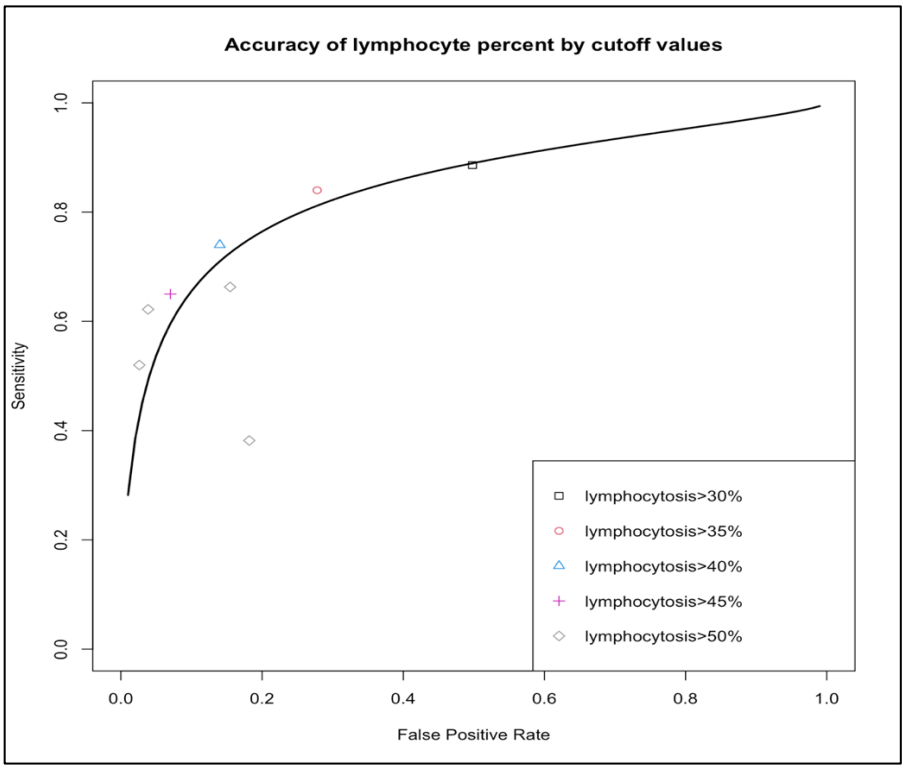
		Patient Selection					Index test				Reference std				Flow & Timing				Overall
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
Study, year	**PubMed ID**	Consecutive	Not case control or retrospective-cohort	Exclusion criteria	Risk of bias	Applicability	Index blinded	Threshold pre-specified	Risk of bias	Applicability	EBV-specific test confirmed by heterophile antibody test or the commercial kits of the heterophile antibody test	Reference blinded	Risk of bias	Applicability	All got reference std	All had same ref std	All accounted for	Risk of bias	L = 0, M = 1, and H = 2+ with high likelihood of bias
Balfour, 2005	16206064	Y	Y	Y	L	L	N	Y	H	L	N	U	H	L	Y	Y	Y	L	H
Biggs, 2013	23918629	Y	Y	Y	L	L	U	Y	L	L	Y	U	L	L	Y	Y	Y	L	L
Grotto, 2003	12948368	Y	Y	Y	L	L	U	Y	L	L	Y	U	L	L	Y	Y	Y	L	L
Hossain, 1989	2547087	Y	Y	Y	L	L	U	U	H	L	Y	U	L	L	Y	Y	Y	L	M
Llor, 2012	22248336	Y	Y	Y	L	L	U	Y	L	L	Y	U	L	L	Y	Y	Y	L	L
Lennon, 2010	21108750	Y	Y	Y	L	Y	U	Y	L	L	Y	U	L	L	Y	Y	Y	L	L
Sumaya, 1985	2987784	N	Y	Y	H	L	U	Y	L	L	Y	U	L	L	Y	Y	Y	L	M
Ventura, 2004	15282662	Y	Y	Y	L	L	U	U	H	L	Y	U	L	L	Y	Y	Y	L	M
Rea, 2001	11458965	Y	Y	Y	L	L	U	Y	L	L	Y	U	L	L	Y	Y	Y	L	L
Brigden, 1999	10506437	Y	Y	Y	L	L	U	Y	L	L	Y	U	L	L	Y	Y	Y	L	L
Ginsburg, 1977	189099	Y	Y	Y	U	L	U	U	H	L	Y	U	L	L	Y	Y	Y	L	M
Fleisher, 1983	6304142	Y	Y	Y	L	L	U	Y	L	L	Y	U	L	L	Y	Y	N	H	M
Krabbe, 1981	6258487	Y	Y	Y	L	L	U	Y	L	L	Y	U	L	L	Y	Y	Y	L	L
Chretien, 1977	191934	U	Y	Y	L	L	N	Y	H	L	Y	N	H	L	Y	Y	Y	L	H
Aronson, 1982	6895981	Y	Y	Y	L	L	U	Y	L	L	Y	U	L	L	Y	Y	Y	L	L
Gartzonika, 2012		U	Y	Y	L	L	U	Y	L	L	Y	U	L	L	Y	Y	Y	L	L
Ho-Yen, 1981	6309974	U	Y	Y	L	L	U	N	H	L	Y	N	H	L	Y	Y	N	H	H

Table 4.3 Diagnostic accuracy for individual elements of the clinical symptom, sign, and hematologic parameter sort by positive likelihood ratios (LR+) within each category. Where the LR+, negative likelihood ratio (LR-) or diagnostic odds ratio differed significantly from 1.0, the value is shown in bold face.

	Studies (patients)-case series	Studies (patients)-cohort studies	Sensitivity (95% CI)	Specificity (95% CI)	LR+ (95% CI)	LR- (95% CI)	Diagnostic odds ratio (95% CI)	AUC
Symptoms								
Nausea or vomiting	2 (262)[32,89]	3 (326)[34,85,92]	0.3 (0.22, 0.39)	0.72 (0.34, 0.93)	1.88 (0.54, 6.04)	0.99 (0.63, 1.24)	2.43 (0.25, 9.30)	0.40
Headache	2 (262)[32,89]	4 (1036)[9,34,85,92]	0.59 (0.40, 0.76)	0.37 (0.23, 0.54)	**1.19 (1.01, 1.45)**	**0.72 (0.50, 0.98)**	**1.72 (1.03, 2.77)**	0.59
Sore throat	2 (262)[32,89]	5 (474)[34,85,92-94]	0.81 (0.68, 0.90)	0.25 (0.17, 0.35)	**1.12 (1.01, 1.25)**	**0.67 (0.41, 1.01)**	**1.78 (1.00, 2.98)**	0.52
Malaise or fatigue	3 (472)[32,89,95]	5 (1137)[9,34,85,92,94]	0.72 (0.59, 0.82)	0.24 (0.11, 0.43)	1.02 (0.91, 1.20)	0.99 (0.67, 1.48)	1.09 (0.62, 1.74)	0.57
Loss of appetite	0	2 (883)[9,34]	0.47-0.74	0.24-0.64	0.86-1.86	0.64-1.54	0.57-2.64	
Cough	1 (140)[32]	1 (709)[9]	0.22-0.40	0.36 (0.32, 0.40)	0.63 (0.31, 1.01)	1.66 (0.99, 2.28)	0.43 (0.14, 1.02)	
Abdominal pain	0	2 (197)[34,85]	0.33-0.38	0.00-0.56	0.58-1.17	0.90-1.50	0.40-1.31	
Myalgia or arthralgia	2 (262)[32,89]	2 (303)[34,92]	0.23 (0.09, 0.49)	0.39-0.59	0.45-1.35	0.47-1.39	0.34-2.76	
Signs								
Lymphadenopathy								
Posterior cervical	2 (253)[32,86]	1 (709)[9]	0.67 (0.51, 0.80)	0.87 (0.84, 0.89)	**3.16 (1.45, 5.20)**	**0.68 (0.41, 0.93)**	**5.18 (1.55, 12.6)**	
Axillary or inguinal	1 (113)[86]	1 (632)[9]	0.23 (0.09, 0.47)	0.82-0.91	**3.05 (1.85, 4.70)**	**0.67 (0.36, 0.91)**	**4.97 (2.05, 10.5)**	
Anterior cervical	2 (253)[32,86]	1 (709)[9]	0.74 (0.59, 0.85)	0.43 (0.39, 0.47)	1.27 (0.80, 1.58)	0.65 (0.25, 1.28)	2.47 (0.63, 6.17)	
Any	4 (445)[32,86,89,95]	6 (1014)[9,34,88,90,93,94]	0.93 (0.86, 0.97)	0.21 (0.07, 0.49)	**1.26 (1.05, 1.65)**	**0.37 (0.20, 0.67)**	**3.77 (1.61, 7.55)**	0.81
Hepatomegaly	1 (140)[32]	5 (971)[9,34,88,90,93]	0.32 (0.07, 0.75)	0.84 (0.18, 0.99)	2.42 (0.95, 6.36)	0.78 (0.48, 1.10)	3.25 (0.84, 8.34)	0.63
Splenomegaly	2 (262)[32,89]	5 (972)[9,34,88,90,93]	0.45 (0.20, 0.73)	0.74 (0.30, 0.95)	**2.39 (1.11, 5.51)**	**0.66 (0.50, 0.84)**	**3.63 (1.38, 7.77)**	0.65
Palatal petechiae	1 (122)[89]	2 (838)[9,92]	0.14 (0.06, 0.28)	0.94-1.00	**1.32-11.4**	**0.57-0.94**	**1.48-155**	

Exudate								
Tonsillar	1 (122)[89]	2 (747)[9,90]	0.47 (0.30, 0.64)	0.78-0.84	**1.39-4.13**	**0.23-0.93**	**1.49-17.0**	
Pharyngeal	0	2 (752)[9,88]	0.13-0.50	0.54-0.93	0.72-4.35	0.62-1.29	0.57-5.39	
Fever								
Measured fever >37.5C	3 (332)[32,89,95]	4 (938)[5,6,14,17]	0.64 (0.37, 0.84)	0.46 (0.17, 0.79)	1.20 (0.91, 1.84)	0.88 (0.63, 1.33)	1.45 (0.67, 2.68)	0.56
Subjective fever	0	2 (122)[14,24]	0.67 (0.48, 0.82)	0.40 (0.24, 0.60)	1.14 (0.85, 1.60)	0.85 (0.49, 1.32)	1.50 (0.64, 3.15)	
Jaundice	0	1 (177)[34]	0.17 (0.10, 0.25)	0.79 (0.68, 0.88)	0.85 (0.45, 1.51)	1.06 (0.91, 1.23)	0.83 (0.37, 1.65)	
Diarrhea	0	1 (175)[34]	0.15 (0.09, 0.23)	0.79 (0.67, 0.89)	0.79 (0.39, 1.41)	1.07 (0.94, 1.26)	0.76 (0.31, 1.50)	
Rash	4 (445)[32,86,89,95]	4 (323)[34,85,88,93]	0.12 (0.06, 0.21)	0.75 (0.59, 0.86)	0.48 (0.18, 1.12)	1.24 (0.98, 1.66)	0.41 (0.11, 1.15)	0.22
Rhinorrhea	0	2 (838)[9,92]	0.14-0.47	0.37-0.83	0.40-1.53	0.85-2.20	0.25-1.78	
Pharyngitis	2 (210)[32,95]	0	0.94 (0.68, 0.99)	-	-	-	-	
Hematologic parameters								
>50% lymphocytes and >10% atypical lymphocytes	0	3 (1361)[9,13,44]	0.45 (0.29, 0.62)	0.99 (0.92, 1.00)	**50.4 (8.43, 162)**	**0.58 (0.38, 0.76)**	**81.2 (19.1, 216)**	0.82
Atypical lymphocytosis (%)								
>40%	0	1 (362)[44]	0.25 (0.19, 0.32)	1.00 (0.98, 1.00)	**50.3 (38.6, 64.1)**	**0.75 (0.68, 0.82)**	**355 (7.43, 622)**	
>20%	0	1 (362)[44]	0.56 (0.49, 0.64)	0.98 (0.94, 0.99)	**28.1 (9.68, 61.4)**	**0.45 (0.38, 0.52)**	**63.9 (10.5, 148)**	
>10%	0	5 (888)[13,34,44,87,93]	0.55 (0.38, 0.70)	0.94 (0.91, 0.96)	**8.97 (3.39, 19.5)**	**0.48 (0.31, 0.65)**	**19.3 (6.48, 44.4)**	0.83
Lymphocytosis (>4×109/L lymphocytes)	0	3 (1235)[44,46,87]	0.59 (0.27, 0.84)	0.94 (0.93, 0.96)	**10.2 (4.79, 16.0)**	**0.44 (0.16, 0.75)**	**30.6 (6.53, 99.4)**	0.64
Lymphocytosis (%)								

>45%	0	1 (1000)[96]	0.65 (0.61, 0.69)	0.93 (0.90, 0.95)	**9.46 (6.76, 13.10)**	**0.38 (0.34, 0.42)**	**25.2 (16.5, 36.8)**	
>50%	0	4 (1740)[13,34,44,96]	0.56 (0.46, 0.65)	0.93 (0.84, 0.97)	**8.52 (2.86, 19.90)**	**0.49 (0.36, 0.63)**	**18.7 (4.83, 51.1)**	0.76
>40%	0	1 (1000)[96]	0.74 (0.70, 0.78)	0.86 (0.83, 0.89)	**5.31 (4.24, 6.68)**	**0.3 (0.26, 0.35)**	**17.7 (12.8, 24.3)**	
>35%	0	1 (1000)[96]	0.84 (0.80, 0.87)	0.72 (0.68, 0.76)	**3.02 (2.63, 3.50)**	**0.22 (0.18, 0.27)**	**13.8 (10.1, 18.8)**	
>30%	0	1 (1000)[96]	0.88 (0.85, 0.91)	0.5 (0.46, 0.55)	**1.78 (1.62, 1.97)**	**0.23 (0.18, 0.30)**	**7.94 (5.63, 10.8)**	
Monocytosis (>1×10^9/L monocytes)	0	2 (1088)[44,46]	0.14-0.72	0.89-0.95	**1.46-7.89**	**0.11-0.98**	**1.49-55.3**	
Atypical lymphocytosis (>1×109/L atypical lymphocytes)	0	1 (38)[90]	0.93 (0.76, 0.98)	0.11 (0.01, 0.48)	1.11 (0.82, 1.78)	1.20 (0.07, 5.69)	4.16 (0.15, 25.2)	
Leukocytosis								
>10×10^9/L	0	4 (1281)[34,44,46,88]	0.43 (0.37, 0.49)	0.83 (0.76, 0.88)	**2.55 (1.58, 3.99)**	**0.68 (0.56, 0.82)**	**3.85 (1.93, 7.01)**	0.63
>5×10^9/L	0	2 (193)[34,88]	0.93-0.94	0.15-0.21	**1.05-1.36**	**0.15-0.70**	**1.51-8.73**	
Neutrophilia (>7.5×10^9/L neutrophils)	0	1 (726)[46]	0.02 (0.00, 0.13)	0.88 (0.85, 0.90)	0.25 (0.02, 1.03)	1.11 (0.99, 1.16)	0.23 (0.02, 1.04)	



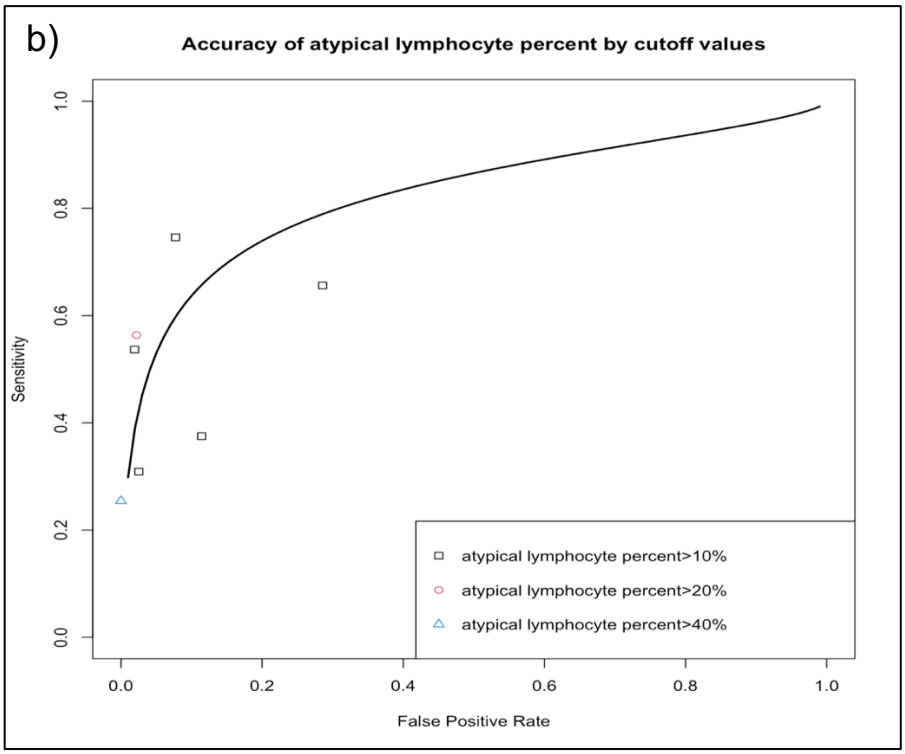
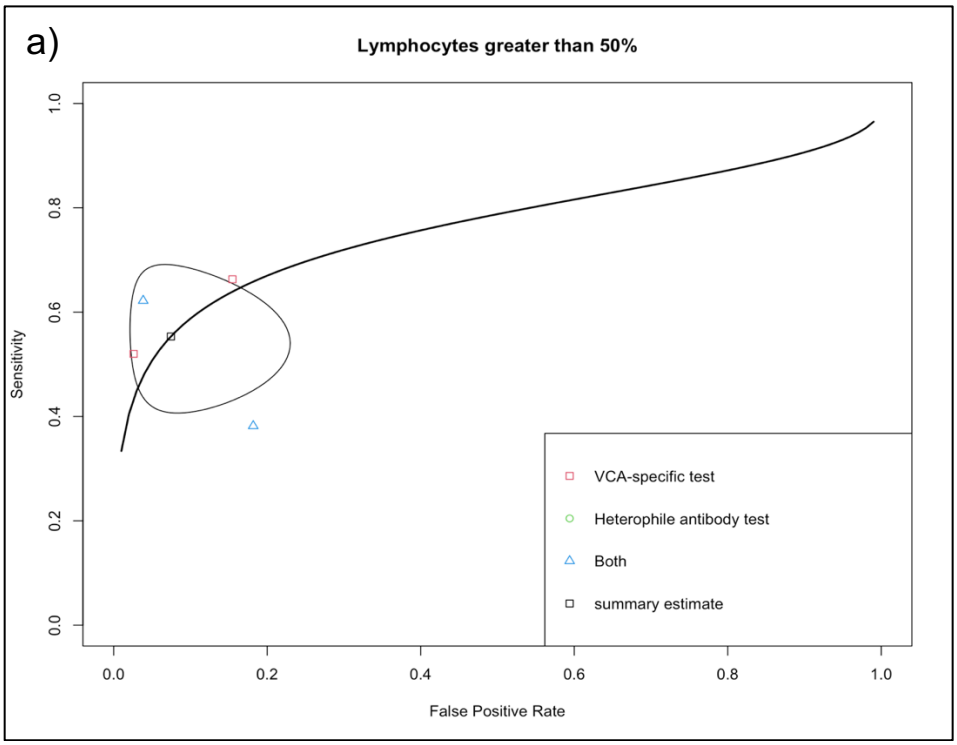


Figure 4.2 Receiver operative characteristic curve by cutoff values a) lymphocyte percent; b) atypical lymphocyte percent



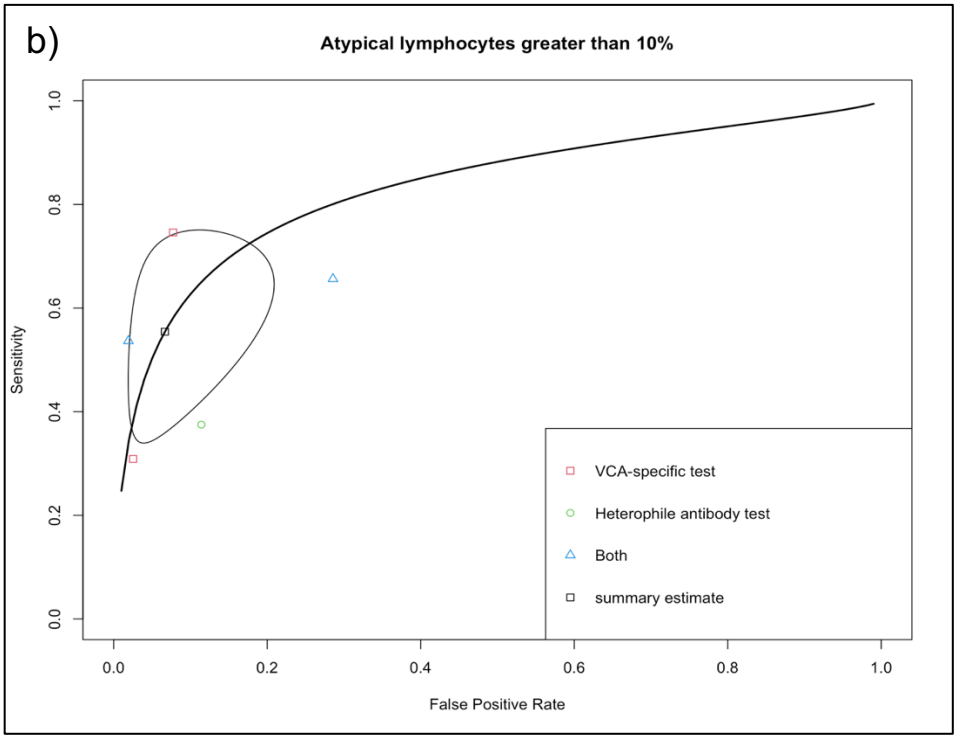
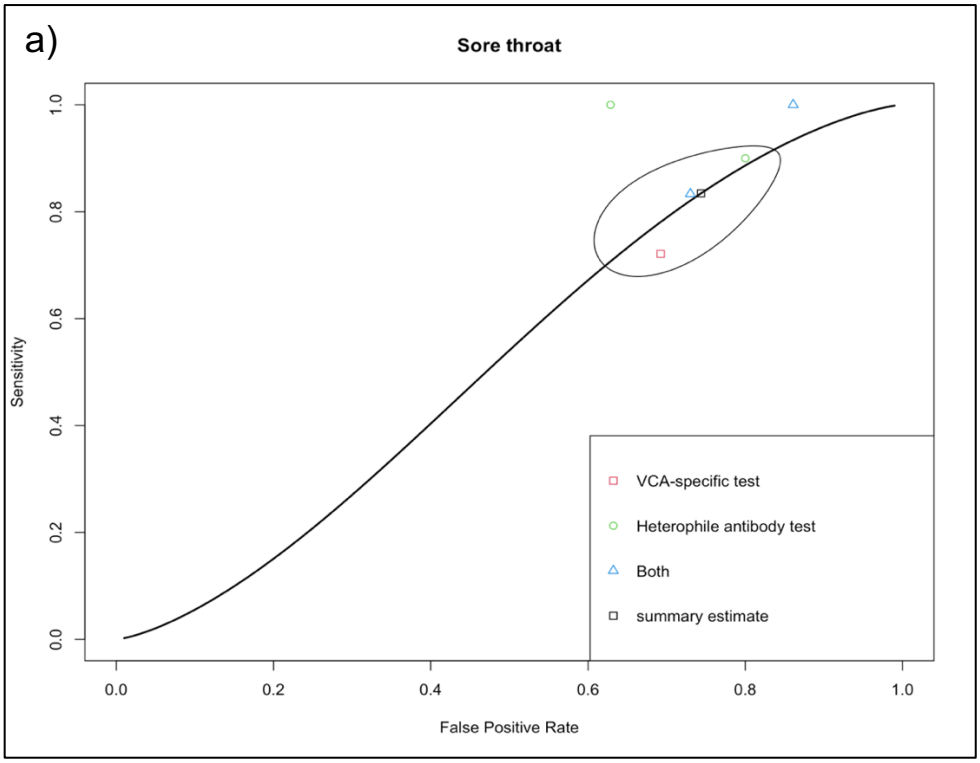
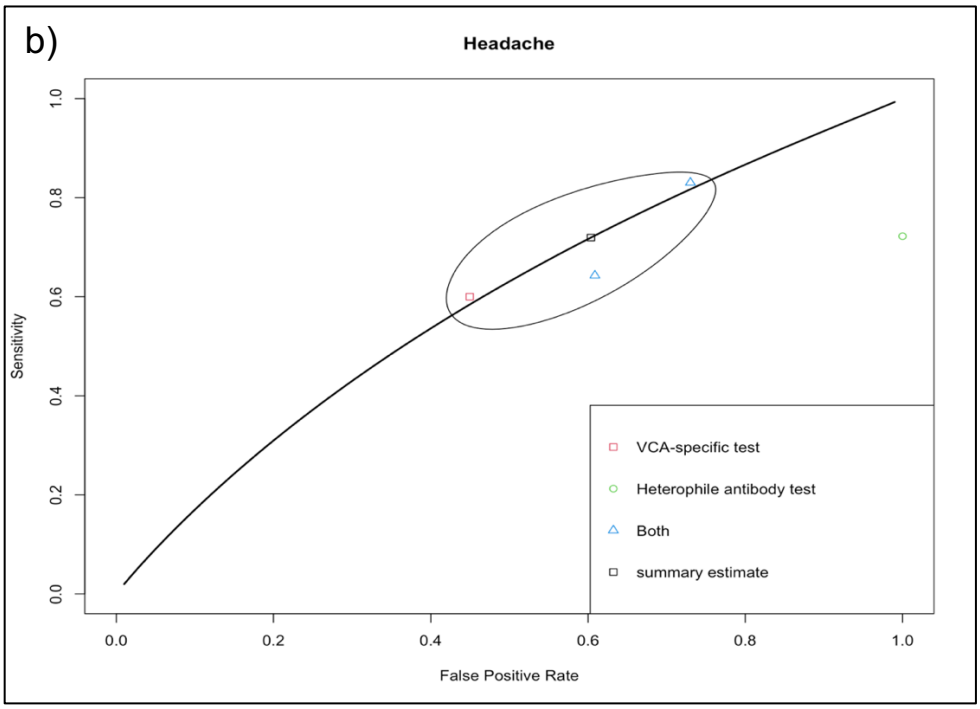


Figure 4.3 Receiver operative characteristic curve by the use of reference standard test a) lymphocytes greater than 50%; b) atypical lymphocytes greater than 10%


a)
Sore throat
Sensitivity
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
VCA-specific test
Heterophile antibody test
Both
summary estimate


b)
Headache
Sensitivity
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
VCA-specific test
Heterophile antibody test
Both
summary estimate

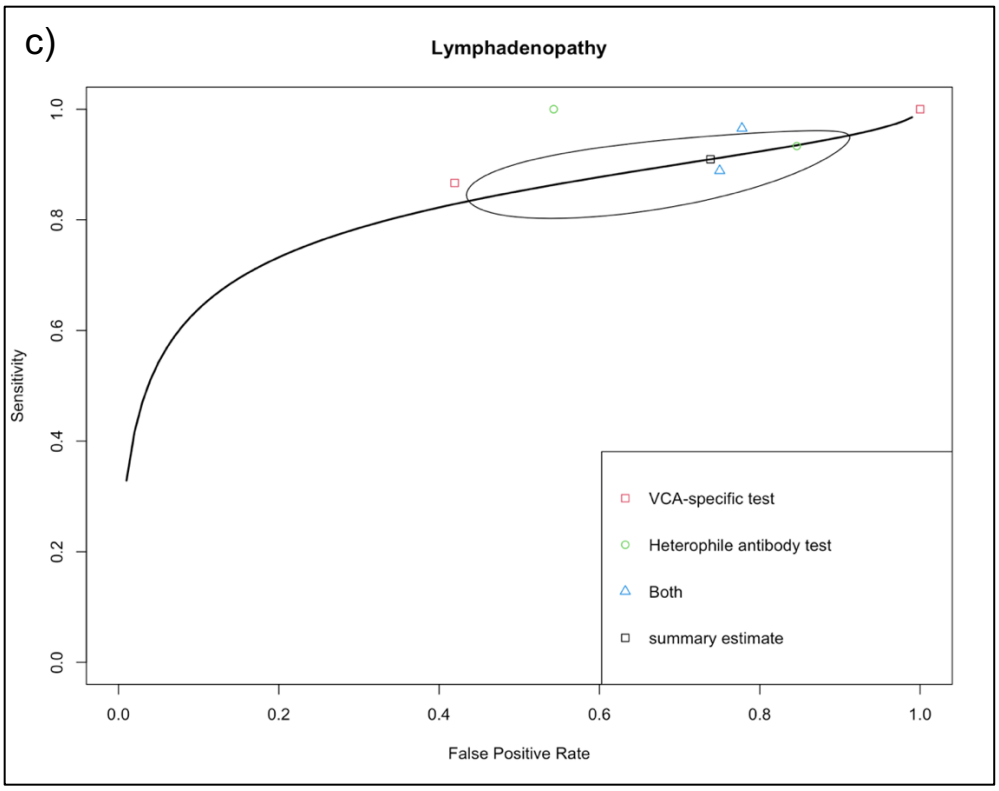
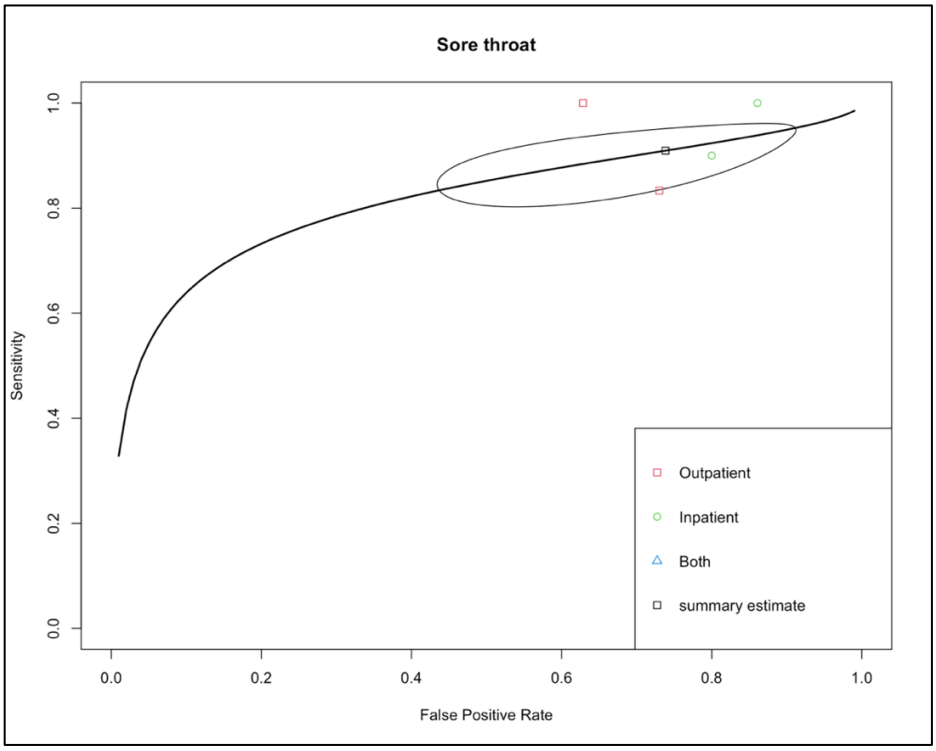
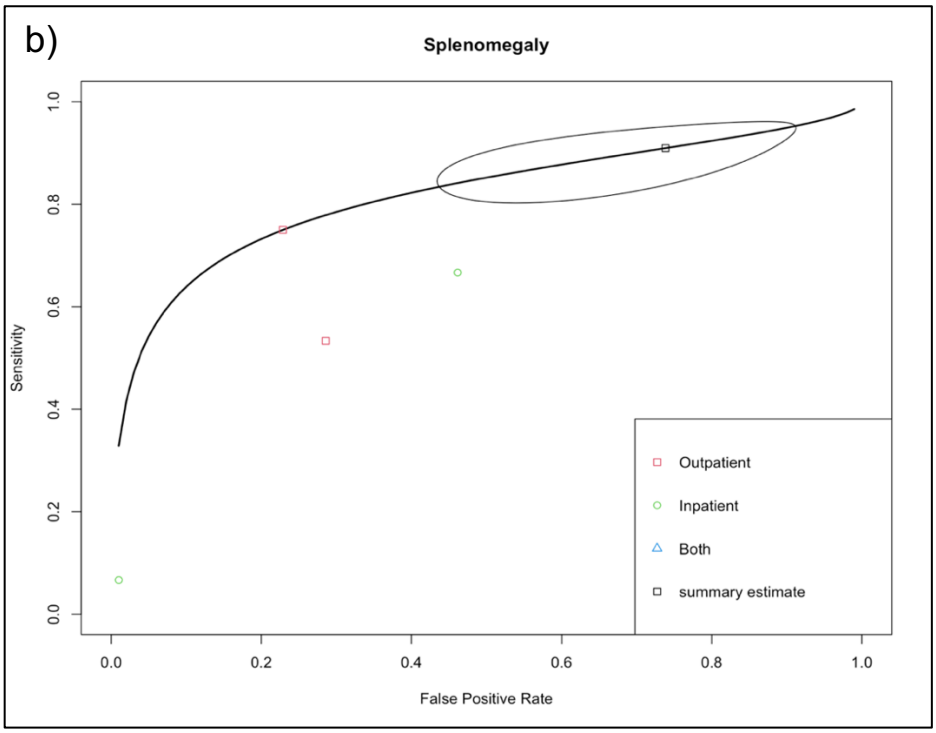


Figure 4.4 Summary ROC curves stratified by the reference standard test for the diagnosis of IM: a) sore throat; b) headache; c) lymph adenopathy.


Sore throat
Sensitivity
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
Outpatient
Inpatient
Both
summary estimate

b)


Splenomegaly
Sensitivity
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
Outpatient
Inpatient
Both
summary estimate

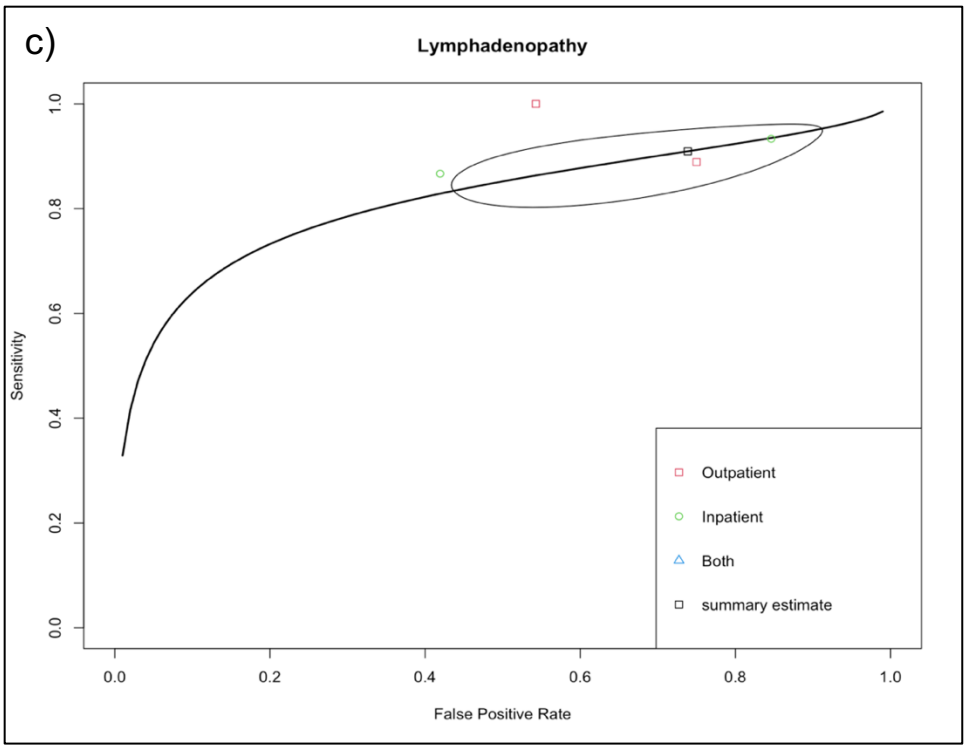


Figure 4.5 Summary ROC curves stratified by patient settings for the diagnosis of IM: a) sore throat; b) splenomegaly; c) lymph adenopathy.

CHAPTER 5

WHAT IS THE BEST ALTERNATIVE TEST TO INDIRECT IMMUNOFLUORESCENCE ASSAY FOR THE DIAGNOSIS OF IM SECONDARY TO EBV INFECTION?[2]

[2] Cai, X., Ebell, M. H., & Haines, L. To be submitted to *the journal of infectious diseases.*

Abstract

Background: No previous study has rigorously compared the diagnostic accuracy of different serological tests to detect patients with primary Epstein-Bar virus (EBV) infection.

Methods: This is a systematic review of the accuracy of EBV immunoassays and heterophile antibody tests among patients with suspected IM that used an immunofluorescent test (IFA) as the reference standard. Two reviewers reviewed all studies in parallel and assessed the quality of the studies using the QUADAS-2 criteria. The summary estimates of the accuracy of each test were calculated by bivariate meta-analysis, which included sensitivity, specificity, likelihood ratios, the diagnostic odds ratios, and the area under the receiver operative characteristic curve.

Results: The overall prevalence of primary EBV infection in the studies was 35% (95% CI: 25% to 45%). Using IFA as the reference standard, the most accurate tests were chemiluminescent microparticle immunoassay (CLIA) (LR+: 17.80, LR-: 0.06, DOR: 367) and enzyme-linked immunosorbent assay (LR+: 19.90, LR-: 0.12, DOR: 192). CLIA could also accurately identify acute and past infections. The heterophile antibody tests were less accurate for ruling out IM when these tests are negative than EBV immunoassays. Both slide agglutination tests (LR+: 7.39, LR-: 0.69, DOR: 11.0) and solid-phase assays (LR+: 6.13, LR-: 0.71, DOR: 9.20) were much less sensitive in children.

Conclusions: Our study found that a positive result of the heterophile antibody test is a strong evidence of IM secondary to primary EBV infection, but that negative results, especially in children, are less reliable and require diagnostic follow-up.

Key words: infectious mononucleosis; Epstein-Barr virus; meta-analysis; test accuracy

Background

Patients with infectious mononucleosis (IM) often present with sore throat, pharyngitis, cervical lymph node enlargement, and fatigue. The incidence of IM is highest among adolescents aged 15 to 24 years (six to eight cases per 1000 person-years).[14] Most IM cases are caused by the Epstein-Barr virus (EBV). EBV is a gamma herpes virus that infects epithelial cells and B lymphocytes.[98] Approximately 50% to 90% of adults are seropositive for EBV.[21]

To diagnose IM caused by EBV infection, the Paul-Bunnell test was developed to detect heterophile antibodies through an agglutination of the erythrocytes from various animals.[10] The latex agglutination test was then modified from the traditional Paul-Bunnell test by adding guinea pig kidney serum absorption to the test, which significantly improves the specificity of the test. Heterophile antibody tests provide a rapid and cost-effective diagnosis of IM, but they are insensitive, particularly in the first weeks of the illness, with a false-negative rate as high as 25%.[7] Moreover, the heterophile antibody tests can detect only 25% to 50% of EBV infections among patients aged younger than 12 years.[12,13]

More sensitive serological immunoassays were then developed to test for antibodies to EBV capsid antigens, which include the viral capsid antigens (VCA)-IgM and VCA-IgG and the EBV nuclear antigen (EBNA)-IgG.[56] These immunoassays are commonly used for routine diagnosis through various techniques, such as enzyme-linked immunosorbent arrays (ELISA), multiplex flow immunoassays, chemiluminescent microparticle immunoassays, or immunoblot based assays. The VCA tests are highly sensitive and have a greater ability to rule out IM when negative compared to heterophile antibody tests.[14] The VCA-IgM appears in the early phase of EBV infection, and it is treated as an indicator of acute infection. However, the VCA-IgM may disappear after a short time. EBNA-IgG is typically not detectable until six to eight weeks after the onset of symptoms and therefore can help distinguish between acute and prior EBV infections.[54,55]

Although many serological tests for IM are available for use in clinical settings, uncertainty prevails about which serological test is the most accurate to detect primary EBV infection. So far, there has been no systematic review or meta-analysis that rigorously compares and evaluates the diagnostic accuracy of different serological tests to identify patients with primary EBV infection. Our objective is to systematically evaluate and compare the diagnostic accuracy of different serologic tests to detect EBV infection in the serum of patients with suspected IM. Currently, an indirect immunofluorescence assay (IFA) is the most widely used technique in the reference laboratories for the detection of EBV-specific antibodies in primary EBV infections and they are commonly used as the reference standard.[99] Performing and interpreting an IFA is both labor intensive and subjective, and also required highly skilled and trained personnel to read the fluorescence images.[100] Therefore, evaluating the diagnostic accuracy of the laboratory immunoassays and the heterophile antibody tests is necessary to help clinicians to find the most appropriate alternative detection methods to reference laboratory IFA test for the detection of EBV infection.

Method

This systematic review and meta-analysis followed the preferred reporting items for systematic reviews and meta-analyses (PRISMA) protocol as checklist to report findings.[84]

Inclusion criteria

This meta-analysis includes published studies that performed original data collection. The inclusion criteria were as follows: prospective or retrospective cohort studies; provided sufficient information to estimate both sensitivity and specificity for the accuracy of at least one serological test to detect primary EBV infection; and sera samples were drawn from patients with suspected EBV infection or IM within any age group. We limited our index tests to two categories: heterophile antibody tests to detect the presence of heterophile antibodies; serological immunoassays to detect VCA IgM, VCA IgG with or without EBNA IgG.

The reference standard tests for included studies were limited to IFA for the detection of VCA IgM, VCA IgG and EBNA IgG. The same reference standard test should be performed on all included sera to avoid the verification bias. No stipulation as to language, country, or year of publication for included studies was made.

We excluded studies that were case reports, comments, editorials, case-control studies, modeling studies, or studies with fewer than 5 sera samples. We also excluded studies if a majority of the subjects were immunodeficient or presented with another special condition (i.e., with another acute illness consistent with EBV-related IM, transplants recipients, or HIV-positive subjects). Studies of IM caused by toxoplasmosis or cytomegalovirus infections were excluded as well.

Reference standard test

Past studies have used a variety of reference standard tests to confirm either the presence or absence of primary EBV infection, but we were unclear at the start of the review about exactly what reference standard tests we would encounter. The performance characteristics of EBV-specific assays depend strongly on the substrates, the nature of the antigen, and differing interpretations of assay results by various manufacturers.[99] Antibody detection by IFA tests using EBV-infected lymphocytes was first introduced by Henle,[101] and has been used as the “gold standard” in routine EBV diagnostics until the present.[99] The VCA IgM, VCA IgG and EBNA IgG are EBV-serological markers that are used to determine the serological profiles characterizing the primary infection. The IFA test was used to detect anti-VCA IgM and IgG, and the anticomplement immunofluorescence assay (ACIF) was used to detect anti-EBNA IgG. A primary EBV infection is indicated if there is a positive result for the VCA IgM and/or VCA IgG with no or a weak antibody response to EBNA IgG.[74]

Search strategy and selection criteria

We used PubMed as the primary electronic source of records reviews and searched for the publications through 30 September 2020, with no restrictions on language. The search

concepts used in our study include: “diagnostic tests”; “infectious mononucleosis or Epstein-Barr virus”; and “accuracy of diagnosis” linked by the Boolean operator “AND”. We also limited the search to human subjects. The complete search strategy for the PubMed database was described in the method section. We also manually searched the relevant articles from the reference lists of the included studies. Only published and peer-reviewed articles were considered in the systematic review.

The titles and abstracts identified using the above search strategy were screened in parallel by two independent reviewers (XC, LH). If there were any discrepancies, a third reviewer (ME) was consulted. The full-text paper for each included abstract was then independently reviewed by the authors.

Assessment of methodological quality

Two reviewers (XC, LH) independently assessed the risk of bias using a quality assessment of diagnostic accuracy studies 2 (QUADAS-2) checklist tailored to this review.[69] The third reviewer (ME) became involved if there was any disagreement. The four key domains of the QUADAS-2 tool include patient selection, evaluation of the index test performance, evaluation of the reference test performance, and flow of patients through the study with corresponding timing. Ideally, studies should either prospectively or retrospectively recruit a representative serum samples that were suspected with EBV infection or to have susceptible IM, either in a primary care or hospital setting. Index tests should be performed and interpreted without knowledge of the reference standard test and vice versa. The reference standard test must correctly classify patients with primary EBV infection, and the same reference standard should be applied to all patients. We defined the quality of each domain as low, unclear, or high risk of bias. The proportion of studies with different degrees of risk is displayed, and the QUADAS-2 instruments adapted for this study and the corresponding questions are shown in Appendix B.

Data abstraction

Two investigators (XC, LH) independently extracted the aggregate data using standardized data entry forms. Any discrepant results were resolved through discussion between the investigators. The following data were extracted from each study: characteristics (country, year of recruitment, setting), populations (mean or median age, gender, sample size, inclusion and exclusion criteria), the index test used (commercial test name, class of immunoglobulin measured), the reference standard test used, and the results from each study to construct 2 by 2 tables. If the absolute numbers were not reported, they were estimated based on the total number of patients, the sensitivity, and the specificity. If more than one commercial test was used in the study, we considered each particular commercial test as an individual study. For example, a study that examined three commercial immunoassays of ELISA from the same population was considered to have 3 studies.

Statistical analysis

We performed random effects meta-analyses of the prevalence of EBV-infected patients or patients with IM stratified by age group using the metaprop procedure in R (version 3.5.2). In the first analysis, we evaluated the diagnostic accuracy for different categories of serologic tests for the detection of primary EBV infection. To evaluate the accuracy of serological immunoassays, all patients testing with a negative result for EBNA IgG and either with positive VCA IgM only or with positive VCA IgM and VCA IgG were considered to present a primary EBV-infection profile for the reference standard IFA test. We also assessed the diagnostic accuracy for the detection of the anti-VCA IgM or anti-VCA IgG alone for the immunoassays being studied using the IFA results as the reference standard. The diagnostic accuracy for the detection of anti-VCA IgM helps to distinguish acutely infected patients from those having had past infections and from seronegative patients. The diagnostic accuracy for the detection of anti-VCA IgG helps to distinguish patients having had past infections from primarily infected and seronegative patients.

We first extracted the numbers needed for the construction of 2 by 2 tables, which compared the diagnostic performance of each index test with the reference standard test. The pooled measures of the diagnostic performance of each element including sensitivity, specificity, and positive (LR+) and negative likelihood ratios (LR-) with 95% confidence intervals (CI), and the diagnostic odds ratio (DOR) were then calculated.[70] The DOR was calculated by LR+ divided by LR-. If there was only one study for an index test, the point estimate and a 95% confidence interval were calculated using the binomial exact method; if there were two studies, the ranges were presented; if there were more than two studies, the summary estimate was calculated based on bivariate analysis.[75]

To describe heterogeneity among studies, we calculated the area under the receiver operating characteristics curve (AUC) for each index test. If the index tests had multiple cutoffs, we also evaluated the presence of threshold effects during the analysis to check whether the sensitivity decreased, and the specificity increased with increasing diagnostic cutoffs. The threshold effect was determined by visually inspecting the receiver operating characteristic (ROC) curves for the cutoffs of the index test. The AUC was calculated as an overall measure of accuracy if a threshold effect existed. To assess design characteristics as potential determinants of diagnostic accuracy, we visually inspected the ROC curves to explore the sources of heterogeneity.

We imported the data into R (version 3.5.2) using the R Studio framework (version 1.1.463). The univariate meta-analysis on a single study or the bivariate analysis on multiple studies for the index tests were performed via the mada package (version 0.5.8).[71] The calculation of AUCs and DORs with 95% CIs were also calculated using this package.

Result

Study Characteristics

There were 1173 abstracts identified in our initial screening (the search was a combined search for this systematic review and a separately reported systematic review of clinical signs

and symptoms). We also identified 14 additional studies from the reference review. After a full-text assessment, a total of 22 studies met our inclusion and exclusion criteria and were used for our final analysis. This detailed search process is summarized in Figure 5.1.

The characteristics of the included studies are summarized in Table 5.1. Among 22 studies, four enrolled only children or adolescents,[102-104] eight enrolled patients of all ages,[12,54,56,105-109] and the rest did not report the patients' ages. Four of the studies were prospective cohort studies,[102,104,106,110] and the remainder were retrospective cohort studies. Five studies were set in the United States,[106,107,110-112] two in Canada,[113,114] one in Australia,[105] one in China,[115] and the remainder were from Europe. The oldest study was published in 1983,[110] and the most recent studies were published in 2018.[104] The index tests being studied included enzyme-linked immunosorbent assays (ELISA), chemiluminescent microparticle immunoassays (CLIA), multiplex follow immunoassays (MFI), immunoblot based assays (IB), immunofiltration assays (IMFA), immunochromatographic assays (ICA), enzyme-linked fluorescent assays (ELFA), and heterophile antibody tests. The heterophile antibody tests included both slide agglutination tests and solid-phase immunoassays.

Quality assessment

Of the 22 studies, one was judged to have a high risk of bias, and the remaining were at moderate risk of bias (n=10) or low risk of bias (n=11). Table 5.2 shows a detailed description of the quality evaluations for each study using the QUADAS tool. Blinding the person performing the index test from the reference standard test or vice versa was explicitly stated in only one study; other studies either did not mask or did not mention about masking process. Given that most of the included studies stated that applying the index tests and the reference standard tests was in accordance with the manufacturer's instructions, the blinding issues were unlikely to have affected the diagnostic accuracy of the index tests. As for the flow and timing domains, most of the studies having a high or moderate risk of bias were due to the omission of inconclusive index test results from the final analysis.

Prevalence of EBV infection

The prevalence of primary EBV infected patients in each study is summarized in Table 5.3. It is stratified by the age groups. In studies enrolling a mix of adults and children with suspected EBV infection, the prevalence ranged from 7.9% to 89.4%, with a pooled prevalence of 36% (95% confidence interval (CI): 25% to 48%). Studies enrolling only children had a prevalence between 14.0% and 40.0%, with a pooled prevalence of 30% (95% CI: 19% to 43%). The overall disease prevalence is among patients with clinically suspected IM is 35% (95% CI: 25% to 45%).

Accuracy of the diagnostic tests for detection of primary EBV infection

The accuracy of index tests for the detection of primary EBV infection stratified by test category is summarized in Table 5.4 (see Appendix D for detailed estimations of each study). Among EBV-specific immunoassays, the most accurate tests to detect primary EBV infection based on the diagnostic odds ratio (DOR) include CLIA (LR+: 17.8, LR-: 0.06, DOR: 367), ELISA (LR+: 19.9, LR-: 0.12, DOR: 192), and IMFA (LR+: 11.4, LR-: 0.09, DOR: 160). The IMFA was evaluated in only one study with a moderate risk of bias;[116] thus, the test accuracy of the IMFA method still needs to be confirmed. The MFI, IB, ICA, and ELFA were fairly specific, but they were less sensitive than ELISA. Therefore, the MFI (LR+: 28.5, LR-: 0.28), IB (LR+: 12.8, LR-: 0.18), ICA (LR+: 35.2, LR-: 0.60), and ELFA (LR+: 7.92 to 33.3, LR-: 0.14 to 0.57) are more accurate for ruling in patients when these tests yield positive results, but they are not ideal for ruling out patients compared to ELISA, CLIA, and IMFA when the test results are negative. The ROC curves for the accuracy of ELISA, CLIA, IB and ICA are shown in Figure 5.2.

The accuracy of heterophile antibody tests is also summarized in Table 5.4. In general, these tests are less accurate than EBV-specific immunoassays. The slide agglutination tests (LR+: 19.3, LR-: 0.27) and the solid-phase assays (LR+: 19.5, LR-: 0.29) have similar abilities to rule in and to rule out the disease for patients in all age groups. Both slide agglutination tests and solid-phase immunoassays were found to have lower sensitivity than immunoassays;

therefore, they are less accurate for ruling out IM when these tests are negative. Furthermore, the slide agglutination tests (LR+: 7.39, LR-: 0.69, DOR: 11.0) and solid-phase assays (LR+: 6.13, LR-: 0.71, DOR: 9.20) tend to be much less sensitive in children. The ROC curves for both forms of heterophile antibody tests to detect primary EBV infection stratified by age group are shown in Figures 5.3.

Accuracy of immunoassays for detection of anti-VCA IgM or anti-VCA IgG

The diagnostic accuracy of the EBV-specific immunoassays using IFA as the reference standard test for detection of the anti-VCA IgM or anti-VCA IgG is summarized in Table 5.5 (see Appendix D for detailed results). Based on these results, we found that CLIA and IB are more accurate in detecting anti-VCA IgM than other immunoassays; therefore, they are helpful for distinguishing early, acute EBV infection. CLIA and ELFA are relatively more accurate in detecting anti- VCA IgG than other tests; hence, these two tests are helpful in identifying previously infected patients.

Through an inspection of the summary ROC curves, no clear patterns of heterogeneity for ELISA were found among different study designs. Since there are limited studies testing children using EBV-specific immunoassays, we did not evaluate the diagnostic accuracy of the EBV immunoassays stratified by age groups.

Discussion

Based on our evaluation of the diagnostic accuracy of different EBV immunoassays, ELISA, CLIA, and IMFA are the most accurate tests to detect primary EBV infection. Instead of using the reference laboratory IFA test for determining anti-EBV profiles, which required approximately 5 to 6 hours and had ambiguous endpoints, the procedure of the ELISA and CLIA tests to determining the same antibody profiles have a shorter turnaround time (4 to 5 hours) with clearer endpoints, and are more suitable for large-scale testing; [117,118] thus, they can potentially be used as an alternative to the reference laboratory IFA method. Since the authors identified only one study that compared the diagnostic accuracy of IMFA with that of the IFA

test, the performance of the IMFA method still needs to be confirmed by well-designed studies. Most of the studies evaluated the immunoassays among patients in any age group, so the accuracy of these tests among children or adolescents remains unclear.

We also compared the EBV immunoassays that are used to detect antibodies against EBV with the standard IFA method. Our results demonstrated that CLIA and IB were in agreement with the reference standard IFA for anti-VCA IgM detection, which help to distinguish early, acute infected patients from those with past infections and seronegative patients.[100] The CLIA and ELFA methods are relatively more accurate for detecting anti-VCA IgG than other tests; thus, these two tests are helpful in identifying past EBV-infected patients from current infected and non-infected patients. The ELISA results were less accurate for detecting anti-VCA IgM or IgG compared to IFA; therefore, we conclude that CLIA may be the most suitable method for the diagnosis of EBV infection at different stages.

Heterophile antibody tests for use at the point of care are commonly used as a diagnostic test for patients with suspected IM in outpatient settings. The rapid heterophile antibody tests required less time and effort to complete, and the kits ranged in price from $2 to $5 per patient.[117] A positive result of the heterophile antibody test is a strong evidence of IM. However, our findings show both slide agglutination and solid-phase immunoassay tests are less sensitive than EBV immunoassays for diagnosing EBV-infected IM, especially in children, thus, they are relatively inaccurate for ruling out disease when the test is negative. This conclusion is consistent with a previous review established by one of the authors.[14] Therefore, if a patient was tested negative for heterophile antibody test and the clinical symptoms were not resolved for a week, a EBV immunoassay may be selected as the confirmatory test. A negative result of a EBV immunoassay is a strong evidence against the diagnosis of IM.

We found one systematic review that demonstrated that PCR may be helpful in the diagnosis of IM secondary to EBV (sensitivity: 77%, 95% CI: 66% to 86%; specificity: 98%, 95% CI: 93% to 100%).[119] Although the PCR tests are relatively specific, they had a lower sensitivity

than the CLIA, IMFA, and ELISA tests and are less accurate in ruling out IM when the PCR tests are negative. Furthermore, this study did not list the reference standard test used for the comparison. Based on our investigation, we did not find any study that compared PCR tests to the IFA test as the reference standard. Thus, the PCR test is not included as one of the index tests in our analysis.

Strengths and limitations

Strengths of the current study include that it is the first systematic review of the EBV immunoassays for the diagnosis of IM, and that our review covered almost all diagnostic techniques that are currently being used in the market. Our study included a comprehensive literature search to identify all the relevant studies for each test and used a contemporary bivariate meta-analysis to evaluate the test accuracies.

Our study also has some limitations. First, our conclusions about the serologic tests are limited by the relatively poor quality of some of the older studies, which often omitted descriptions of key aspects of study design or of the study participants. Second, we compared the summary estimates of accuracy among different study populations, which generated significant unexplained heterogeneity. For example, the differences in the timing of the sampling in relation to the onset of symptoms, which were not clearly stated in most of the studies, might have resulted in differences in sensitivity or specificity. Third, some studies deliberately classified the equivocal results as false positives or false negatives during the statistical analysis, which would lead to the estimated sensitivity and specificity being lower than the true values. Furthermore, most of the studies did not perform the cost-effectiveness analysis or reported the price for the immunoassays, thus we only evaluated the accuracy of each immunoassay without considering their cost-effectiveness.

Implications for the research

Future, well-designed studies to evaluate individual tests for IM that avoid the major limitations of the existing evidence base are needed. This can be accomplished by adhering to

the fundamentals of design for the study of diagnostic tests: a well-defined study population with a prospective, consecutive sampling approach; performance of the index tests and the reference standard test in a standardized and blinded manner; and use of an accurate reference standard tests that is uniform across all participants. To reduce variability in estimates and enhance generalizability, the estimates of the test accuracy should also be stratified by the study setting and by the characteristics of the study population; and the number of days from symptom onset should be documented.

Ultimately, serial testing that combines different serologic tests on the same sample has not been proposed to guide the evaluation of patients with suspected IM, which, if done, may provide a useful diagnostic strategy. A previous study[14] proposed an algorithm for the management of suspected IM that considers the results of the hematologic parameters, the heterophile antibody test, and the VCA IgM test to decide whether IM can be ruled out. Further study is needed to validate this algorithm.

Conclusion

Our study found that EBV immunoassays are generally more accurate for the diagnosis of IM secondary to EBV than heterophile antibody tests. The ELISA, CLIA, and IMFA are the most accurate immunoassays to detect primary EBV infection. The CLIA method could also more accurately detect anti-VCA IgM and anti-VCA IgG compared to the standard IFA method; therefore, we conclude that CLIA would be the most suitable alternative method to IFA for the diagnosis EBV infection. On the other hand, the decisions made regarding serologic testing for IM should be driven not only by the diagnostic accuracy, but by the levels of test complexity, the patient's preferences, the length of time to get a result, and the cost of the test, all of which should play an important role in determining which tests are to be used.

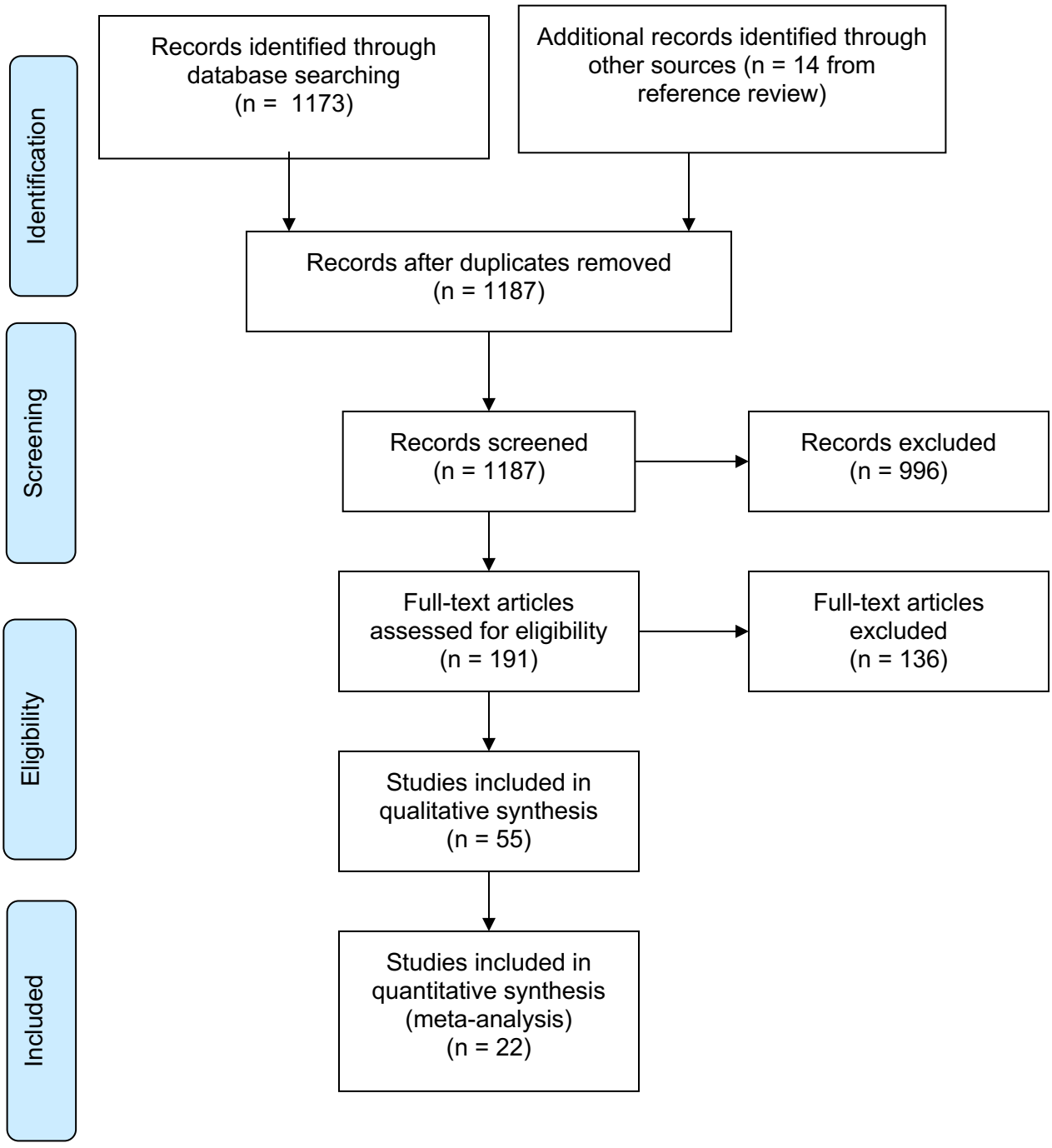


Figure 5.1 PRISMA Flow Diagram Describing the Search Process

Table 5.1 Characteristics of Included Studies

Author, Year	Study Design	Number of patients	Study population	Mean or median age	Gender	Index test	Name of the commercial serological kit	Antibodies detected by the index test	Country
Fleisher, 1983	Prospective cohort	500	Sera from consecutive patients seeking treatment at university health service with illness suggestive of IM were included for EBV-specific serological test. WBC and differential counts were performed uniformly during weekdays and sporadically at other times.	NR	NR	Slide agglutination test	Monospot	Heterophile antibodies	US
Ho, 1989	Both retrospective and prospective cohort	1672	Sera were collected retrospectively and prospectively from patients aged between 3 and 63 years with clinical histories suggestive of a recent EBV infection.	NR	NR	ELISA	Behring-werke AG	IgM	Australia
Samlley, 1990	Retrospective cohort	167	A total of 73 patients who were tested by heterophile antibody test as the reference standard test, and another 94 patients who were tested by the indirect fluorescent test that specifically detected IgG antibodies to EBNA as reference test were included in the study.	NR	NR	ELISA	Monolert (Ortho Diagnostic System)	IgM and IgG	US

Levin, 1991	Prospective cohort	298	Patients and healthy individuals aged 3 to 69 years with suspected IM were examined at health centers and had blood drawn for determination of complete blood count, heterophil antibody, and EBV-specific ELISA antibody levels.	NR	NR	ELISA	Monolert	IgM and IgG	US
Wiedbrauk, 1993	Retrospective cohort	128	Sera from patients aged between 3 and 76 years were collected to evaluate the performance of each EIA kit. The sera were selected from specimens that were sent to the lab for EBV testing.	Mean 30 years	66% female	ELISA	Gull, Incstar, Ortho, Sigma, Bio Whittaker	IgM and IgG	US
Linderholm, 1994	Retrospective cohort	108	Sera from 103 patients aged between 2 and 60 years with clinically suspected IM were included.	Median 19 years	44% female	slide agglutination assays, solid-phase immunoassays	im.absorption (Mercia), Monospot (Ortho), Monosticon Dri-Dot (Organon Teknika), Mono-Plus (Dominion), Monolatex (Biokit); Cards Mono, Cards OS Mono (Pacific Biotech), Preview Mono (Lecco), Monolert (Ortho)	Heterophile antibodies	Sweden
Elgh, 1996	Retrospective cohort	100	Sera were drawn from 50 patients with serological verified primary EBV infected patients and from 47 patients who were either EBV immune or suspectable patients with other viral diseases in 6 months.	Median 20 years	NR	slide agglutination assays, solid-phase immunoassays	Monolatex (Biokit), Mono-Lex (Trinity lab), Mono-Latex (Wampole), Mono-Plus (Wampole), IM-Check (Veda lab), Clearview IM (Unipath)	Heterophile antibodies	Sweden

Svahn, 1997	Retrospective cohort	214	Consecutive serum samples from patients aged between 4 to 84 years, with symptoms that compatible with primary EBV infection were sent to a microbiology lab	median 18 years	NR	ELISA, Heterophile antibody test	Gull, Biotest, Behring, Monolatex (Biotest), Mono-Lex (Trinity lab)	IgM and IgG	Sweden
Chan, 1998	Retrospective cohort	117	Sequential patient sera from hospital lab for the diagnosis of primary EBV infection were used for this study. One sample from a 2-month child was excluded from further analysis because passive transfer of maternal antibody could potentially confound the serological interpretation.	NR	NR	ELISA	NA	IgM	China
Gomez, 2000	Retrospective cohort	350	Sera from patients aged between 1 and 80 years with clinically suspected IM were included in the study.	Median 26 years	43% female	ICA, Latex agglutination tests (heterophile antibody test)	BIFA-MI (Sumilab), Monoslide test (Bio Merieux), IM absorption kit (Microgen Bioproducts)	IgM and IgG	Spain
Fung, 2002	Retrospective cohort	152	Sera from consecutive samples that received comprehensive EBV serology, from both immunocompetent and immunocompromised patients were used in the study.	NR	NR	ELISA	Wampole	IgM and IgG	US
Kleines, 2006	Prospective cohort	349	Sera collected from pediatric in outpatient setting aged 0.5 to 18 years were included in the study.	NR	NR	ELISA	Enzygnost (Dade-Behring)	IgG	Germany
Altuglu, 2007	Retrospective cohort	45	Sera collected from patients aged 1 to 62 years sent to hospital microbiology lab for EBV testing.	Mean 25 years	NA	multiplex flow immunoassay (MFI),	Athena Multi-Lyte (Zeus), Euroline (Euroimmun)	IgM and IgG	Turkey

						immunoblot based assay			
Devanthery, 2010	Retrospective cohort	393	Sera collected from patients suspected of acute or latent EBV infection and had EBV testing by IFA between 1998 and 2005. The sample population also included a mix of specimens from immunocompromised (organ transplant) patients.	NR	NR	EIA, MFI	Novitec EBV-EIA (Genbio), Athena Multi-Lyte (Zeus)	IgM and IgG	Switzerland
de Ory, 2011	Retrospective cohort	125	Sera collected from clinically suspected cases of IM were studied.	NR	NR	Immunofiltration, ELISA, CLIA	Immunoquick filtration (ALL.Diag), Liason (Diasorin S.p.A), Immulite 2000 CLIA (Siemens), ETI-EBV-M reverse assay (Diasorin S.p.A)	IgM and IgG	Spain
De Ory, 2012	Retrospective cohort	50	Serum samples came from a collection previously used in the evaluation of EBV infection were studied. Using IFA method, samples were classified as recent, past, or no EBV infection.	NR	NR	ELFA	VIDAS (Biomérieux)	IgM and IgG	Spain
Koçoğlu, 2014	Retrospective cohort	101	Sera obtained from clinically suspected EBV infected patients.	NR	NR	immunoblot based assay, enzyme-linked fluorescent assay (ELFA), EIA, immunochromatographic assay (ICA)	Euroline profil 2 Immunoblot kit (Euroimmun), ELFA (VIDAS), micro-ELISA (Euroimmun), ICA (VIRapid)	IgM and IgG	Turkey
de Ory, 2014	Retrospective cohort	117	Sera collected from clinically suspected IM cases were included in the study. The samples were classified as EBV recent infection, past infection, and EBV negative	NR	NR	Immunoblot	Euroline profile 2 (Euroimmun)	IgM and IgG	Spain

			based on the results of VCA IgM and IgG as well as EBNA.						
Corrales, 2014	Retrospective cohort	365	Sera collected from samples for routine EBV-specific antibody testing were included in the study. Most samples belonged to children or young adolescents with fever, rash or clinical suspicion of IM.	Median 8 years	39% female	CLIA	Architect, Liaison (DiaSorin), Immulite 2000 (Siemens)	IgM and IgG	Spain
Lapierre, 2016	Retrospective cohort	444	Sera were collected from university lab for routine EBV serology testing. The sera were randomly selected on the basis of their results by standard ELISA or IF assays.	NR	48% female	CLIA	Architect (Abbott)	IgG	Canada
Al Sidairi, 2017	Retrospective cohort	65	Anonymized sera from people with various serological profiles unlinked to clinical information were randomly selected	NR	NR	MFI, chemiluminescent microparticle immunoassays (CLIA)	BioPlex 2200 (Bio-Rad Lab), Architect i2000SR (Abbott)	IgM and IgG	Canada
Kasifoglu, 2018	Prospective cohort	178	Consecutive pediatric patients with lymphadenopathy, pharyngitis, fever, hepatomegaly and/or splenomegaly were sent to hospital microbiology laboratory.	Median 72 months	43.3% female	ELISA, immunoblot based assay, PCR	Euroimmun AG, Euroline profil 2 Immunoblot kit (Euroimmun), Exicycler 96 thermal block (Bioneer)	IgM and IgG	Turkey

Table 5.2 Detailed Results of the Study Quality Based on QUADAS-2 Assessment for Each Individual Study

		Patient Selection					Index test				Reference std				Flow & Timing				Overall
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
PMID	First Author, Year	Consecutive	Not case control study	Exclusion criteria	Risk of bias	Applicability	Index blinded	Threshold pre-specified	Risk of bias	Applicability	Immunofluorescence based assay (IFA) as reference standard method	Reference blinded	Risk of bias	Applicability	All got reference std	All had same ref std	All accounted for	Risk of bias	L = 0, M = 1, and H = 2+ with high likelihood of bias
18080674	Altuglu, 2007	Y	Y	Y	L	L	U	Y	L	L	Y	U	L	L	Y	Y	Y	L	L
24687237	de Ory, 2014	Y	Y	Y	L	L	U	Y	L	L	Y	U	L	L	Y	Y	Y	L	L
16545933	Kleines, 2006	Y	Y	Y	L	L	U	Y	L	L	Y	U	H	L	Y	Y	Y	L	L
10150041	Samlley, 1990	Y	Y	Y	L	L	U	Y	L	L	Y	U	L	L	Y	Y	Y	L	L
1653677	Levin, 1991	Y	Y	Y	L	L	U	Y	L	L	Y	U	L	L	Y	Y	Y	L	L
22854635	De Ory, 2012	Y	Y	Y	L	L	U	Y	L	L	Y	U	L	L	Y	Y	Y	L	L
11922922	Fung, 2002	Y	Y	Y	L	L	U	Y	L	L	Y	U	L	L	Y	Y	Y	L	L
10973466	Gomez, 2000	Y	Y	Y	L	L	U	Y	L	L	Y	U	L	L	Y	Y	Y	L	L
2545744	Ho, 1989	Y	Y	Y	L	L	U	Y	L	L	Y	U	L	L	Y	Y	Y	L	L
8126196	Linderholm, 1994	Y	Y	Y	L	L	U	Y	L	L	Y	U	L	L	Y	Y	Y	L	L
8388892	Wiedbrauk, 1993	Y	Y	Y	L	L	U	Y	L	L	Y	U	L	L	Y	Y	Y	L	L
30015428	Kasifoglu, 2018	Y	Y	Y	L	L	N	Y	H	L	Y	U	L	L	Y	Y	Y	L	M

21191077	de Ory, 2011	N	Y	Y	H	L	U	Y	L	L	Y	U	L	L	Y	Y	Y	L	M
25552141	Koçoğlu, 2014	Y	Y	Y	L	L	N	Y	H	L	Y	U	L	L	Y	Y	Y	L	M
9350722	Svahn, 1997	Y	Y	Y	L	L	U	Y	L	L	Y	U	L	L	L	Y	N	H	M
9077426	Elgh, 1996	Y	N	Y	H	L	U	Y	L	L	Y	U	L	L	Y	Y	Y	L	M
6304142	Fleisher, 1983	Y	Y	Y	L	L	U	Y	L	L	Y	U	L	L	Y	Y	N	H	M
20219083	Devanthery, 2010	Y	Y	Y	L	L	U	Y	L	L	Y	U	L	L	Y	Y	N	H	M
24623623	Corrales, 2014	Y	Y	Y	L	L	U	Y	L	L	Y	U	L	L	Y	Y	N	H	M
9774594	Chan, 1998	Y	Y	Y	L	L	U	Y	L	L	Y	Y	L	L	Y	Y	N	H	M
27258036	Lapierre, 2016	N	Y	Y	H	L	U	Y	L	L	Y	U	L	L	Y	Y	Y	L	M
29034860	Al Sidairi, 2017	N	Y	Y	H	L	N	Y	H	L	N	N	H	U	Y	N	Y	H	H

Table 5.3 Prevalence of Epstein-Barr virus (EBV) infection in the included studies, by population and the reference standard test

Names	Total patients	EBV infected cases	Prevalence
Adults and children with clinically suspected EBV infection			
Fleisher, 1983	500	124	24.8
Ho, 1989	1672	353	21.1
Samlley, 1990	94	84	89.4
Levin, 1991	298	77	25.8
Wiedbrauk, 1993	127	56	44.1
Linderholm, 1994	108	46	42.6
Elgh, 1996	53	100	53.0
Svahn, 1997	203	40	19.7
Chan, 1998	107	23	21.5
Gomez, 2000	350	98	28
Fung, 2002	164	13	7.9
Altuglu, 2007	43	7	16.3
Devanthery, 2010	387	94	24.3
De Ory, 2011	125	82	65.6
De Ory, 2012	43	25	58.1
Koçoğlu, 2014	96	17	17.7
De Ory, 2014	117	70	59.8
Lapierre, 2016	144	105	72.9
Pooled subtotal:		36 [25 to 48]	
Children with clinically suspected EBV infection			
Kasifoglu, 2018	178	25	14
Corrales, 2014	368	118	32.1
Gomez, 2000	110	46	40
Linderholm, 1994	20	8	40
Pooled subtotal:		30 [19, 43]	
Overall total		35 [25, 45]	

Note: The prevalence of primary EBV infection was evaluated based on the age groups. If a study reports different numbers of patients among tests being studied, the data for the greatest number of patients reported were used. All studies used the indirect immunofluorescence test as the reference standard.

Table 5.4 The accuracy of the index test for the detection of primary Epstein-Barr virus (EBV) for serological immunoassays and heterophile antibody tests sorted by the diagnostic odds ratio (DOR)

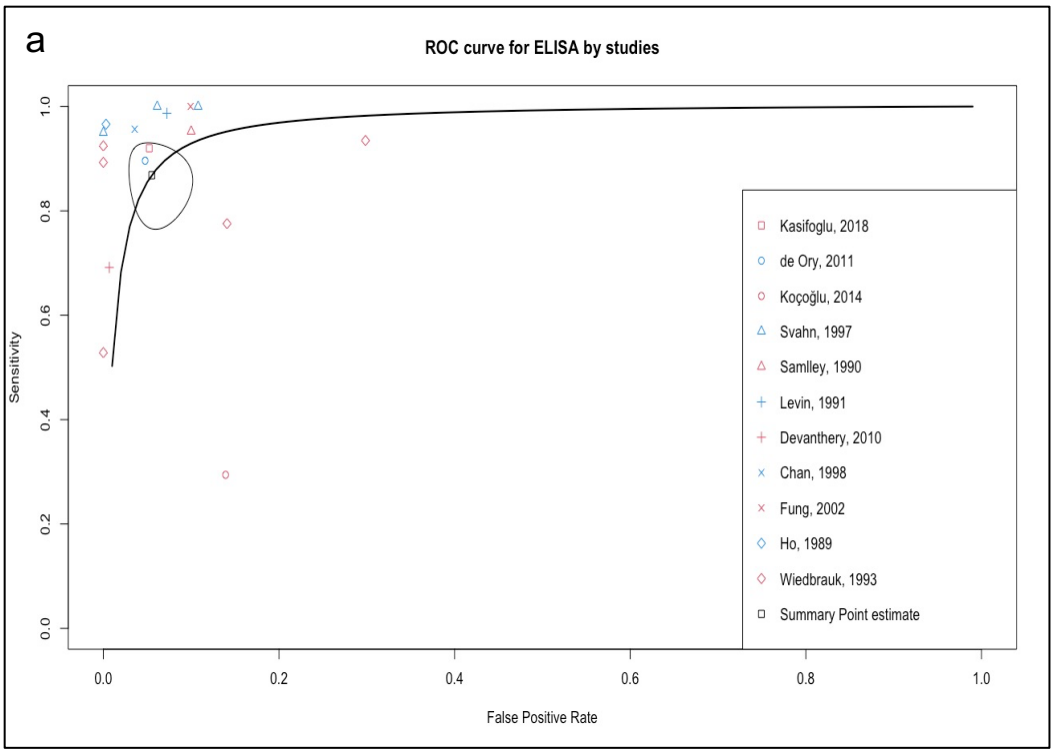
Index tests	Total patients, n (studies, n)	Sensitivity (95% CI)	Specificity (95% CI)	LR+ (95% CI)	LR- (95% CI)	DOR (95% CI)	AUC
Serological immunoassays							
Chemiluminescent microparticle immunoassay (CLIA)	610 (3)	0.95 (0.83, 0.98)	0.94 (0.86, 0.98)	17.8 (6.85, 39.8)	0.06 (0.01, 0.17)	367 (108, 967)	0.97
Children	368 (1)	0.99 (0.94, 1.00)	0.88 (0.83, 0.91)	8.36 (5.92, 11.5)	0.02 (0.00, 0.06)	140 (115, 574)	0.98
Enzyme-linked immunosorbent assay (ELISA)	4311 (17)	0.90 (0.82, 0.94)	0.95 (0.91, 0.97)	19.9 (9.68, 37.2)	0.12 (0.06, 0.19)	192 (62.5, 462)	0.97
All ages	3166 (10)	0.92 (0.83, 0.96)	0.97 (0.90. 0.99)	32.1 (9.79, 79.8)	0.09 (0.04, 0.18)	428 (91.1, 1280)	0.98
Children	178 (1)	0.92 (0.73, 0.98)	0.95 (0.90. 0.97)	18.8 (8.57, 34.9)	0.10 (0.02, 0.29)	308 (42.4, 1130)	0.97
Immunofiltration assay (IMFA)	125 (1)	0.93 (0.85, 0.97)	0.91 (0.78, 0.96)	11.4 (4.25, 25.5)	0.09 (0.04, 0.18)	160 (34.0, 489)	0.97
Multiplex flow immunoassay (MFI)	387 (1)	0.73 (0.64, 0.81)	0.97 (0.95, 0.98)	28.5 (14.2, 53.4)	0.28 (0.20, 0.38)	107 (43.4, 221)	0.96
Immunoblot based assay (IB)	381 (3)	0.86 (0.56, 0.97)	0.91 (0.62, 0.98)	12.8 (2.59, 40.4)	0.18 (0.05, 0.45)	70.2 (25.0, 149)	0.94
Immunochromatographic assay (ICA)	542 (3)	0.41 (0.30, 0.52)	0.98 (0.95, 0.99)	35.2 (7.67, 113)	0.60 (0.48, 0.71)	61.4 (11.3, 210)	0.80
Enzyme-linked fluorescent assay (ELFA)	135 (2)	0.44-0.88	0.89-0.99	7.92-33.3	0.14-0.57	56.6-58.3	0.94
Heterophile antibody test							
Slide agglutination test							
All ages	753 (12)	0.75 (0.66, 0.82)	0.96 (0.92, 0.98)	19.3 (10.2, 33.7)	0.27 (0.19, 0.35)	75.1 (33.8, 142)	0.93
Children	136 (7)	0.35 (0.24, 0.47)	0.95 (0.89, 0.98)	7.39 (3.29, 14.6)	0.69 (0.57, 0.80)	11.0 (4.40, 22.4)	0.77
Solid-phase assay							
All ages	192 (4)	0.73 (0.65, 0.79)	0.95 (0.92, 0.97)	19.5 (10.8, 31.7)	0.29 (0.22, 0.37)	70.5 (33.7, 134)	0.96
Children	33 (4)	0.34 (0.20, 0.51)	0.93 (0.81, 0.98)	6.13 (1.80, 17.3)	0.71 (0.52, 0.88)	9.21 (2.07, 28.4)	0.65

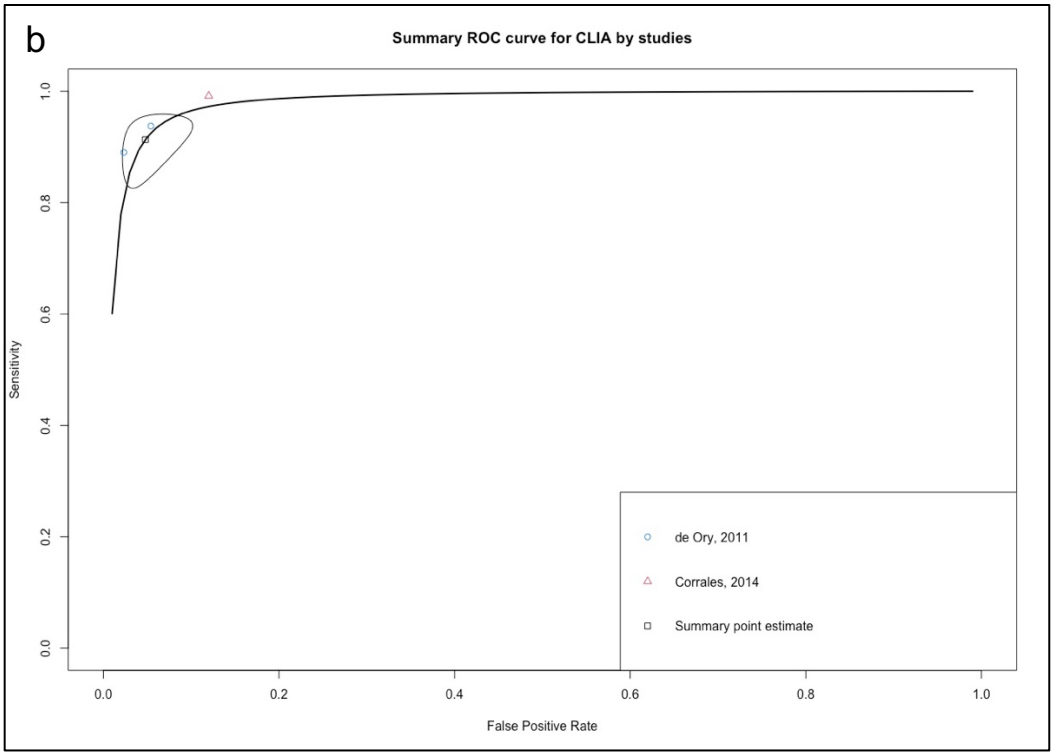
Note: AUC=area under the receiver operating characteristic curve. LR+=positive likelihood ratio. LR-=negative likelihood ratio. DOR=diagnostic odds ratio, calculated by LR+/LR-.

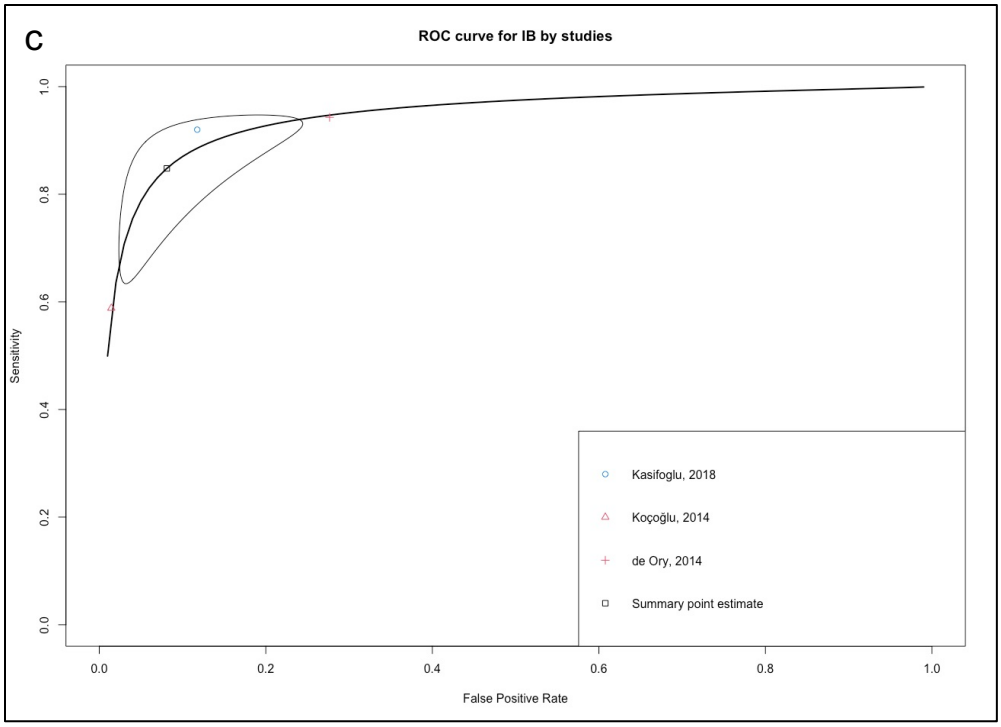
Table 5.5 The diagnostic accuracy for the detection of the anti-viral capsid antigen (VCA) IgM or anti-VCA IgG alone between the EBV specific immunoassays and reference standard

Index method	Total patients, n (studies, n)	Sensitivity (95% CI)	Specificity (95% CI)	LR+ (95% CI)	LR- (95% CI)	DOR (95% CI)	AUC
Assays to detect anti-VCA IgM							
Immunofiltration assay (IMFA)	125 (1)	0.74 (0.63, 0.82)	0.99 (0.86, 1.00)	169 (5.00, 984)	0.28 (0.19, 0.39)	670 (15.1, 4030)	0.98
Chemiluminescent microparticle immunoassay (CLIA)	663 (4)	0.88 (0.79, 0.93)	0.97 (0.93, 0.99)	38.9 (12.3, 94.8)	0.13 (0.07, 0.21)	356 (66.7, 1030)	0.97
Immunoblot based assay (IB)	279 (2)	0.83-0.92	0.90-0.97	8.56-27.2	0.09-0.18	46.3-315	0.86
Enzyme-linked immunosorbent assay (ELISA)	657 (6)	0.75 (0.53, 0.89)	0.92 (0.84, 0.96)	9.60 (4.14, 17.9)	0.29 (0.12, 0.53)	42.9 (8.30, 130)	0.93
Immunochromatographic assay (ICA)	101 (1)	0.75 (0.45, 0.92)	0.89 (0.80, 0.94)	6.81 (3.14, 12.5)	0.30 (0.09, 0.62)	31.2 (5.50, 98.0)	0.90
Enzyme-linked fluorescent assay (ELFA)	151 (2)	0.67-0.79	0.86-0.96	5.76-14.8	0.25-0.35	23.0-42.4	0.85
Assays to detect anti-VCA IgG							
Immunofiltration assay (IMFA)	125 (1)	0.79 (0.71, 0.86)	0.97 (0.69, 0.99)	74.8 (2.56, 435)	0.23 (0.15, 0.33)	372 (8.50, 1890)	0.96
Chemiluminescent microparticle immunoassay (CLIA)	820 (5)	0.96 (0.90, 0.99)	0.92 (0.85, 0.95)	12.2 (6.83, 20.4)	0.05 (0.02, 0.12)	324 (105, 728)	0.97
Immunoblot based assay (IB)	396 (3)	0.99 (0.95, 1.00)	0.70 (0.56, 0.81)	3.4 (2.26, 5.32)	0.03 (0.01, 0.08)	205 (32.4, 704)	0.97
Enzyme-linked fluorescent assay (ELFA)	144 (2)	0.90-0.91	0.87-1.00	6.88-91.0	0.09-0.12	57.0-1011	0.91
Enzyme-linked immunosorbent assay (ELISA)	652 (6)	0.89 (0.82, 0.94)	0.82 (0.74, 0.88)	4.92 (3.40, 7.20)	0.14 (0.08, 0.23)	39.5 (18.1, 77.7)	0.83
Immunochromatographic assay (ICA)	101 (1)	0.85 (0.75, 0.91)	0.83 (0.62, 0.93)	5.51 (2.19, 13.2)	0.20 (0.11, 0.33)	31.9 (7.70, 97.5)	0.90

Note: AUC=area under the receiver operating characteristic curve. LR+=positive likelihood ratio. LR-=negative likelihood ratio. DOR=diagnostic odds ratio, calculated by LR+/LR-.


a
ROC curve for ELISA by studies
Sensitivity
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
Kasifoglu, 2018
de Ory, 2011
Koçoğlu, 2014
Svahn, 1997
Samlley, 1990
Levin, 1991
Devanthery, 2010
Chan, 1998
Fung, 2002
Ho, 1989
Wiedbrauk, 1993
Summary Point estimate


b
Summary ROC curve for CLIA by studies
Sensitivity
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
de Ory, 2011
Corrales, 2014
Summary point estimate



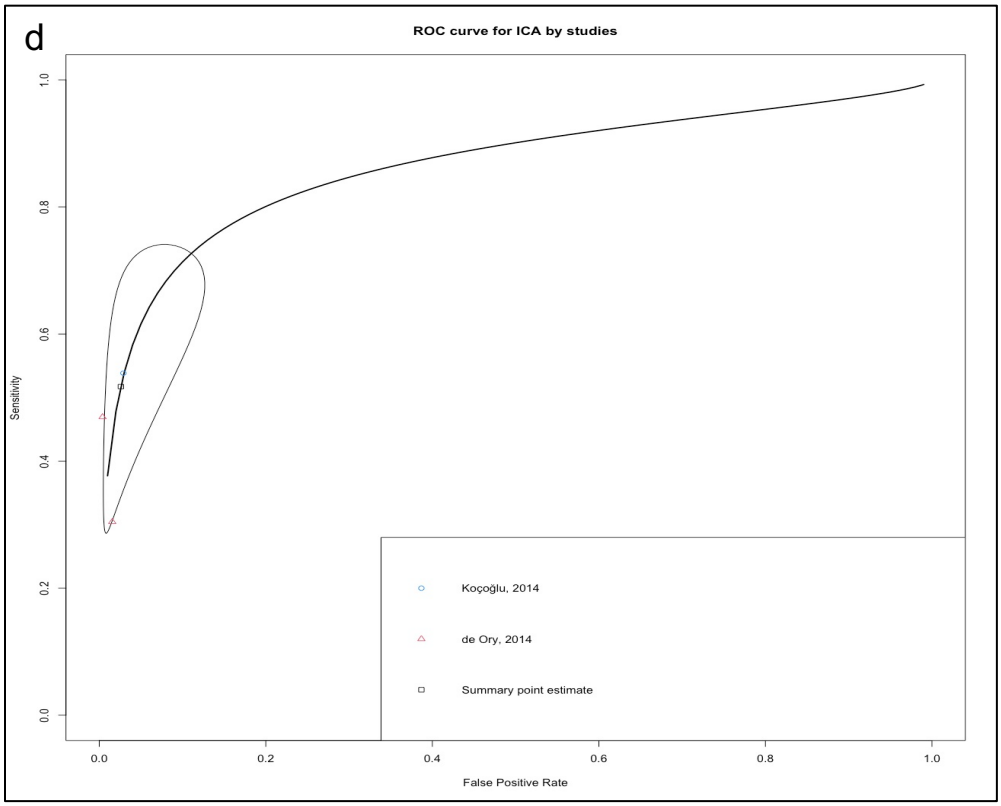


Figure 5.2 Summary receiver operating characteristic curve: a) enzyme-linked immunosorbent assay (ELISA); b) chemiluminescent microparticle immunoassay (CLIA); c) Immunoblot based assay (IB); d) Immunochromatographic assay (ICA)

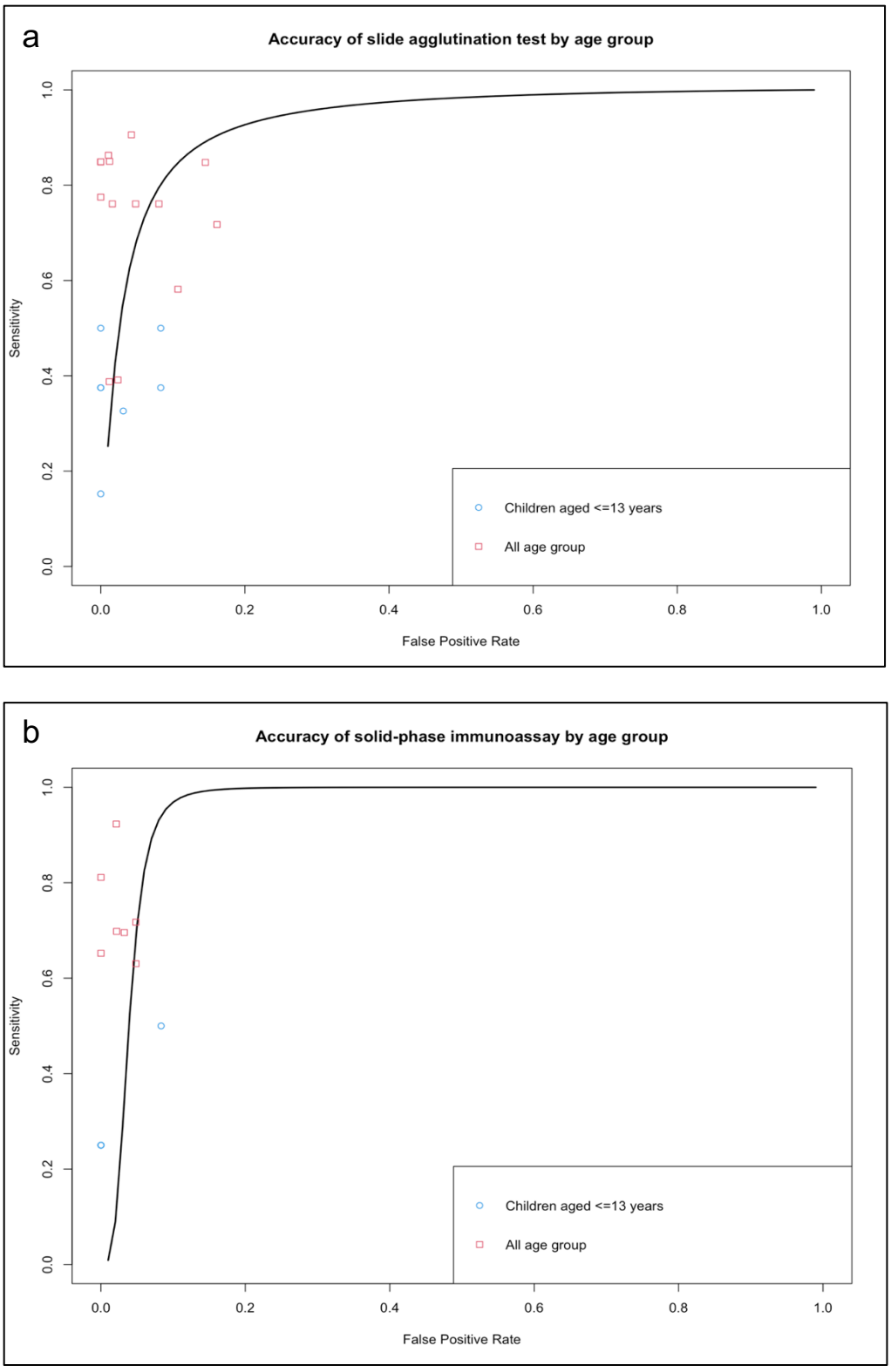


Figure 5.3 Summary receiver operating characteristic curve with accuracy stratified by population for a) slide agglutination test; b) solid-phase immunoassay

CHAPTER 6

A QUANTITATIVE STUDY OF THE TEST THRESHOLD FOR INFECTIOUS MONONUCLEOSIS[3]

[3] Cai, X., & Ebell, M. H. To be submitted to Journal of the American Board of Family Medicine.

Abstract

Background: Ordering a serological test for infectious mononucleosis (IM) in all patients with sore throat is costly and impractical. The test threshold to determine when to order a diagnostic test for IM based on the patient's symptoms has not been previously studied.

Objective: To determine the test threshold for IM in the management of patients with sore throat.

Design and setting: Online surveys were sent to a convenience sample of US primary care clinicians regarding their decision-making regarding whether or not to order a test for IM in a patient with sore throat.

Method: Clinical vignettes were created for seven patients with sore throat that varied by different combinations of symptoms and signs. The probability of IM for each vignette was estimated by the investigator based on a plausible range. Clinicians were then asked to decide whether to test or not test for IM. Mixed-effect logistic regressions were then used to determine the test threshold for IM where half of physicians chose to test and half chose not to test.

Results: A total of 136 clinicians responded to the survey and provided their clinical decisions for a total of 819 clinical vignettes. The overall test threshold for IM as estimated using the logistic regression was 9.5% (95% CI: 8.2% to 10.9%). The test threshold for clinicians practicing greater than 10 years was significantly greater than for those practicing less or equal to 10 years (10.5% vs. 7.3%, p=0.02). No significant differences between specialties and practice sites were found with respect to the test threshold.

Conclusion: This study identified a test threshold for IM of approximately 10% based on the realistic clinical vignettes. This threshold was stable regarding the clinician's specialty and practice sites. The estimation of the test threshold in our study will potentially increase the efficiency in terms of the diagnostic testing for IM among patients with a sore throat.

Background

Infectious mononucleosis (IM) is a common disease caused by Epstein-Barr virus (EBV) among adolescents and children. The most common symptoms and signs associated with IM include sore throat, fever, cervical adenopathy, and fatigue.[24,32,33,120] Other clinical symptoms and signs of IM include rash, headache, nausea, jaundice, axillary or inguinal adenopathy, muscle or joint pain, and splenomegaly. The diagnostic accuracy of individual symptoms and signs of IM have been inconsistent among studies.[9,34] The heterophile antibody tests are rapid and cost-effective test for IM, but the heterophile antibody tests are less sensitive compared to VCA tests, especially in EBV-infected children. The EBV specific tests have a greater ability to rule out IM when the test is negative compared to the heterophile antibody tests and are useful in detecting the early stages of an acute infection.[7,34] However, it would be costly and impractical to order a serological test in all patients with sore throat in routine practice.[3]

The disease threshold model proposes two clinical decision thresholds: the test threshold and the treatment threshold (Figure 6.1).[121] The test threshold is the stage at which the clinician decides to either rule out the disease or to obtain additional data. The treatment threshold is the stage at which clinicians decide to either gather additional data or to initiate treatment. There are three options for patients with suspected illness: if the probability of disease falls below the test threshold, the disease will be ruled out. If the probability falls between the test threshold and treatment threshold, then more information is needed for the diagnosis. If the probability falls above the treatment threshold, then treatment should be initiated.

One of the authors in a previous study[61] developed a novel approach that determined the decision thresholds by presenting physicians with a series of clinical vignettes, each with a different plausibility and disease prevalence that represent a range of probability of disease. Physicians were asked to choose among ruling the disease out, ordering an additional test, or initiating therapy. The thresholds were then estimated using a logistic regression model. The test threshold is the “tipping point” at which a decision was made to either rule out disease

versus gathering more information, while the treatment threshold is the tipping point at which a decision was made to either gather more information or initiate treatment. Since there is no specific medication available to treat IM, and antibiotics do not work against EBV infection, the treatment threshold for IM was not considered in this study.

Yet, the determination of the test threshold for IM has not been studied in clinical practice. It is also uncertain if the thresholds for IM differ by medical practitioner's specialty and their site. By identifying the test threshold for IM, physicians can easily decide if a diagnostic test for IM is needed for patients based on their reported conditions. This study will adopt the previous technique[61] and use detailed and realistic online vignettes that dynamically vary the likelihood of IM among patients presenting with a sore throat to explore the effect of physician decisions regarding testing for IM.

Method

Participants

We recruited a convenience sample of primary care physicians and emailed a participation invitation nationwide from spring to fall 2020. Attached to this email was a link to an online questionnaire about an IM test threshold. The investigators administered the survey using Qualtrics, provided through the University of Georgia (UGA), which provides a secure and automated method for data collection. Respondents could access the online survey either on a mobile phone or on a personal computer. All participants who submitted the Qualtrics survey were included in the study. All surveys were completed anonymously, and each IP address could only be used once to avoid duplicate responses.

Study design

This was a cross-sectional study of clinicians regarding their clinical decision-making in the diagnosis of IM in an outpatient setting with the availability of the office-based examination for IM. After the informed consent was given, each clinician was asked about their medical specialty (family medicine, internal medicine, physician assistant, or nurse practitioner), years in

practice, practice site (primary care, urgent care, or emergency department), and whether they are working at a university health center. Then, each clinician was presented with seven separate clinical scenarios of patients with a sore throat and different combinations of signs and symptoms, corresponding to an estimated likelihood for IM ranging from 1% to 30%. Participants were told to assume that each scenario took place in 2019, and COVID-19 was not a potential cause of the patient's symptoms and signs. The clinical presentations for each scenario given to clinician is shown in Table 6.1. For each scenario, the likelihood of IM was selected from a plausible range of the probabilities predetermined by the investigator based on previous published meta-analysis on the accuracy of the symptoms and signs for the diagnosis of IM.[35] Each clinician was then asked to select from one of the clinical decisions below:

- You feel that IM is unlikely, and you will not order any tests for IM.
- You feel that more information is needed, and you will order a "Monospot" test.

In the threshold model, these two options are consistent with being below the test threshold and being above the test threshold, respectively. The clinicians were told that the heterophile antibody test, the "Monospot" test, is the only confirmatory test available to the physicians for diagnosing IM. The "Monospot" test is considered to be 80% sensitive in the first 7 days of infection and 95% sensitive after 7 days.

Analysis

For the univariate analysis, we summarized the characteristics of each participating clinician descriptively. We also summarized the frequencies and the percentage of the clinical decision for each clinical vignette.

The test threshold was determined by adopting the method described in a previous study.[61] This method is based on a logistic regression analysis of the physician decision regarding the disease probability. The following logistic regression equation was used to determine the test threshold:

$$ln\left[\frac{p}{1-p}\right] = a + bx$$

Equation 1

where *p* is the probability that a clinician decided not to rule out the disease and to order a diagnostic test when the test threshold is being estimated; the value *x* is the probability of IM predetermined by the investigator for each vignette, while *a* and *b* are the model coefficients.

The test threshold of disease is defined as the probability of disease where 50% of the clinicians would decide to rule out disease and the other 50% decide to order diagnostic test or initiate treatment. Thus, we defined the probability of ruling out IM (*p*) as 0.5. At this probability, 50% of the clinicians would rule out IM without ordering additional test and another 50% of clinicians would decide to order the "Monospot" test to confirm IM. The following equation is obtained after inverting Equation 1 with respect to *x* and replacing *p* with $\bar{x}$,

$$X_{test}=-a/b$$

Equation 2

where $\hat{a}$ and $\hat{b}$ were the coefficients estimated from Equation 1. The resulting X_{test} was deemed as a test threshold. Since each clinician needed to evaluate seven scenarios, mixed-effect logistic regression models were used, and we applied a random intercept term to Equation 1 in order to adjust for inter-physician variability. The confidence intervals of the test threshold for IM were determined by using the covariance matrix for the estimated coefficients in the model.[76]

To achieve the subgroup comparison, the test threshold model was also adjusted by years of medical practice (<=10 years vs. >10 years), practice sites (primary care vs. non-primary care), clinician specialties (family physician vs. non-family physician), and whether they work at a student health center.

All statistical analyses were performed using R software version with version 3.0.2.[72] The mixed-effect logistic regression was implemented by glmer() function from lme4 package.

Ethical approval

This study was approved by the Human Subjects Committee of the University of Georgia, and we received informed consent from all participating clinicians.

Results

Characteristics of participants

A total of 136 clinicians responded to the invitation and answered the survey, of whom 122 provided useable data for analysis. The remainder did not provide any information regarding their training or background, so the response rate for this study was 89.7%. The characteristics of the 122 participants are summarized in Table 6.2.

Overall, participating clinicians provided their clinical decisions for a total of 819 clinical vignettes. Most clinicians worked in a primary care setting (82.8%), 83.6% were in family medicine, and approximately 72% of the clinicians had been practicing medicine for over 10 years. Only 11.5% of clinicians worked in a student health center. The percentage of clinicians not ordering a test for IM ranged from 92.6% for scenario 1 (1% estimated probability of IM) to 0.8% for scenarios 6 and 7 (25% to 30% estimated probability of IM).

The distribution of probabilities for clinical decisions among clinicians Figure 6.2 and a comparison of the probabilities for clinical decisions among different subgroups are displayed in box plots Figure 6.3. The box plots show the number of clinicians choosing each clinical decision and the quartiles of probabilities for each decision. The probability of IM was less than 8% for most physicians who ruled out the diagnosis, and among those who ordered a diagnostic test, the probability of disease ranged from 12% to 25%. For physicians to rule out IM, the range of the disease probability is wider for primary care physicians compared to non-primary care physicians. The box plots also show that the median probability of IM is higher among non-family physicians and clinicians with greater than 10 years of practice.

Test thresholds estimation

The overall test threshold for IM as estimated using the mixed-effect logistic regression model was 9.5% (95% CI: 8.2% to 10.9%) (Table 6.3 and Figure 6.4). For the stratified analysis of test thresholds, the test threshold for clinicians practicing greater than 10 years was significantly greater than for those practicing less or equal to 10 years (10.5% vs. 7.3%, p=0.02). No significant differences among other subgroups were found with respect to the test threshold. The test thresholds derived for different subgroups are presented in Figure 6.5.

Discussion

This is the first study to determine the test threshold for IM using a set of realistic clinical vignettes with the probability of IM varying from 1% to 30%. In the context of an outpatient visit for a patient with sore throat, clinicians had an implicit test threshold of just under 10% for ordering a test for IM in a patient presenting with sore throat.

This study found that the estimated test threshold for IM did not differ between primary care physicians and non-primary care physicians, especially among family physicians and non-family physicians. The test threshold for IM was stable across different specialties and practice sites. Thus, this threshold can be used by clinicians regardless of their specialties and/or work site. On the other hand, the test threshold was found to be significantly higher if clinicians had been in practice for greater than 10 years (10.5% vs 7.3%). This may reflect that late-career physicians are more confident in ruling-out disease without testing. However, given the small sample size in this study, this conclusion may be considered as hypothesis generating, and a study with a larger sample size would be required to support this finding.

The test threshold identified from our study can guide the clinician's decision-making process for IM diagnosis. If the pre-test probability of IM for a patient with a sore throat is much below 10% (i.e., 5%), we would assume this patient has a low risk of IM, and the physicians can rule out IM without further testing for IM; if the pre-test probability is much greater than 10% (i.e., 15%), we would assume this patient has a high risk of IM, and a confirmatory test for IM should

be ordered. On the other hand, if the pre-test probability of IM for a patient is judged to be close to 10% (i.e., the estimated probability of IM is between 5% and 15%), we would assume this patient has a high risk of IM, and the physicians need to use their judgement to make their final decision. The authors have developed a simple risk score based on the symptoms and signs for the diagnosis of IM among college students (X. Cai & M.H. Ebell, to be submitted for publication). By using our estimated test threshold of 10%, we identified patients in low- (LR-: 0.5) and high-risk (LR+: 2.46) groups for IM. Given an estimated probability of IM of 8% for adolescents aged 16 to 20 years with a sore throat based on an Australian primary care study,[97] our risk scores would reduce the probability of IM to 4% for patients in the low-risk group, obviating the need for a confirmatory test; and would increase the likelihood of IM to 17% for patients in the high-risk group, for whom IM testing is recommended.

Future studies should investigate the test threshold for IM in a real patient setting. We could directly observe the clinicians' decision-making process in the course of consultations with real patients in front of them. We would ask the physicians to estimate the probability of IM and to make their clinical decisions on IM testing for each visit. Observing the clinicians' behavior in the real patient setting would allow clinicians to include other factors in the decision, such as patient's attitude or expectations, that could not be measured using the simulated vignettes.

Strength and limitation

This is the first study to use realistic clinical vignettes to study clinician decision-making regarding IM testing for patients with sore throat. The study design also allows us to compare the clinicians' clinical decisions across different practice settings and helps us to estimate the threshold with greater precision.

However, several limitations exist in this study. First, the response rate for this study could not be calculated due to our study used a convenience sample. The invitations to participate were extended to the author's personal networks, and we couldn't measure the number of clinicians who received the invitations. Primary care physicians with more

experiences in IM diagnosis might be more interested in this project and they were more likely to participate, which would lead to the selection bias. Second, this study provided only seven vignettes to each clinician, with the likelihood of IM ranging from 1% to 30%. In order to achieve a more accurate estimate of the test threshold, observing clinicians behavior with real patients in front of them, and asking clinicians to estimate the probability of IM for each visit should be considered. This would allow the clinicians to account for the severity of symptoms, as well as the patients' attitude. Also, the probability of IM in each vignette was provided instead of letting the physicians to estimated it, and the stated disease probability in each vignette might potentially have influenced the physicians' clinical decisions. Intuitive clinical decision-making is a complex process and involves different levels of uncertainty, and this process might not be sufficiently captured by the probability of disease.

Third, instead of directly observing real patients, this study summarized the patient's information in each vignette in writing. Although physicians could learn of the patient's clinical presentations directly from the text, this study failed to consider the patient's complaints, such as each patient's expectations for the disease, as well as the patient's concerns when ordering the diagnostic test for IM, which might have biased the physician's clinical decision-making process.[122]

Furthermore, this study might not have fully captured all management options by offering only two options (rule out, test) for each vignette and limiting the diagnostic testing options for IM to the Monospot test. Some physicians may prefer a viral capsid antigen tests as a diagnostic test for IM, since the heterophile antibody tests are relatively specific yet somewhat insensitive (as many as 24% of false negatives), especially in the first week of IM illness. Future studies would provide more options for diagnostic testing and provide the opportunity to compare the decision-making in regard to different serological test options.

Conclusion

This study used realistic clinical vignettes to study clinical decision-making regarding diagnostic testing for patients with sore throat by varying the likelihood of IM among patients in the vignettes. Using this approach, we identified a test threshold of IM of approximately 10%. This threshold was stable in regard to clinician's specialty and practice sites. We also found a slightly higher test threshold for IM among physicians having been in practice for greater than 10 years.

The results from the current threshold study can be used as guidance to develop a clinical decision rule for the management of IM. The CDR can classify patients at low and high risk of IM, corresponding to IM probabilities of less than 10% and greater than or equals to 10%, which are consistent with the test thresholds identified in our study. The estimation of test threshold based on the realistic vignettes will not only increase the acceptability of the score development and classification, but also increase the efficiency in terms of the diagnostic testing in suspected IM cases.

Chapter 6 Tables and Figures

Table 6.1 The clinical presentation for each clinical vignette given to clinician

	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7
Age	30	30	24	18	18	18	18
Days of symptoms	10	6	6	10	10	10	10
Sore throat	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Rash	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cough		Yes	Yes	Yes	Yes	Yes	Yes
Sore muscles			Yes	Yes	Yes	Yes	Yes
Sore joints			Yes	Yes	Yes	Yes	Yes
Sleeping too much				Yes	Yes	Yes	Yes
Nausea				Yes	Yes	Yes	Yes
Fever					Yes	Yes	Yes
Headache					Yes	Yes	Yes
Enlarged tonsils with exudate						Yes	Yes
Posterior cervical adenopathy							Yes
Probability of infectious mono	1%	4%	7%	12%	18%	25%	30%

Table 6.2 Demographic characteristics of the participating clinicians

Characteristic of clinicians (n=122)	**n (%)**
Type of clinical setting	
Family Medicine	102 (83.6%)
Internal Medicine	11 (9%)
Physician Assistant	4 (3.3%)
Nurse Assistant	2 (1.6%)
Other	1 (0.8%)
No response	2 (1.6%)
Time in practice, years	
<=5	24 (19.7%)
6 to 10	10 (8.2%)
11 to 20	41 (33.6%)
>20	47 (38.5%)
Practice site	
Primary care	101 (82.8%)
Urgent care	7 (5.7%)
Emergency medicine	5 (4.1%)
Other	7 (5.7%)
No response	2 (1.6%)
Student health clinic setting	
Yes	14 (11.5%)
No	108 (88.5%)
Clinical decision: clinician ruled out IM without ordering a test	
Scenario 1 (1%)	113 (92.6%)
Scenario 2 (4%)	103 (84.4%)
Scenario 3 (7%)	67 (54.9%)
Scenario 4 (12%)	28 (23%)
Scenario 5 (18%)	18 (14.8%)
Scenario 6 (25%)	1 (0.8%)
Scenario 7 (30%)	1 (0.8%)

Table 6.3 Estimation of overall test threshold and by subgroups

Test Thresholds	Probability of IM (95% CI)	p-value
All participants	9.5 (8.2, 10.9)	
Practice type		0.47
Primary care	9.8 (8.7, 10.9)	
Non-primary care	8.7 (7.1, 10.3)	
Time in practice, years		0.02
0-10	7.3 (5.7, 8.5)	
>10	10.5 (9.2, 11.8)	
Specialty		0.62
Family physician	9.8 (8.5, 11.1)	
Non-family physician	9.0 (6.5, 11.5)	
Student health center		0.59
Yes	9.7 (7.6, 11.3)	
No	10.7 (9.1, 11.8)	

Note: The test threshold is estimated based on probability estimation from mix-effect logistic regression models.

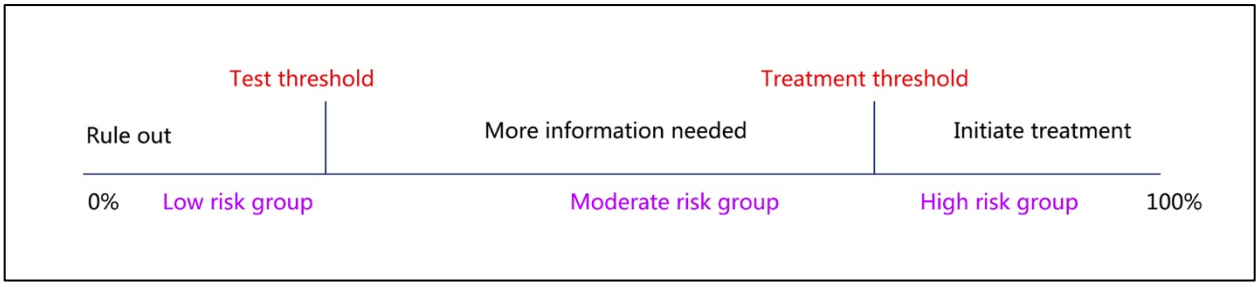


Figure 6.1 Illustration of the threshold model

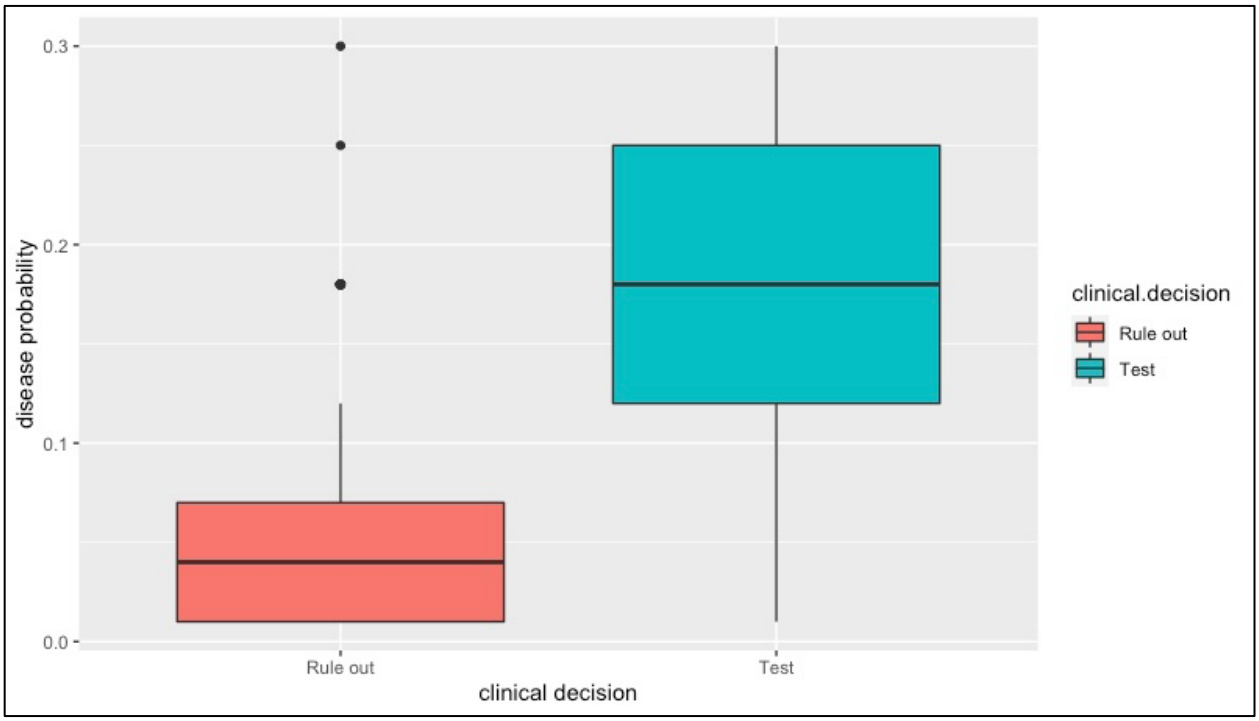
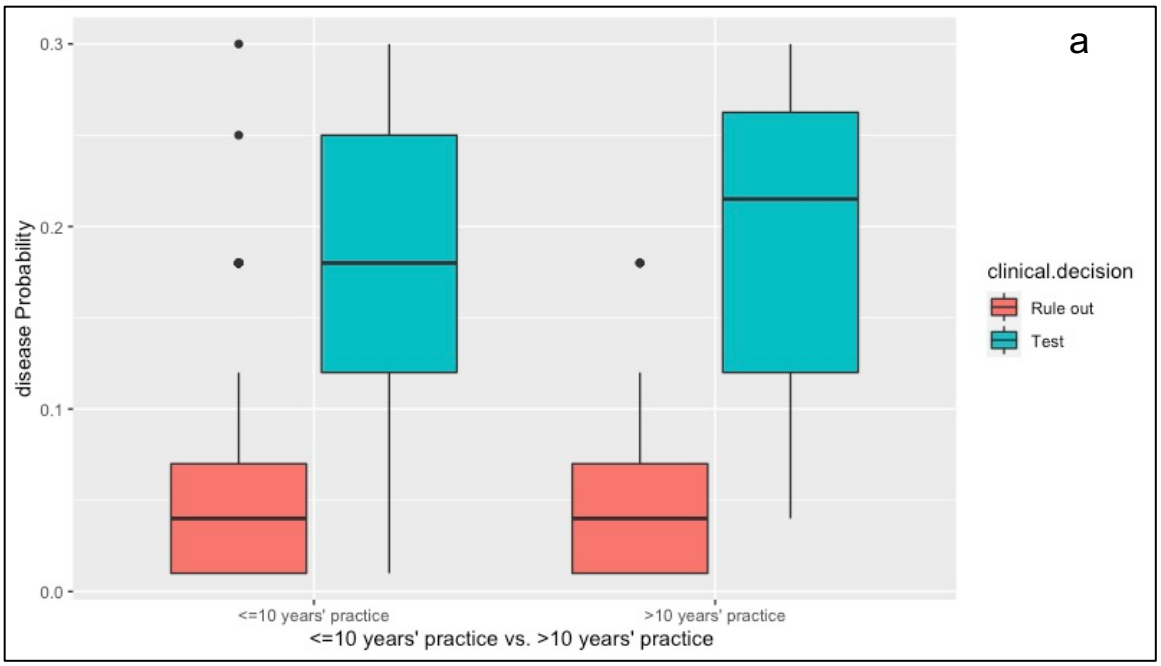
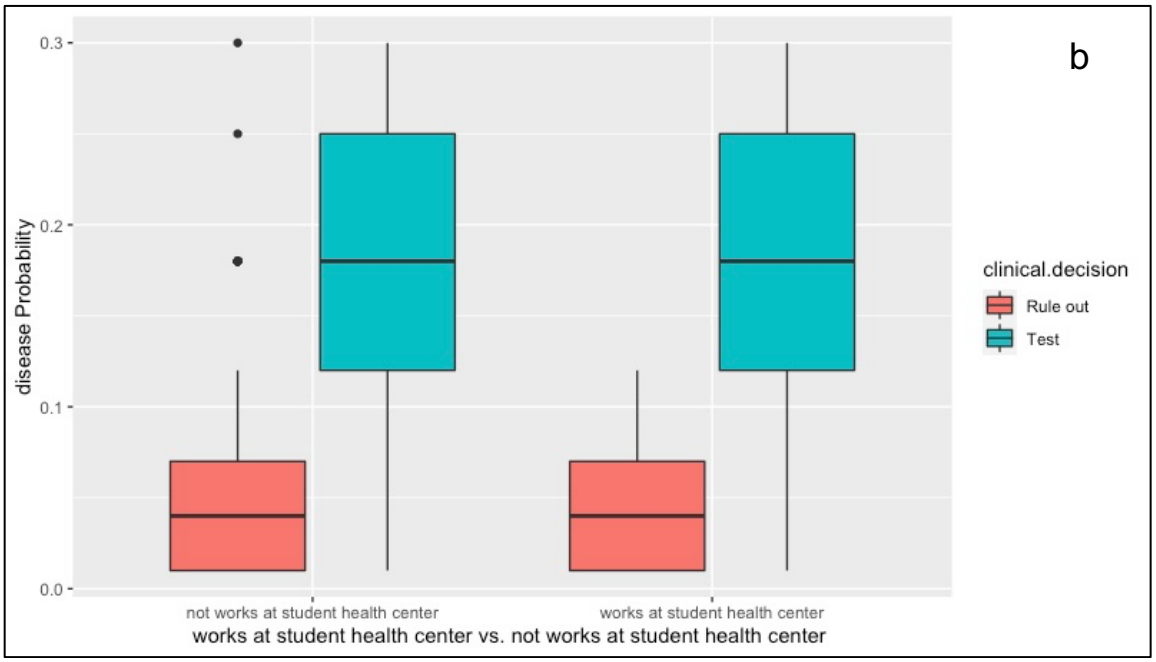


Figure 6.2 The distributions of probabilities for clinical decisions among clinicians


a
0.3
0.2
0.1
0.0
disease Probability
clinical.decision
Rule out
Test
<=10 years' practice
>10 years' practice
<=10 years' practice vs. >10 years' practice


b
0.3
0.2
0.1
0.0
disease Probability
clinical.decision
Rule out
Test
not works at student health center
works at student health center
works at student health center vs. not works at student health center

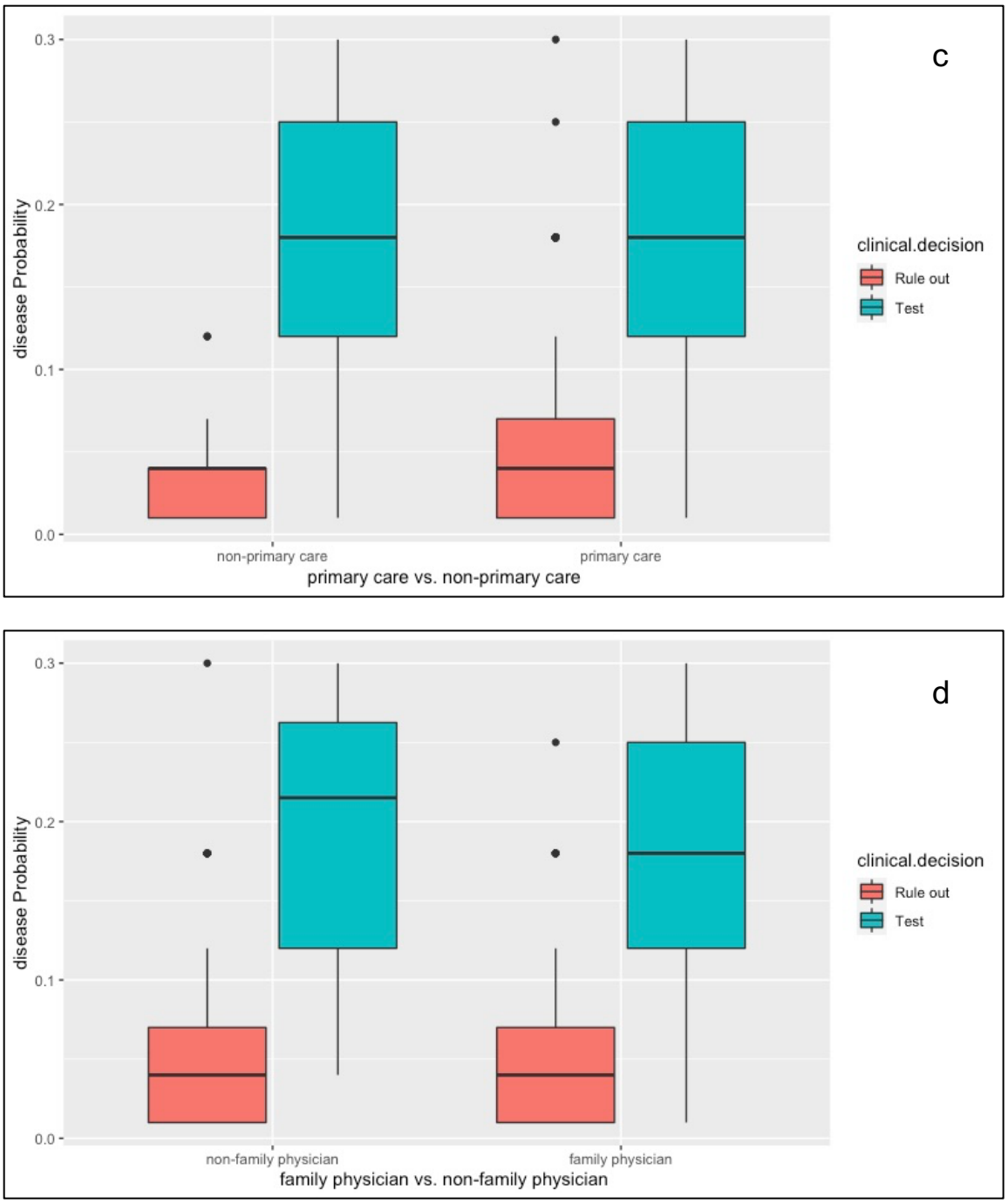


Figure 6.3 Box plots displaying the distribution of probabilities for IM by decision with subgroups: a) <=10 years' practice versus >10 years' practice; b) works at student health center versus not works at student health center; c) primary care physician versus non-primary care physician; d) family physician versus non-family physician

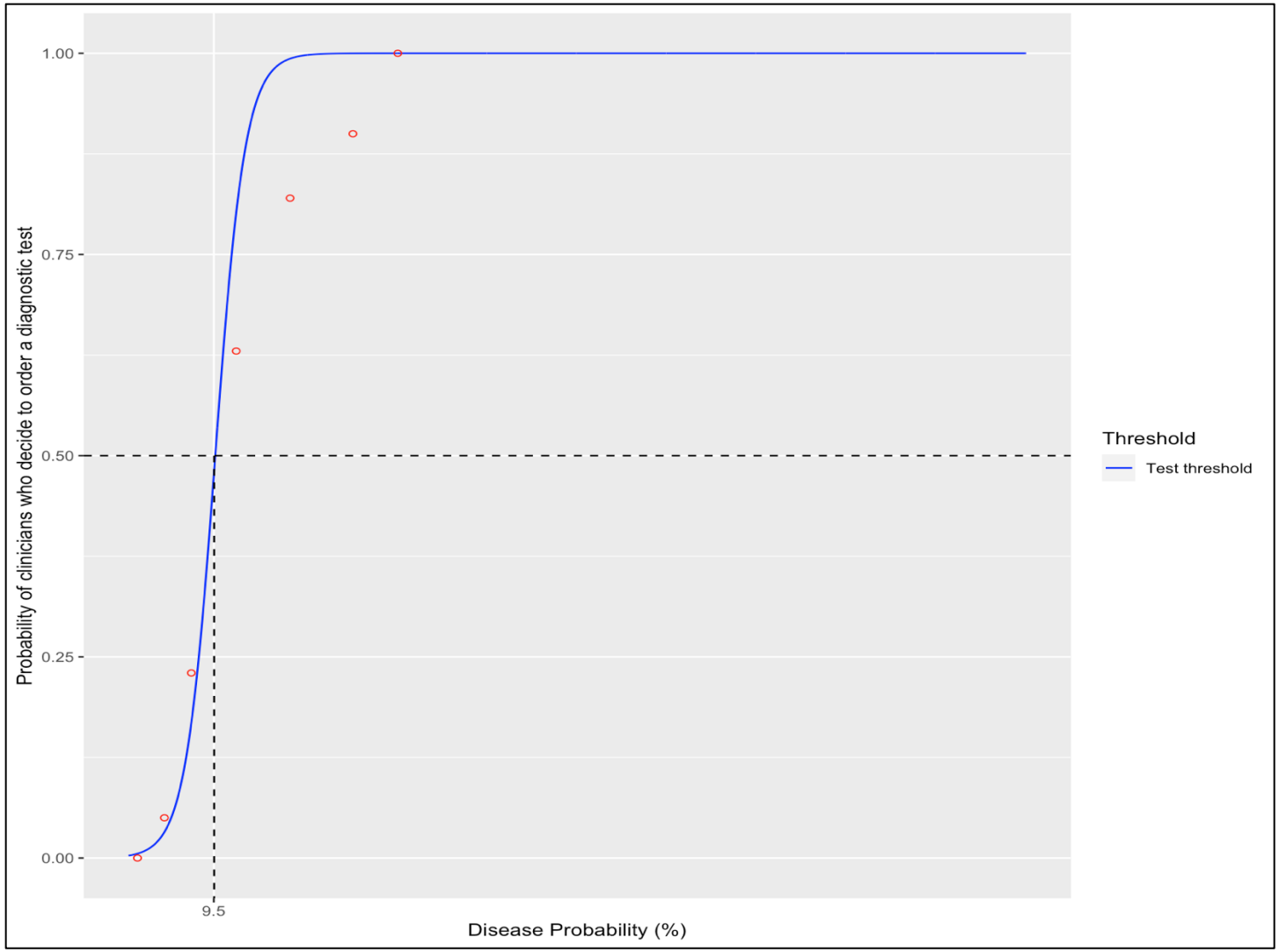
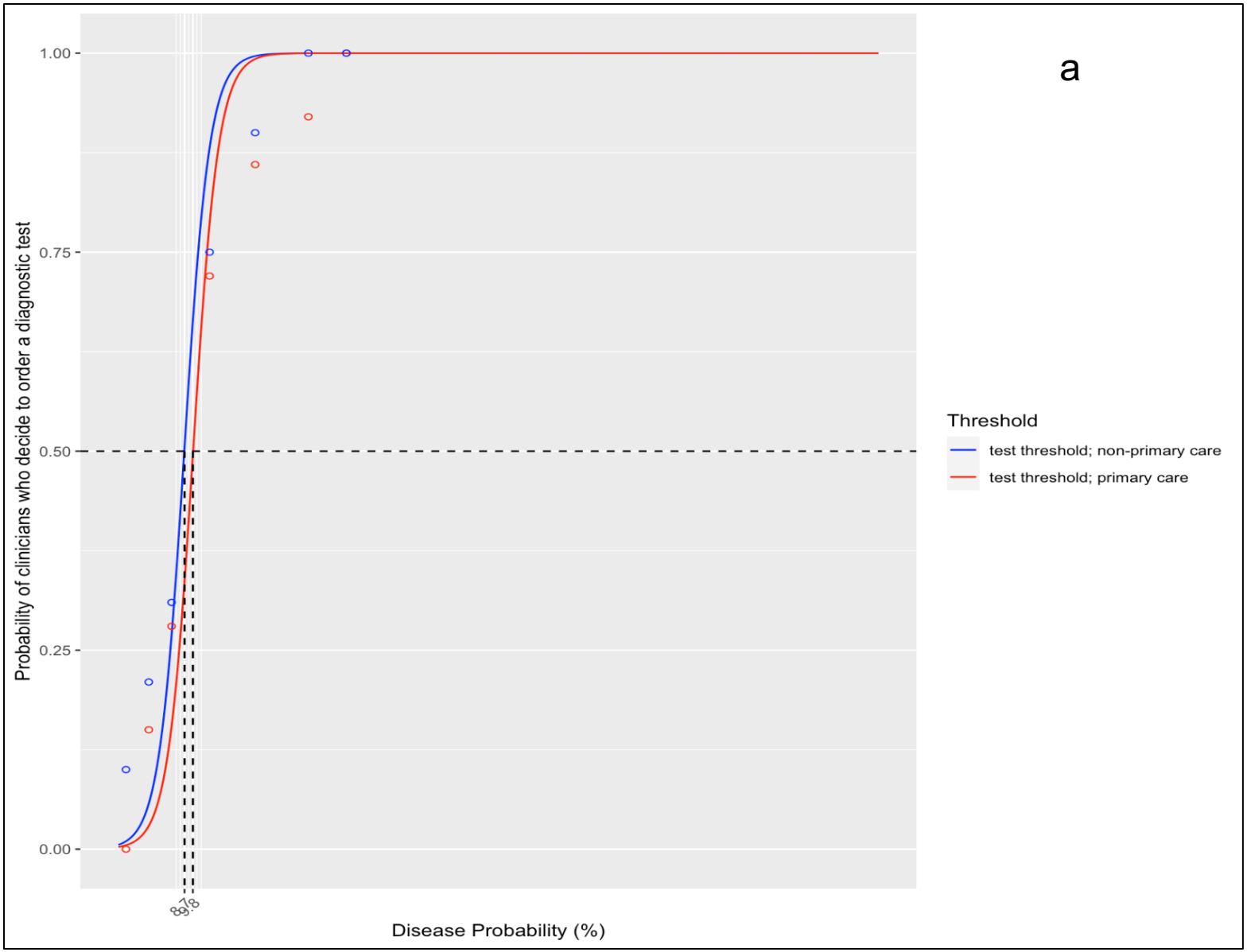
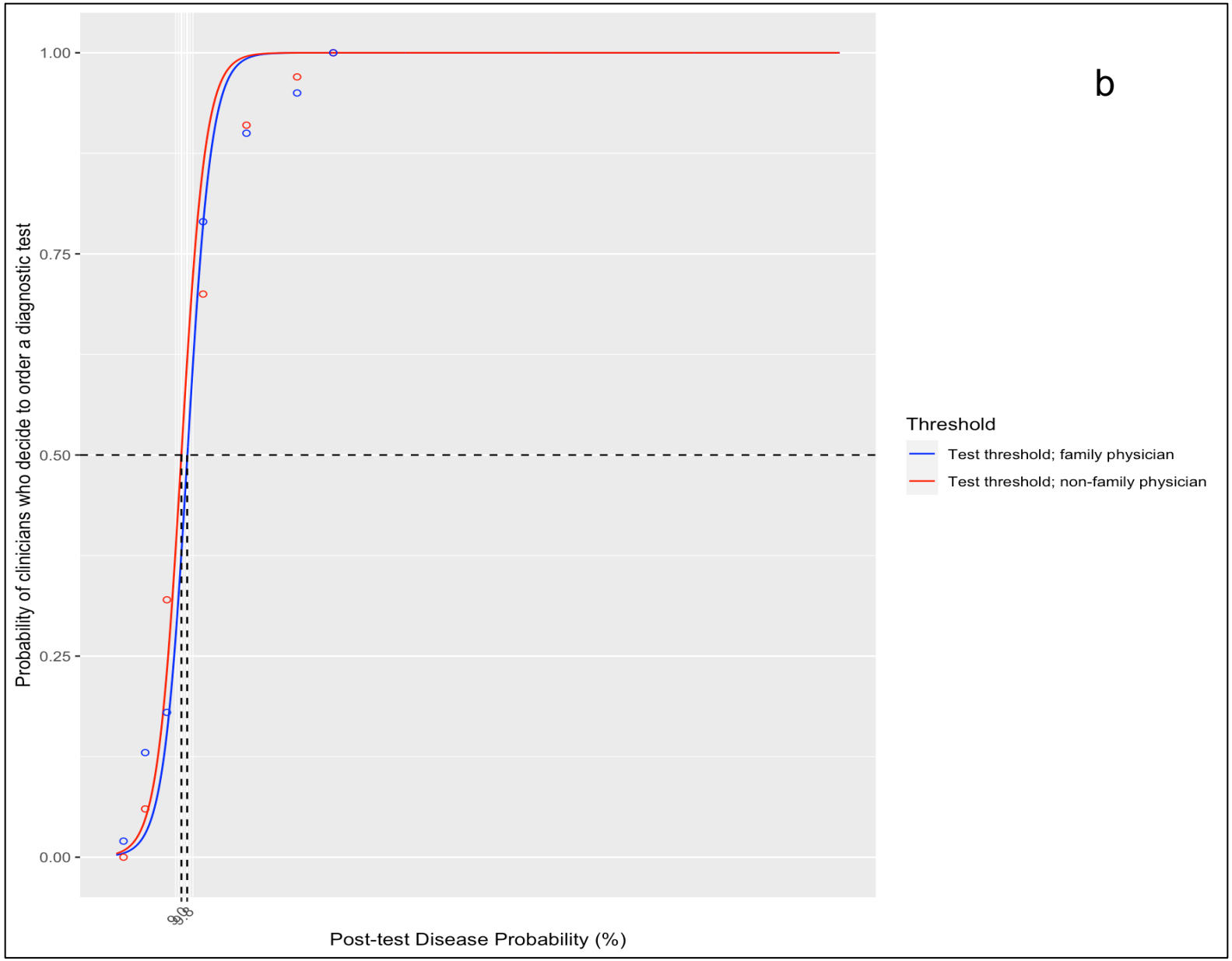
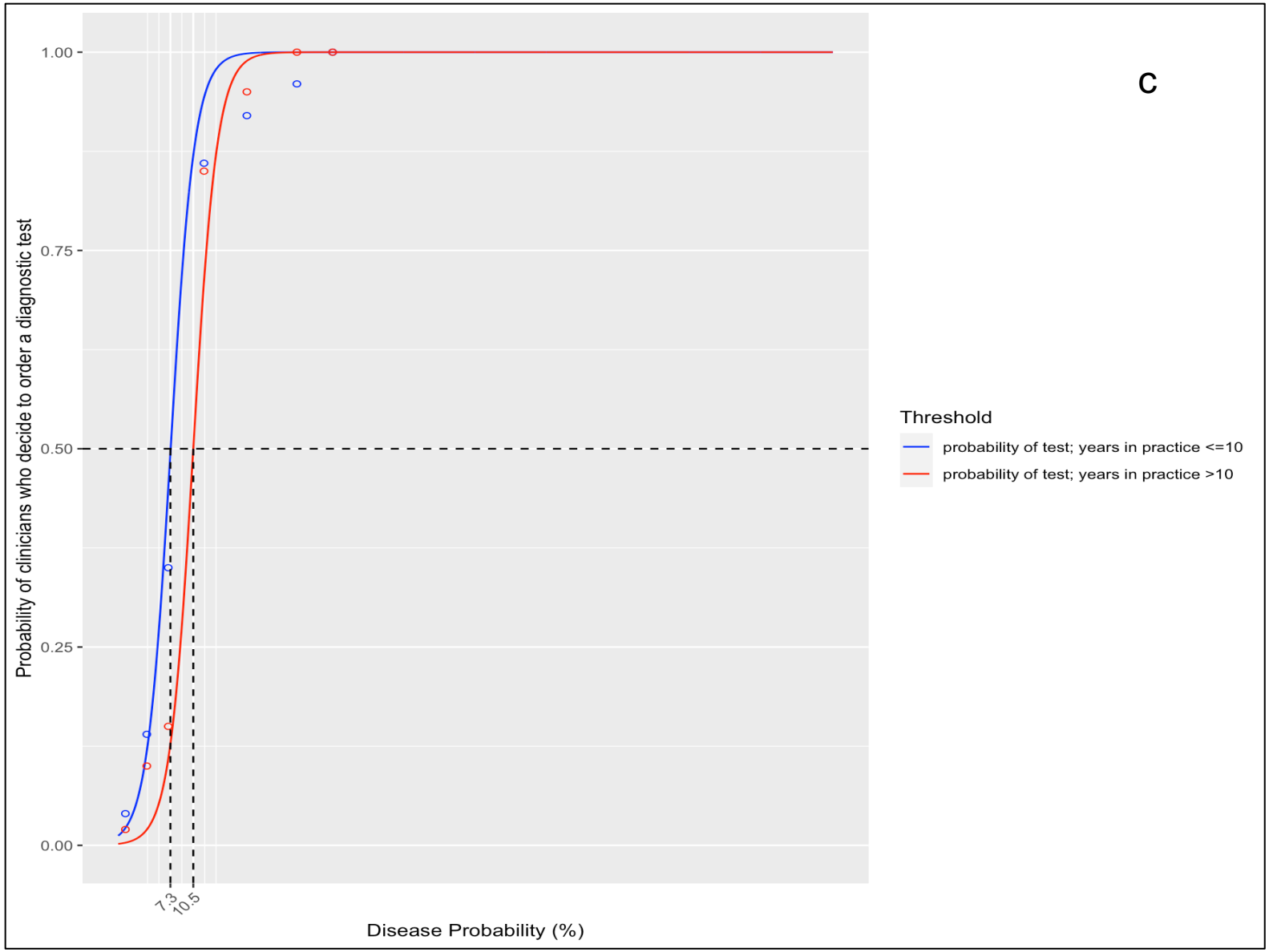
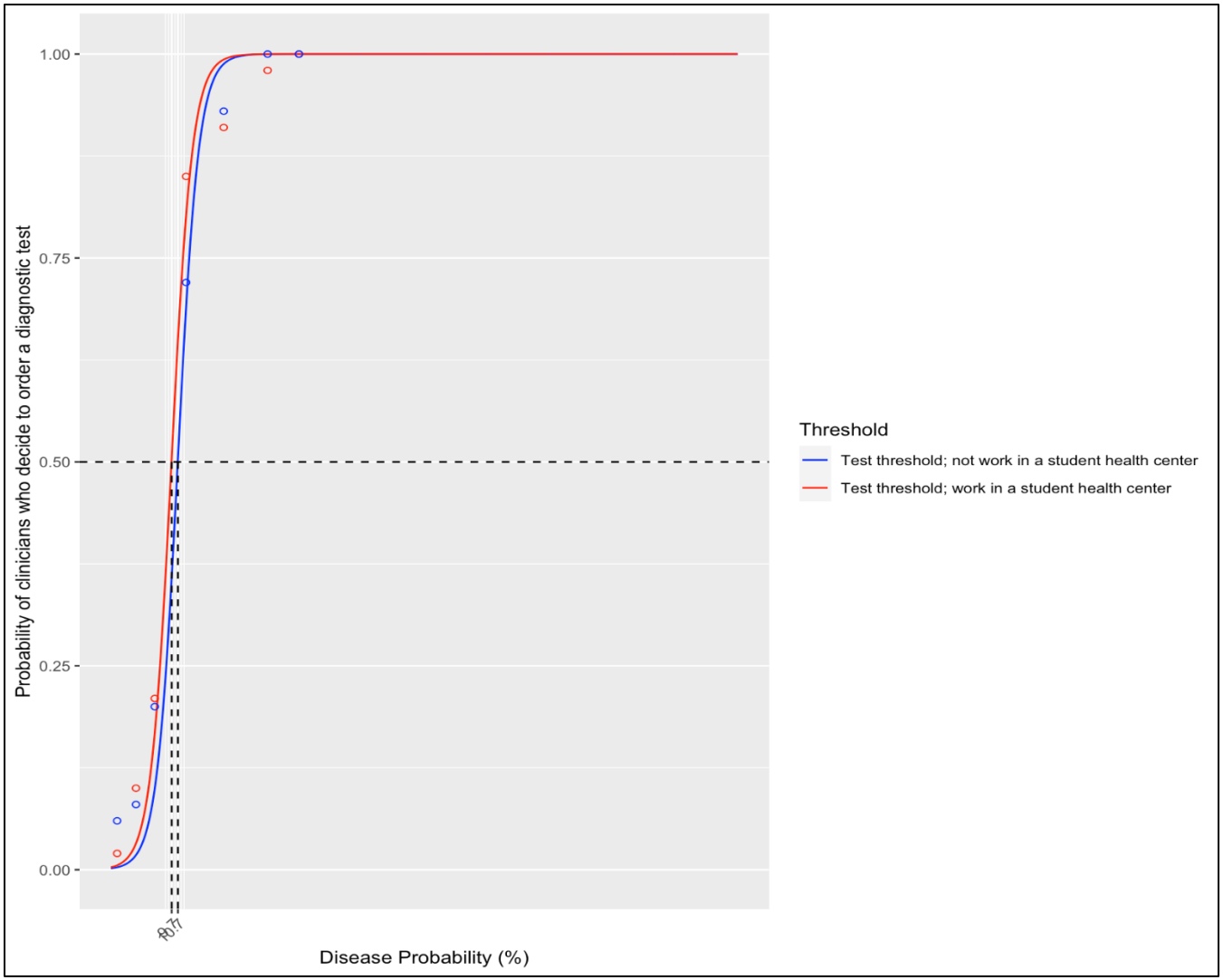


Figure 6.4 Test (blue solid line) threshold based on the logistic regression model, obtained equaling to 0.5 of the probability of not ruling out IM (test threshold) estimated according to model 2. Points (circles) represent the true probability of clinicians that decided to rule in IM and to order a diagnostic test for each scenario.


a
1.00
0.75
0.50
0.25
0.00
Probability of clinicians who decide to order a diagnostic test
8.7
8.8
Disease Probability (%)
Threshold
test threshold; non-primary care
test threshold; primary care


b
1.00
0.75
0.50
0.25
0.00
Probability of clinicians who decide to order a diagnostic test
Post-test Disease Probability (%)
Threshold
Test threshold; family physician
Test threshold; non-family physician


C
1.00
0.75
0.50
0.25
0.00
Probability of clinicians who decide to order a diagnostic test
7.3
10.5
Disease Probability (%)
Threshold
probability of test; years in practice <=10
probability of test; years in practice >10



d

Figure 6.5 Test thresholds of IM by subgroups: a) primary care physician versus non-primary care physician; b) family physician versus non-family physician; c) ≤10 years in practice versus >10 years clinicians; d) working in a student health center versus not working in student health center.

CHAPTER 7

DEVELOPMENT AND INTERNAL VALIDATION OF RISK SCORES TO DIAGNOSE INFECTIOUS MONONUCLEOSIS AMONG COLLEGE STUDENTS[4]

[4] Cai, X., Ebell, M. H., Russo, G., Dobbin, K. K., & Cordero, J. *To be submitted to Clinical infectious diseases.*

Abstract

Background: Individual symptoms or signs of infectious mononucleosis (IM) are of limited value in determining the presence of the disease. Clinicians would benefit from a clinical prediction rule (CPR) in the diagnosis of IM using a combination of clinical symptoms, signs, and hematologic parameters.

Objective: To develop and to internally validate simple risk scores based on clinical symptoms and signs both with and without hematologic parameters in order to diagnose IM among college students.

Method: From 2015 to 2019, data were extracted from the electronic health records of a university health center and were divided into derivation and validation sets. Independent predictors for the diagnosis of IM were identified in the univariate analysis using the derivation set. The multivariate logistic regression models were used to develop two prediction models: one with only symptoms and signs (IM-Nolab) and one adding hematologic parameters to the model (IM-Lab). The point scores were then created based on the regression coefficients, and we grouped patients into low- and high-risk groups. These scores were then validated in the validation group.

Results: We identified 1498 patients who had been tested for IM in the derivation cohort, 243 of whom were confirmed to have the disease. The IM-NoLab risk score included posterior cervical adenopathy, rash, myalgias, headache, and tonsillar exudate and identified patients in the low-risk (8.8% IM) and high-risk (31.2%) groups in the validation dataset (AUC=0.76). The IM-Lab risk score included the ratio of atypical lymphocytes to white blood cell counts (WBC) greater than 10%, the ratio of lymphocytes to WBC greater than 40%, tonsillar erythema, posterior cervical adenopathy, and swollen lymph nodes and identified low- and high-risk groups of 4% and 79.4% respectively in the validation group (AUC=0.94).

Conclusion: Both risk scores were well internally validated and provided a useful tool for clinicians to make a rapid diagnosis of IM. The IM-Nolab score has potential utility in telehealth

visits, but the IM-Lab score provides a more accurate result. When externally validated, they would be useful for improving diagnosis of IM and helping clinicians prioritize diagnostic testing.

Introduction

Infectious mononucleosis (IM) is a common disease among young adults, especially among college students, with the incidence rate ranging from 11 to 48 cases per 1000 persons.[4] Previous studies have shown that the most common symptoms and signs of IM include sore throat, lymph node enlargement, fever, tonsillar enlargement, pharyngeal inflammation, transient palatal petechiae, and rashes.[7] However, the individual symptoms and signs are of limited value for the diagnosis of IM.[9,34] Previous studies[34,41,45] found that using complete blood count (CBC) and leukocyte differential count are also helpful in diagnosing IM in clinically suspected patients.

Laboratory tests, such as the heterophile antibody tests and Epstein-Barr virus (EBV) specific antibody tests are promising, but each of these tests has their own inadequacy. The heterophile antibody test is rapid, cost-effective, and simple compared to other serologic tests.[14] The heterophile antibodies can be detected within one week after the onset of IM. However, these tests are less sensitive than viral capsid antigen (VCA) tests, especially among younger children infected with EBV.[24] The VCA-IgM, VCA-IgG, and EBNA tests were developed for the detection of EBV-specific antibody responses.[3] The VCA-IgM tests are used to detect the early stages of acute infection since the antibodies for VCA-IgM can be detected earlier than those for VCA-IgG.[57] However, the VCA-IgM antibodies may not be produced at all or may only be produced for a short time; therefore, the results of the tests for these antibodies may be inaccurate. Furthermore, the presence of EBNA antibodies cannot usually be detected until six to eight weeks after the onset of clinical symptoms and is used to distinguish between acute and previous infections.

Clinical prediction rules (CPRs) are clinical tools to quantify the diagnosis, prognosis, or response to treatment among patients based on their medical history, physical examination, and basic laboratory results.[58] CPRs may provide evidence for treatment and further diagnostic tests based on the probability of the disease. So far, there have been CPRs developed for strep

throat, pneumonia, influenza, and other infections to aid in diagnosis.[47-50] Although IM is a common condition, there are limited well-designed prospective studies to address the diagnosis of IM, and there has been no published study attempting to develop and validate a clinical prediction rule for IM. Therefore, the primary goal of the current study is to develop and internally validate a CPR for IM that combines symptoms and signs with CBC and leukocyte differential counts to help clinicians make an early diagnosis of IM.[35] To accomplish this goal, we identified the clinical information in a cohort of suspected IM patients from the University Health Center (UHC) at the University of Georgia (UGA), and investigate the risk factors as well as develop a CPR to identify patients with at low- and high-risk of IM.

Method

Study settings

This is a secondary analysis of a previously collected dataset. The lead investigator obtained the de-identified data from the UHC at UGA for all patients in which IM was clinically suspected between September 1st, 2015 and January 1st, 2019, based on the fact that a heterophile antibody test (Monogen test) for IM was ordered. The UHC at UGA provides primary care, specialty health care, education and prevention-focused services to approximately 35,000 students enrolled at the university each year. The UHC has four primary care clinics with approximately 20 primary care clinicians available during regular business hours. The UHC is accredited by the Joint Commission, the nation's most prestigious accrediting board, as well as by the Commission on Colleges of the Southern Association of Colleges and Schools (SACS).[77] Thus, this health care facility is a suitable and ideal place to study IM illness in a college-health population.

Data collection

The standardized dataset of the patients' symptoms, signs, and laboratory parameters was assembled using extant literature. We included any patients cared for between September 1st, 2015 and January 1st, 2019 at the UGA health center who had a diagnostic test for IM. The

study excluded patients with no record available regarding signs, symptoms, CBC information or diagnostic test results. The primary outcome was a positive result of IM as diagnosed by the Monogen test. The Monogen test is the heterophile antibody test in a convenient latex agglutination form.[78] The sensitivity and specificity of Monogen are 94.2% (95% CI: 87.9% to 97.9%) and 91.3% (95% CI: 84.7% to 95.8%), when using a hemagglutination test as the reference standard, and are 99% (95% CI: 89.5% to 99.5%) and 93.3% (95% CI: 85.7% to 96.4%), when using an EBV-specific test to resolve the non-discrepant results.[78] The heterophile antibody test is rapid and cost-effective; therefore, it is widely used for IM diagnosis in the health clinics.[24]

The UHC uses an electronic health record system (EHR) to record and maintain the patients' symptoms, signs, and laboratory test results for each clinical visit. The UGA health center staff, who were not study team members, were responsible for linking the clinical and laboratory data and for removing any identifier/personal information, including name, age in years, birthdate, address, contact information, and student ID number. Each patient was assigned a random ID number only known by the staff in order to maintain confidentiality. The de-identified data was securely transferred from the UHC to the study investigators for analysis. The study investigator then merged the dataset for clinical presentation and laboratory parameters by each patient's ID number, as created by health center staff, and the date of visit.

Independent predictors

The independent predictors in this study included demographic data, symptoms, signs, as well as laboratory findings. The clinical symptoms and signs were recorded as the presence of one or more of the following conditions: fever, diarrhea, vomiting, fatigue, headache, joint pain, myalgia, nausea, rash, sore throat, swollen lymph nodes, cough, anterior cervical lymphadenopathy, posterior cervical lymphadenopathy, pharyngeal erythema, tonsillar erythema, exudative pharyngitis, tonsillar enlargement, and/or tonsillar exudate. The clinical symptoms were reported by patients in the portal and the clinical signs were evaluated by

physicians during the physical examination. The fields in the EHR are optional, where patients and physicians only need to check the symptoms and signs the patient had. The investigators therefore assumed that all the blank inputs in the symptoms or signs for each patient as negative. The laboratory findings included a lymphocyte counts and its percentage, a neutrophil count and its percentage, a monocyte counts and its percentage, a white blood cell count, as well as the percentage of the atypical lymphocytes.

Derivation and validation cohort

This study used a temporal split sample approach to derive the derivation and validation datasets. The data collected from consecutive patients visiting the UGA UHC from January 1st, 2017 to January 31st, 2019 was used as the derivation cohort, which we used to build the model and develop the point scores (N=1498, 64%). We then used the data collected from September 1st, 2015 to December 31st, 2016 as a validation cohort (N=844, 36%) to evaluate the accuracy of the point scores.

Statistical analysis

For a baseline assessment, we presented the continuous variables in medians with interquartile ranges (IQRs) and summarized the dichotomous variables by the frequencies of occurrence with their corresponding proportions. The baseline assessment was stratified by derivation and validation datasets. We then compared the values variables with and without a positive test for IM in the derivation set. The student's t-test was used to compare the continuous variables, and the Pearson χ^2 statistics or Fisher's exact test was used to compare the proportion of the categorical variables between patients with and without a positive test for IM, as appropriate. The study considered that the association between predictors and outcome was statistically significant if the two-tailed p-values were less than 0.05. We selected variables that were significantly associated with our outcome at $p<0.1$ for the inclusion in multivariable analysis.

To develop the point scores using the derivation set, we first used the patient's symptoms and signs as independent predictors (IM-Nolab). Then, we added to the models the hematologic parameters as independent predictors (IM-Lab). If two predictors were statistically correlated, we selected the one that had a stronger association with the outcome in the univariate analysis in order to avoid multicollinearity. The hematological parameters were converted to binary predictors to simplify calculations for the final risk score. The cutoffs were decided based on the inspection of histograms, as well as the biological plausibility from previous systematic review.[35] To build a more parsimonious model, we applied a forward model selection guided by the Akaike Information Criterion (AIC).[14,15] The beta coefficients were determined from the final multivariate models. To improve the model prediction, a uniform shrinkage factor was applied to the regression coefficients to move them toward zero. We then assigned a point score to each predictor in the model by transforming its corresponding beta coefficient. Each beta coefficient was divided by the smallest beta value and then rounded to the nearest integer. The study population was then categorized into low-risk and high-risk groups based on a pre-determined testing threshold of 10% from a previous study. In that study, we surveyed a group of 136 clinicians about their testing preferences regarding IM. Based on the results of that analysis, we estimated a test threshold of 10%, above which serologic testing for IM was indicated.

To assess the performance of the model, we first validated the model internally by using 10-fold cross-validation. The area under the receiving operating characteristic curve (AUC) and the associated 95% confidence intervals (CIs) were calculated from the cross-validation samples. The point scores were then validated using the independent validation cohort. The overall ability of discrimination was evaluated by a receiving operating characteristic curve (ROC) and was measured by AUC. The calibration of the scores was evaluated using the Hosmer-Lemeshow (H-L) test and a calibration plot, which measures how well the predicted outcome matched the observed outcome. We also evaluated the accuracy of the scores based

on the probability of patients being classified into low- and high-risk groups in the validation cohort. All the analysis in our study was performed with Stata version 11.0 (College Station, TX).

Ethical considerations

The University of Georgia's Institution Review Board (IRB) approved this project. It was deemed to be exempt research due to the data in this study was de-identified, and previously collected and extracted retrospectively from electronic health record system.

Results

Characteristics of the study population

The derivation cohort included 1498 patients, of whom 243 of whom were confirmed to have IM (Table 7.1). The validation cohort included 844 patients, and 136 patients tested positive for IM (Table 7.2). The most common symptoms and signs included headache, fatigue, sore throat, anterior cervical lymphadenopathy, and swollen lymph nodes for both patients with and without IM. The distribution of the hematologic parameters are generally similar between the derivation cohort and validation cohort.

The characteristics of the patients were stratified by whether or not they had a positive test for IM, and a bivariate analysis of the association between independent variables and the positive test for IM is shown in Table 7.1. The symptoms and clinical signs significantly associated with IM ($p < 0.05$) included fatigue, headache, myalgias, nausea, rashes, sore throat, swollen lymph nodes, wheezing, coughing, posterior cervical lymphadenopathy, any lymphadenopathy, exudative pharyngitis, tonsillar enlargement, tonsillar exudate and fever (≥100.04°F). Laboratory parameters positively associated with IM included increased lymphocyte count and lymphocyte percentage, increased WBC count, increased atypical lymphocyte percentage, decreased monocyte count and monocyte percentage, and decreased neutrophil count.

Development of the risk scores

Table 7.3 summarizes the two multivariate models used to predict the likelihood of IM. In the model with only symptoms and clinical signs as independent variables, five predictors (posterior cervical adenopathy, headache, tonsillar exudate, myalgia, and rash) were included (Table 2). Posterior cervical adenopathy was a strong predictor of the risk (OR, 4.85; 95% CI,3.6-6.55), followed by rashes (OR, 2.46; 95% CI,1.48-4.06), tonsillar exudate (OR, 2.03; 95% CI,1.45-2.86), myalgias (OR, 1.93; 95% CI,1.42-2.64), and headache (OR, 1.57; 95% CI, 1.15-2.13). Adding hematologic parameters as predictors to the model, the independent variables in the final model included atypical lymphocytes greater than 10% (OR, 28.9; 95% CI, 15.3-54.58), lymphocytes greater than 40% (OR, 4.62; 95% CI, 2.46-8.7), posterior cervical adenopathy (OR, 1.98; 95% CI, 1.14-3.45), tonsillar erythema (OR, 1.75; 95% CI, 1-3.11), and swollen lymph nodes (OR, 1.5; 95% CI, 0.86-2.6).

The risk scores for both models were then created based on the beta-coefficients and by applying a shrinkage coefficient. The total IM-Nolab scores ranged from 0–9, and the IM-Lab scores ranged from 0-16 in the derivation cohort. After grouping patients into low- and high-risk groups based on visual inspection of point scores in the derivation cohort, the prevalence of the IM cases in the corresponding risk group was 8.6% and 32.3% respectively for the IM-Nolab scores (Table 7.4) and was 4.4% and 79.6% respectively for the IM-Lab scores (Table 7.5).

Validation of the risk scores

In conducting a 10-fold cross-validation, the optimally adjusted AUC was 0.73 (95% CI, 0.69-0.75) for the IM-NoLab score (Table 7.4). By adding hematologic parameters to the model, the optimally adjusted AUC increased to 0.94 (95% CI, 0.91-0.95) for the IM-Lab score (Table 7.5). The low-risk groups in the cross-validation samples for both models were below 10%, consistent with an established test threshold (X. Cai & M.H. Ebell, submitted for publication).

In the validation cohort collected between the year 2015-2016, the calibration for both models was good, with non-statistically significant values for the Hosmer-Lemeshow goodness

of fit test (p = 0.1 for the IM-Nolab model and p = 0.52 for the IM-Lab model). The total scores ranged from 0–10 for the points derived from the IM-Nolab model, and the scores ranged from 0-16 for the points derived from the IM-Lab model. For the prediction rule with only symptoms and signs, the probability of IM in the low-risk and high-risk groups was 8.8% and 31.2%, respectively (Table 7.4). After adding hematologic parameters to the rule, the probability of IM was classified as 4% and 79.4% (Table 7.5). The probability of IM in the low- and high-risk groups was consistent with those in the derivation cohort. The AUC in the derivation cohort was 0.76 for the IM-Nolab model, and 0.94 for the IM-Lab model. The ROC plots in the validation cohort are shown in Figure 7.1, and the calibration plots are shown in Figure 7.2.

Discussion

To the best of our knowledge, this study is the first attempt to develop clinical prediction rules in order to diagnose IM. We have developed and internally validated two prediction rules in this study, one of which requires only clinical symptoms and signs (IM-Nolab), and another that requires hematologic parameters commonly available in most clinical settings (IM-Lab). The risk scores performed well in our validation cohort, although external validation using populations from other locations or with different age groups would be desirable. Each developed risk score contained only five predictors, which is simple enough for clinicians to memorize in an outpatient setting. These scores could serve as a useful tool for helping clinicians identify patients who should undergo testing, especially in a low resource or telehealth setting. The IM-Nolab risk score could also be incorporated into a mobile application, where patients could do a self-assessment of their risk for IM and decide if further diagnostic testing is warranted.

Based on a previous systematic review on the independent predictive value of clinical symptoms and signs,[35] posterior cervical adenopathy, headache, and tonsillar exudate help rule in IM. Our analysis also found that myalgia and rash were also independent predictors of IM, which has not been previously reported in the meta-analysis.

For a patient with clinically suspected IM, the IM-Nolab score has potential utility in the context of a telehealth visit, which has become a common venue for assessing respiratory diseases during the COVID-19 pandemic. The observed likelihood ratios for the IM-Nolab scores, which were 0.5 for the low-risk group and 2.5 for the high-risk group, have moderate power to rule in disease. However, since some of the typical symptoms of IM, such as sore throat, were used as the criteria for ordering diagnostic tests in the outpatient setting, this might have limited our ability to use these as predictors due to incorporation bias.

The IM-Lab risk score includes hematologic parameters as predictors and was more accurate than the IM-Nolab risk score. Besides the clinical signs of posterior cervical adenopathy and tonsillar erythema, the IM-Lab risk score also includes the white blood cell differential test results, such as > 40% lymphocytes and > 10% atypical lymphocytes. A previous study[35] has shown that the presence of lymphocytosis (LR+: 5.3, 95% CI: 4.2 to 6.6 for > 40% lymphocytes) and atypical lymphocytosis (LR+: 11, 95% CI: 2.7 to 35 for > 10% atypical lymphocytes) significantly increased the likelihood of IM. The combination of a higher percentage of lymphocytes combined with atypical lymphocytosis is also helpful for diagnosing IM (LR+: 54, 95% CI: 8.4 to 189).[35] In the IM-Lab risk score, the observed likelihood ratio was 0.24 for low-risk group and 20.1 for the high-risk group. Therefore, adding hematologic parameters to our prediction rule significantly increased the accuracy of IM diagnosis, which increased the ability to rule out IM in the low-risk group and to rule in IM in the high-risk group. Both of the risk scores performed well in the validation cohort, which suggests that our scores are feasible and can be further tested in other populations.

Strengths and Limitations

This constitutes the first study to develop and validate a clinical decision tool for IM in any population. With approximately 6 to 8 cases per 1000 persons per year, the incidence rate of IM is highest among adolescents aged 15 to 24 years; IM is also commonly diagnosed in communal living conditions, including military barracks and university accommodations.

Therefore, college students are one of the most appropriate and convenience populations for investigating IM. Our risk scores can be easily memorized for use at the point of care and can successfully classify a clinically meaningful percentage of patients into low- or high-risk of IM.

However, this study has several limitations. First, we were unable to obtain demographic information, such as gender and age. Since most college students are young adults ranging in between 18 and 23 years, age was not assumed to be a confounder in our analysis. However, a previous study[123] showed that males were more frequently diagnosed with IM (the ratio of female to male: 0.48, 95% CI: 0.44 to 0.52 for people aged between 20 and 24 years); further study will need to include gender as a potential risk factor for IM.

Second, response bias may have affected the results due to the clinical symptoms in the portal have been self-reported by patients. Some acute symptoms might have already been resolved before the doctor's visit, so the patients might not have recorded the resolved symptoms. Third, the derivation cohort and the validation cohort were both selected from the UHC at UGA, and the diagnostic test used by the health system, as well as the population, could have potentially biased the evaluation. The next step in our research will focus on an external validation of these PCRs using heterogeneous cohorts with larger sample sizes to determine the generalizability and the validity of the model, for example, to confirm the findings in populations in an outpatient setting consisting of different age groups or in populations in other countries with other health systems. Furthermore, the demonstration of the accuracy of using the office-based lab tests is not sufficient; further studies are needed to examine the impact of the use of such prediction rules based on the rates, the cost, and the outcomes of ordering the lab tests.

Conclusion

The IM-Nolab and IM-Lab risk scores were derived from American university students' health records and provided a useful tool for clinicians to make a rapid diagnosis of IM. The IM-Nolab score used only clinical symptoms and signs to predict the likelihood of IM, and therefore

has potential utility in a telehealth visit. The IM-Lab score was designed to use minimal laboratory tests, and it provided more accurate results than the IM-Nolab. Both of the scores were well validated using the internal validation approach. Upon external validation, these risk scores would be applicable for prioritizing diagnostic testing for IM in the outpatient setting.

Chapter 7 Tables and Figures

Table 7.1 Clinical characteristics of patients in the derivation cohort

	Infectious mononucleosis (N, %)	Not infectious mononucleosis (N, %)	p-value
N	243 (16.2)	1255 (83.8)	
Symptoms and signs (N, %)			
Diarrhea	8 (3.3)	74 (5.9)	0.14
Vomiting	15 (6.2)	65 (5.2)	0.63
Fatigue	190 (78.2)	903 (72.0)	0.05
Headache	147 (60.5)	624 (49.7)	<0.01
Joint pain	3 (1.2)	64 (5.1)	0.02
Myalgia	112 (46.1)	399 (31.8)	<0.01
Nausea	96 (39.5)	269 (21.4)	<0.01
Rash	30 (12.3)	63 (5.0)	<0.01
Sore throat	206 (84.8)	966 (77.0)	0.01
Swollen lymph nodes	154 (63.4)	592 (47.2)	<0.01
Wheezing	23 (9.5)	49 (3.9)	<0.01
Cough	91 (37.4)	647 (51.6)	<0.01
Anterior cervical lymphadenopathy	197 (81.1)	956 (76.2)	0.12
Posterior cervical lymphadenopathy	122 (50.2)	206 (16.4)	<0.01
Any lymphadenopathy	208 (85.6)	984 (78.4)	0.2
Pharyngeal erythema	10 (4.1)	81 (6.5)	0.21
Tonsillar erythema	87 (35.8)	357 (28.4)	0.03
Exudative pharyngitis	73 (30.0)	199 (15.9)	<0.01
Tonsillar enlargement	56 (23.0)	177 (14.1)	0.01
Tonsillar exudate	76 (31.3)	205 (16.3)	<0.01
Fever (Body temperature ≥ 100.04 °F)	9 (3.7)	127 (10.1)	<0.01
Hematologic parameters (Median, IQR)			
Median lymphocyte count, (× 10^9/L)	5.0 (3.2, 7.1)	1.7 (1.3, 2.2)	<0.01
Median lymphocyte percent, (%)	49.9 (37, 63.8)	19.8 (13, 28)	<0.01
Median monocyte Count, (× 10^9/L)	0.7 (0.5, 0.9)	0.7 (0.5, 1)	0.21
Median monocyte percent, (%)	7 (5, 9)	8 (6, 11)	0.11
Median neutrophils, (× 10^9/L)	3.7 (2.6, 6)	6.1 (4, 9.1)	<0.01
Median neutrophils percent, (%)	40.8 (27.9, 54.7)	69.7 (60.7, 77.7)	<0.01

Median atypical lymphocytes percent, (%)	18 (11.5, 28)	2 (0, 4)	<0.01
Median white blood cell count, (× 10^9/L)	9.9 (7.6, 12.5)	9 (6.8, 12.3)	0.03
Hematologic parameters by cutoffs (N, %)			
Lymphocyte count, (× 10^9/L)			<0.01
<4.0	81 (34.2)	1210 (98.5)	
≥4.0	156 (65.8)	18 (1.5)	
Lymphocyte percent, (%)			
>30	200 (83.7)	270 (21.9)	<0.01
>40	169 (70.7)	82 (6.7)	<0.01
>50	119 (49.8)	21 (1.7)	<0.01
Atypical lymphocyte percent, (%)			<0.01
<10	39 (17.8)	548 (94.8)	
≥10	180 (82.2)	30 (5.2)	
Neutrophils count, (× 10^9/L)			<0.01
<8	181 (76.4)	607 (49.4)	
≥8	56 (23.6)	621 (50.6)	
Neutrophils percent, (%)			<0.01
<60	200 (83.7)	289 (23.4)	
≥60	39 (16.3)	947 (76.6)	
Monocyte count, (× 10^9/L)			0.33
<1	184 (77.6)	917 (74.7)	
≥1	53 (22.4)	311 (25.3)	
Monocyte percent, (%)			<0.01
<10	203 (84.9)	889 (71.9)	
≥10	36 (15.1)	347 (28.1)	

Table 7.2 Clinical characteristics of patients in the validation cohort

	Non-IM cases (N, %)	IM cases (N, %)	p-value
N	708 (83.9)	136 (16.1)	
Symptoms and signs (N, %)			
Diarrhea	15 (2.1)	2 (1.5)	0.87
Vomiting	23 (3.2)	3 (2.2)	0.71
Fatigue	389 (54.9)	96 (70.6)	<0.01
Headache	231 (32.6)	73 (53.7)	<0.01
Joint pain	20 (2.8)	0	0.1
Myalgia	175 (24.7)	52 (38.2)	<0.01
Nausea	74 (10.5)	35 (25.7)	<0.01
Rash	74 (10.5)	23 (16.9)	0.04
Sore throat	371 (52.4)	104 (76.5)	<0.01
Swollen lymph nodes	213 (30.1)	70 (51.5)	<0.01
Wheezing	13 (1.8)	1 (0.7)	0.58
Cough	216 (30.5)	34 (25.0)	0.24
Tonsillar enlargement	104 (14.7)	29 (21.3)	0.07
Tonsillar exudate	115 (16.2)	55 (40.4)	<0.01
Anterior cervical lymphadenopathy	562 (79.4)	115 (84.6)	0.2
Posterior cervical lymphadenopathy	113 (16.0)	68 (50.0)	<0.01
Any lymphadenopathy	574 (81.1)	118 (86.8)	0.14
Pharyngeal erythema	22 (3.1)	3 (2.2)	0.77
Tonsillar erythema	259 (36.6)	51 (37.5)	0.92
Exudative pharyngitis	112 (15.8)	53 (39.0)	<0.01
Fever (Body temperature ≥ 100.04 °F)	51 (7.2)	5 (3.7)	0.19
Hematologic parameters (Median, Interquartile range (IQR))			
Median lymphocyte percent, (%)	19.0 [13.0, 27.0]	43.0 [33.7, 54.4]	<0.01
Median lymphocyte count, (× 10^9/L)	1.8 [1.3, 2.2]	5.4 [3.5, 7.7]	<0.01
Median monocyte Count, (× 10^9/L)	0.7 [0.5, 1.0]	0.6 [0.4, 0.8]	<0.01
Median monocyte percent, (%)	8.0 [6.0, 11.0]	7.0 [5.0, 9.0]	<0.01
Median neutrophils, (× 10^9/L)	6.2 [4.3, 9.2]	3.2 [2.3, 4.5]	<0.01
Median white blood cell count, (× 10^9/L)	9.2 [7.3, 12.1]	9.9 [7.1, 13.4]	0.15
Median atypical lymphocytes percent, (%)	2.0 [0.0, 4.0]	20.0 [12.0, 31.0]	<0.01
Median neutrophils percent, (%)	69.0 [61.0, 77.3]	34.0 [24.9, 44.2]	<0.01

Table 7.3 Multivariable analysis result for IM-NoLab and IM-Lab models and risk scores

	Multivariable Odds Ratio (95% CI), P value	β Regression Coefficient	Points Given
IM-NoLab model and risk score			
Posterior cervical adenopathy	4.85 (3.60-6.55), 0.01	1.58	4
Tonsillar exudate	2.03 (1.45, 2.86), 0.02	0.71	2
Rash	2.46 (1.48, 4.06), <0.01	0.90	2
Headache	1.57 (1.15, 2.13), 0.05	0.45	1
Myalgias	1.93 (1.42, 2.64), 0.01	0.66	1
		Range:	**0 - 10**
IM-Lab model and risk score			
Atypical lymphocytes>10%	28.9 (15.3, 54.6), 0.01	3.36	8
Lymphocytes>40%	4.62 (2.46, 8.70), 0.01	1.53	4
Posterior cervical adenopathy	1.98 (1.14, 3.45), 0.02	0.68	2
Tonsillar erythema	1.75 (1, 3.11), 0.06	0.56	1
Swollen lymph nodes	1.50 (0.86, 2.60), 0.15	0.40	1
		Range:	**0 - 16**

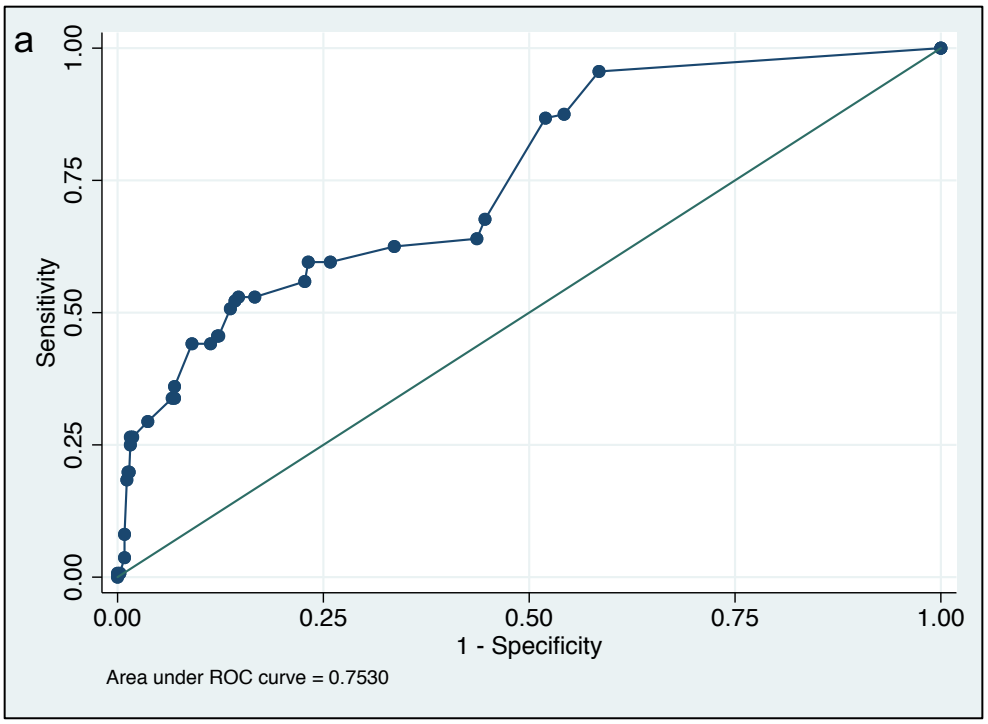
Note: Risk scores were created based on the derivation set data, using beta-coefficients and applying a shrinkage coefficient.

Table 7.4 Risk of IM in the derivation and validation cohorts, according to risk category and summary statistics of IM-Nolab model performances

	Risk Category		
IM-Nolab	**Low (<3 points)**	**High (3+ points)**	**All Participants**
Derivation Cohort, Year 2017-2019			
Number of patients in group, %	1027 (68.6)	471 (31.4)	1498
Number with IM, %	91 (8.6)	152 (32.3)	243 (16.2)
Likelihood Ratio	0.5	2.46	…
Area under the ROC curve	…	…	0.75
Hosmer-Lemeshow test statistics (P value)	…	…	16.48 (0.14)
10–fold Cross-Validation			
Optimism-adjusted AUC (95% CI)	…	…	0.73 (0.67–0.8)
IM % (95% CI)	7.8 (7.4, 8.7)	34.2 (30.1, 36.3)	…
Validation Cohort, Year 2015-2016			
No. (%)	568 (67.3)	276 (32.7)	844
IM, %	50 (8.8)	86 (31.2)	136 (16.1)
Likelihood Ratio	0.5	2.36	…
C–statistic (95% CI)‡	…	…	0.76
H-L test statistics (P value)	…	…	21.6 (0.1)

Table 7.5 Risk of IM in the derivation and validation cohorts, according to risk category and summary statistics of IM-Lab model performances

	Risk Category		
IM-Lab	**Low (<5 points)**	**High (5+ points)**	**All Participants**
Derivation Cohort, Year 2017-2019			
No., %	1263 (84.3)	235 (15.7)	1498
IM, %	56 (4.4)	187 (79.6)	243 (16.2)
Likelihood Ratio	0.24	20.12	…
AUC	…	…	0.93
Hosmer-Lemeshow test statistics (P value)	…	…	9.74 (0.2)
10–fold Cross-Validation			
Optimism-adjusted AUC (95% CI)	…	…	0.94 (0.89, 0.99)
IM % (95% CI)	3.8 (3.4, 4.6)	78.2 (76.3, 79)	…
Validation Cohort, Year 2015-2016			
No. (%)	708 (83.9)	136 (16.1)	844
IM, %	28 (4)	108 (79.4)	136 (16.2)
Likelihood Ratio	0.21	20.8	…
AUC	…	…	0.94
H-L test statistics (P value)	…	…	6.16 (0.52)



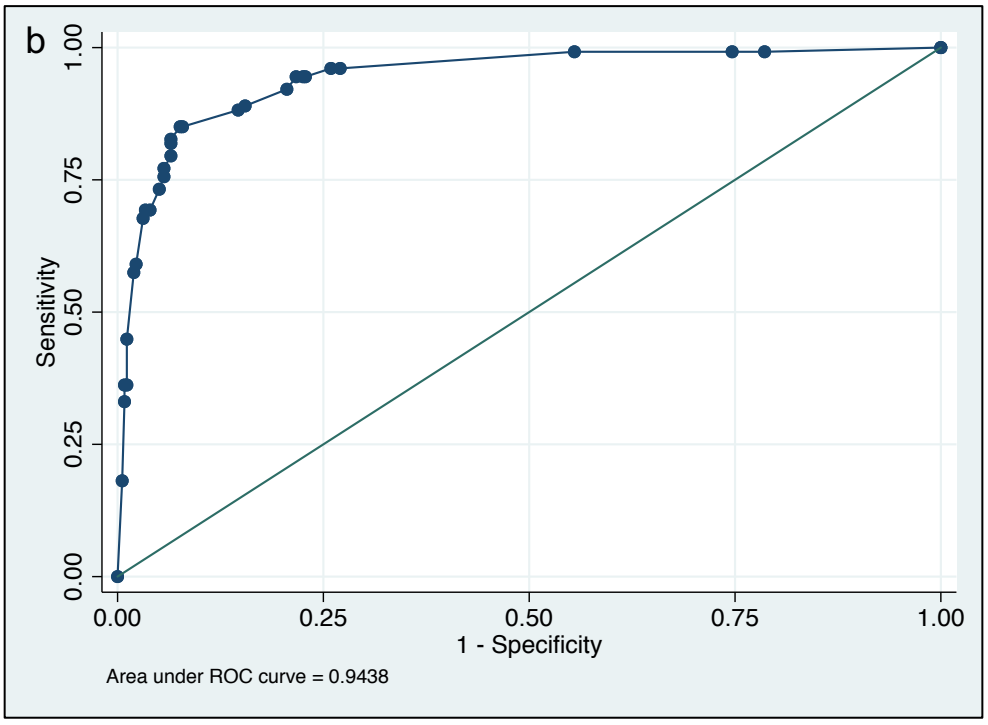


Figure 7.1 ROC plot in the validation cohort: a) IM-NoLab; b) IM-Lab

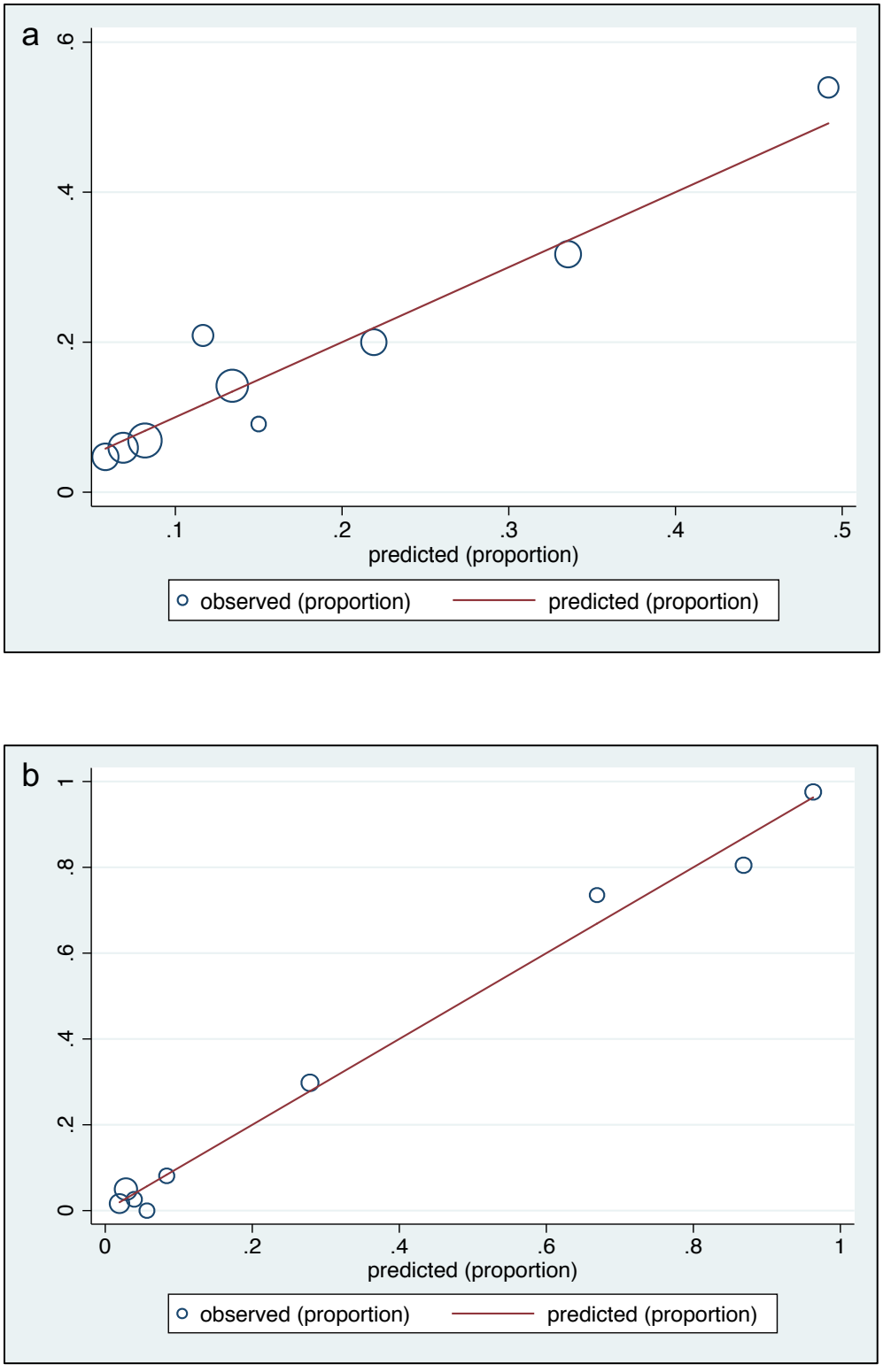


Figure 7.2 Calibration plot in the validation cohort: a) IM-Nolab; b) IM-Lab

CHAPTER 8

THE USE OF INNOVATIVE MODELING STRATEGIES TO PREDICT THE LIKELIHOOD OF INFECTIOUS MONONUCLEOSIS AMONG COLLEGE STUDENTS[5]

[5] Cai, X., Ebell, M. H., Russo, G., Dobbin, K. K., & Cordero, J. *To be submitted to British Journal of General Practice*.

Abstract

Background: Individual symptoms or signs of infectious mononucleosis (IM) are of limited value in determining the presence of the disease.

Objective: To develop and to internally validate decision support tools using three different innovative statistical methods for assessing the likelihood of infectious mononucleosis (IM) among university students using symptoms, clinical signs, and hematologic parameters.

Design and setting: From 2015 to 2019, structured data were extracted from the electronic health records of a university health center and were divided into derivation and validation cohorts.

Method: Independent predictors for the diagnosis of IM were identified by univariate analysis in the derivation cohort. Each statistical method (classification and regression tree (CART), fast and frugal tree (FFT), and artificial neural network (ANN)) was used to develop two models: one with only symptoms and signs (IM-Nolab) and one adding hematologic parameters to the model (IM-Lab). The performance of each model was then internally validated using the validation cohort.

Results: The derivation cohort had 1498 patients (243 with IM) and the validation cohort had 844 patients (126 with IM). The IM-Nolab CART model has only two predictors; the IM-Lab CART model, IM-Nolab FFT model, and IM-Lab FFT model have three predictors; and the ANN models have four predictors. The probability of IM in the low- and high-risk groups in the validation cohort was 7.3% and 32.2% for IM-Nolab CART model (AUC=0.69); 5.0% and 68.2% for IM-Lab model (AUC=0.93); 8.2% and 33.5% for IM-Nolab FFT model (AUC=0.71); 5.9% and 61.8% for IM-Lab FFT model (AUC=0.94); 8.8% and 50.4% for IM-Nolab ANN model (AUC=0.70); and 4.4% and 69.3% for IM-Lab ANN model (AUC=0.97).

Conclusion: The derived IM-Lab and IM-Nolab models based on CART, FFT and ANN methods provided useful tools to help clinicians make rapid diagnosis of IM. Each model was

internally validated, and the discrimination plots showed good discriminations for the IM-Nolab models and excellent discriminations for the IM-Lab models.

Background

Infectious mononucleosis (IM) is a clinical condition mainly caused by the Epstein-Barr virus (EBV). Patients with IM typically present with adenopathy, fatigue, sore throat, and/or fever.[32] Since other respiratory diseases may have a similar clinical presentation,[124] information from individual symptoms and signs has limited value for diagnosing IM. A clinical prediction rule (CPR) that integrates the symptoms, clinical signs, as well as complete blood count (CBC) and leukocyte differential counts could be a useful tool to quantify the likelihood of IM and to classify patients as either low-risk (rule out the disease) or high-risk (test for IM).[121] CPRs have been shown to increase the accuracy of diagnostic assessments and to reduce the cost while maintaining the quality of patient care.[59] Although CPRs have been developed for the diagnosis of strep throat, pneumonia, influenza, and other infections,[47-50] no CPR has been developed or validated to predict the likelihood of IM based on the patients' symptoms and signs. The development of a CPR that combines symptoms and signs with basic lab tests could help clinicians better diagnose IM in the outpatient setting.[35]

In the past year, the authors have performed a series of studies on the diagnosis of IM. In one of these studies, we estimated a test threshold for IM by sending online surveys to a group of 136 US primary care clinicians about their testing preferences regarding IM. Based on the results, we estimated a test threshold of 9.5%, above which serologic testing for IM was indicated. (X. Cai & M. H. Ebell, to be submitted for publication). In this study, we will use this test threshold as the tipping point to classify patients into low- (≤10%) or high-risk (>10%) groups.

Several statistical approaches have been developed as alternatives to logistic regression for the creation of CPRs. The classification and regression tree (CART) approach has been previously applied to the development of diagnostic tools.[80,125,126] CARTs are recursive and partitioning algorithms and have the advantage of making no assumptions regarding the underlying distribution of the predictor variables, and therefore allows inclusion of predictors that

are highly skewed. CARTs also have good face validity for clinicians, and the results are relatively simple for clinicians to interpret.[79,127] Fast and frugal tree (FFT) analysis is another non-parametric method that performs binary classification tasks in sequential order. Compared to CART analysis, FFT is advantageous in terms of frugality (requiring relatively few predictor variables) and simplicity. FFTs can be robust against overfitting and can make accurate prediction even with a small amount of noisy data.[81] The artificial neural network (ANN) is another statistical technique that is suited for complex pattern recognition tasks. ANNs are used to simulate learning and memory processes and have the advantage of modelling complex non-linear relationships.[128] ANNs have been used to diagnose multiple conditions, such as acute myocardial infarction, acute pulmonary embolisms, or survival following cardiopulmonary resuscitation.[129-131] In the present study, we applied the CART, FFT, and ANN methods to create decision support tools for assessing the likelihood of infectious mononucleosis (IM) among university students using their symptoms and clinical signs, as well as hematologic parameters.

Method

Study Design, Participants, and Data Collection

This is a cross-sectional analysis of a previously collected dataset from the university health center (UHC) at the University of Georgia (UGA). We collected clinical information from consecutive patients with clinically suspected IM who had a commercial test of heterophile antibody test (Monogen) at the UHC between September 1st, 2015 and January 31st, 2019. The UHC at UGA has 20 primary care clinicians and provides primary care services to approximately 35,000 university students every year.[77] Respiratory infections are one of the most common reasons for UHC visits.

Information regarding patients in the study cohort was collected using the electronic health records (EHRs) at the time of clinic admission. Recorded information included demographics, symptoms reported by patients, clinical signs assessed by physicians using the

physical examination, as well as the laboratory findings and the diagnostic test result. The data obtained by the investigator were de-identified and any personal information, including age, contact information, physical address, and student ID number were removed by the staff at the UHC in order to maintain confidentiality. For the analysis, the investigator merged the clinical visit data and the lab data by patients' ID numbers, as created by the UHC health staff, and the visit data. The investigators then removed any patients with no record available regarding their clinical presentation or diagnostic test results.

Independent variables and outcome measures

The independent variables are the clinical symptoms as reported by the patients in the portal, clinical signs are recorded by physicians in the physical examination during the visit, as well as laboratory findings of the hematological parameters. The independent variables in our study were selected if they were potentially associated with IM based on a previous systematic review.[35] The hematologic parameters included the total white blood cell count, absolute lymphocyte counts and the percentage of lymphocytes, absolute number of monocytes and the percentage of monocytes, absolute number of neutrophils and the percentage of neutrophils, and the atypical lymphocyte percent. The hematologic parameters in the data were recorded as continuous variables.

The outcome variable was whether each patient had IM as confirmed by the Monogen test. The Monogen test (Biokit, S.A) is a commercial heterophile antibody test in a latex agglutination form.[78] It was performed on all patients by trained nurses at the UHC based on the manufacturer's recommendations. The sensitivity and specificity of the Monogen test are 94.2% (95% CI: 87.9% to 97.9%) and 91.3% (95% CI: 84.7% to 95.8%) respectively compared to a reference laboratory hemagglutination test. When using the EBV-specific tests as the reference standard test performed on non-discrepant sera, the sensitivity and the specificity of the Monogen test increased to 99% (95% CI: 89.5% to 99.5%) and 93.3% (95% CI: 85.7% to

96.4%), respectively.[78] The heterophile antibody test is cost-effective, and patients can obtain the results on the same day as their doctor visit.[24]

Derivation and validation cohort

This study used a split sample approach to create derivation and validation cohorts based on the patient's visit date. We selected the data from consecutive patients visiting the UGA UHC from January 1st, 2017 to January 31st, 2019 as the derivation cohort. The derivation data took up about 64% of the total population and was used to construct and train each statistical model. We then used the data collected from September 1st, 2015 to December 31st, 2016 as the validation cohort (36% of the total population). This validation cohort was used to evaluate the performance of each model developed using the derivation cohort. For each statistical method, we developed 2 separate models in the derivation cohort, one (the IM-NoLab) included only clinical symptoms and signs as candidate predictors, and the second model (the IM-Lab) added the hematologic parameters to the prediction model.

Statistical analysis

Variable selection

To select variables, we summarized the hematologic parameters by medians with interquartile ranges (IQRs) stratified by the diagnostic test result and used the Wilcoxon rank-sum test to compare the results in the derivation data. The clinical symptoms and signs were presented by frequencies of occurrence in patients with and without a positive test for IM in the derivation data, and the Chi-squared test or Fisher's exact test was used to compare the proportions of each predictors. We considered that the differences between the two groups were statistically significant if the two-tailed p-values were less than 0.1, and we only included the predictors that were significantly associated with the diagnosis of IM in the multivariate logistic regression analysis.

CART analysis

The CART approach is a non-parametric statistical method for multivariate analysis. CART is a form of binary recursive partitioning in which each parent node can be divided into two child nodes by identifying the predictors that best differentiate the population into groups with or those without the outcome of interest, in this case laboratory-confirmed IM cases. The child nodes may themselves be divided into additional children in the same manner.[79] CART analysis consists of four basic steps: building the tree, stopping the tree-building process, "pruning" the tree, and identifying the optimal tree.[79] During the tree-building process, a CART approach will check all possible splitting variables and will assign a predicted class to each resulting node that maximizes the sensitivity and specificity of the classification. This process stops when a prespecified minimum number of cases is reached at each terminal node or when all observations within each child node have an identical distribution of predictor variables.

The Gini index method was used to break the parent nodes into child nodes with the default split size set to allow the tree to expand. The logworth statistic was used to rank each candidate predictor in order to identify the optimal split for each node. The logworth statistic is the negative log of adjusted p-values for the chi-square statistic. Splits were then chosen based on the significant values for the candidate predictors.[80] For hematologic parameters, the cutoff values were decided by maximizing the sum of mean squares for the differences between groups. In order to minimize the number of branches without significantly affecting the goodness-of-fit, the tree was then pruned by omitting either predictors that had lower sensitivities or variables that did not classify a substantial number of patients into a risk group.[79] In our analysis, we specified a minimum number of cases at each terminal node of 10 to avoid having unstable estimates with a wide confidence interval. The nodes with a probability of IM less than 10% were grouped together and classified as the low-risk group, and those with a probability greater than or equal to 10% were classified as the high-risk group.

Fast and frugal tree analysis

Next, we applied FFT analysis for multivariate analysis. The FFTs eliminate the need for calculation of a risk score or computation for a full regression model, and are therefore easier to implement than the prediction rules using traditional regression analysis.[81] In the FFT, each level of the tree has one binary classification (cue) and a cue-based question (node). The cues are ranked, and the solution to each cue generates an exit or a further node. The procedure to construct a FFT includes selecting cues, setting the decision threshold for each cue, and determining the sequence of the cues and the exit for each cue. A genetic search (*fan*) algorithm that begins with randomly generated trees and evolved toward an optimal solution is used to address these tasks by selecting the tree in the exit structure, restricting the size of the tree and removing the unnecessary nodes.[81]

In this study, we created and visualized the FFTs and selected the final trees using the *fan* algorithm on the derivation data. To simplify the models for later ease of use, we limited the maximum number of nodes to three. The weighted accuracy (WACC) was defined as the weighted average of sensitivity and specificity dictated by a weight parameter. The weight parameter showed how sensitivity is related to the specificity and was set to 0.5 in order to produce the best model that balance sensitivity with specificity, as the authors believe that false positives are as harmful as false negatives for diagnosing IM. We selected the top three cues with the highest WACC as the final nodes in the FFTs models. Finally, we constructed and visualized the FFT models and calculated the probability of IM as classified at each node. The probability of IM for each node was then calculated and the node was classified as being part of the low- or the high-risk group.

Artificial neural network analysis

ANNs have the ability to learn mathematical relationships from a corresponding output variable among a set of input variables.[82] A typical ANN contains a series of nodes within three layers (input, hidden, output). The input nodes consist of all possible predictors, and the output

node represents the outcome variable.[82] The nodes within the hidden layer are intermediate values that are calculated by the network. The hidden nodes allow the network to model complex, nonlinear relationships between the input variables and the output variable. Similar to the beta coefficient in logistic regression models, each input node is connected to the hidden node by a connection weight, and the connection weight contains the information acquired by trained ANNs.[82] Since there was no existing theory to predetermine the optimal number of hidden layers, we selected the number of hidden layers based on the accuracy of the prediction network.

A training algorithm for an ANN was used to minimize the difference between the value of the predicted output and the actual outcome variable, thereby minimizing the mean square error (MSE) of the network, by estimating the connection weights.[82] The back-propagation training algorithm is one of the most commonly used algorithms to determine optimal connection weights. The learning rate refers to how much each weight changed after the training, with higher learning rates suggesting greater weight changes. The momentum allows the magnitude of each weight change to be proportional.[82] The network applied back-propagation by using the generalized delta learning rule to obtain the optimal connection weights.

We developed ANN models in the derivation cohort using a standard feed-forward, back-propagation ANN. The numbers of hidden-layer neurons were determined to optimize the sensitivity and the specificity of the models. To adjust for back-propagation, we set the learning rate as 0.2 and the momentum as 0.9, respectively.[83] The output neuron had a value of 1 for patients with IM and 0 for patients without IM. The predicted IM cases were calculated through an iterative process. The predicted outputs were then compared to the actual output values, and the connection weights were adjusted according to any errors in the network. This process was completed until the minimum MSE was lower than 0.01. The likelihood of IM for each patients was then generated by ANN from the output neuron.[83]

Model validation

We then evaluated the performance of each model using the validation cohort. We first plotted the receiver operating characteristics (ROC) curves and calculated the area under the curve (AUC) to evaluate the discrimination of each model. We visualized the calibration plot for each model to evaluate how well the predicted outcome matched the observed outcome. We also compared the accuracy of IM probability in the low- and high-risk group in the derivation and validation cohorts.

The univariate analysis was performed using R software, version 3.0.2.[72] The FFTrees package developed with the R software was used for FFT analysis.[81] The CART and ANN models were developed using the Statistical Package for the Social Sciences for Windows version 16.0 (SPSS; Chicago, IL, USA).

Ethical consideration

The University of Georgia Institution Review Board (IRB) approved this project. Since the data in this study was de-identified, previously collected and extracted from electronic health record system, this study was considered to be exempt research.

Result

Characteristics of the study population

Our derivation cohort included 1498 patients, of whom 243 (16.2%) were confirmed to have IM. The validation cohort included 844 patients, of whom 136 (16.1%) were confirmed to have IM.

The distributions of symptoms, clinical signs and hematologic parameters for patients with and without IM based on the Monogen test results in the derivation cohort are shown in Table 7.1. Patients with IM are more likely to report fatigue, headache, myalgia, rashes, sore throat, swollen lymph nodes, nausea, cough, any lymphadenopathy, tonsillar enlargement, posterior cervical lymphadenopathy, exudative pharyngitis, tonsillar exudate and fever (≥100.04°F) than patients without IM. For hematologic parameters, an increased lymphocyte

count and lymphocyte percentage, increased WBC count, increased atypical lymphocyte percentage, and a decreased neutrophil count increase the likelihood of IM. The characteristics of the independent predictors in the validation cohort are shown in Table 7.2.

Structure of the derivate models

Figure 8.1 and Figure 8.2 show the simplified structure of the IM-Nolab and IM-Lab CART models, respectively. The IM-Nolab CART model has 2 predictor variables, while the IM-Lab CART model has 3 predictor variables. For FFT models, we selected the cues with the top three weighted accuracy (WACC) values in each model as the final nodes for the prediction of IM (Table 8.1). The final structures of the IM-Nolab and IM-Lab FFT models developed using the derivation cohort are shown in Figure 8.3 & 8.4. For ANN models, we included the top five important predictors in each model based on the importance scores as the input nodes in order to reduce computational complexity and to achieve transparency of the decision support tool. The final independent predictors included in the IM-Nolab and IM-Lab ANN models are shown in Supplement Table 8.2. The final structures of the IM-Nolab ANN model and IM-Lab ANN model are shown in Figure 8.5.

We then combined the terminal nodes and classified them as being part of the low- or the high-risk group. The probabilities and the stratum specific likelihood ratios of IM for low- and high-risk group in the derivation cohort are shown in Table 8.3.

Model validation

To evaluate the accuracy of the percentage of patients classified as low- and high-risk for the FFT, CART, and ANN models, we reported the probability of patients in the low- and high-risk groups with corresponding likelihood ratios, and the AUC statistics in the validation cohort (Table 8.3). The ROC curves for IM-Nolab models are shown in Figure 8.6 and the ROC curves for IM-Lab models are shown in Figure 8.7. The Calibration plots in the validation groups showed fair agreement between our predicted outcome and the observed test results for both IM-Nolab and IM-Lab models using all statistical methods.

Discussion

This study used three innovative statistical methods to develop decision-support tools for assessing the likelihood of infectious mononucleosis (IM) among university students. The IM-Nolab models are potentially useful in an outpatient setting to help clinicians identify patients who should undergo testing, especially at a time when resources are limited. All IM-Nolab models have similar abilities to discriminate among patients with or without IM (AUC was 0.69 for the IM-Nolab CART model, 0.71 for the IM-Nolab FFT model, and 0.70 for the IM-Nolab ANN model). The IM-Nolab ANN model is more accurate to rule in disease (LR+: 7.52 versus LR+: 2.48 for IM-Nolab CART and LR+: 2.47 for IM-Nolab FFT) and classified a higher proportion of patients in the low-risk group than other methods (85.2% versus 73.5% for IM-Nolab CART and 69.4% for IM-Nolab FFT in the validation cohort). However, the ANN models are referred to as "black-box" decision-making tools because the network have poor transparency and face validity for the decision rules.[131] Thus, the ANN models cannot be directly implemented by clinicians in the practice. The diagnostic accuracy of the IM-Nolab FFT model and IM-Nolab CART model is identical, and they classified a similar proportion of patients in the low-risk group. The IM-Nolab CART model has fewer predictors than the IM-Nolab FFT model (two versus three), making it the simplest one to use by clinicians in the outpatient setting.

The IM-Lab models have higher diagnostic accuracy, better discrimination, and classify a higher percentage of patients in the low-risk group than IM-Nolab models. Among IM-Lab models, the IM-Lab ANN model has the highest AUC and diagnostic odds ratio (DOR) (78.6 versus 22.5 for IM-Lab CART model and 21.7 for IM-Lab FFT model). Since the IM-Lab ANN model has poor transparency, it is currently difficult to implement by physicians in the outpatient setting. The IM-Lab CART model, on the other hand, identified a higher proportion of patients in the low-risk group (86.6% versus 75.1%), and has a higher DOR than the IM-Lab FFT model (22.5 versus 21.7). Therefore, the IM-Lab CART model is the preferred algorithm when blood test results are available.

The CART models and FFT models are intuitive and may be easier to understand by both patients and clinicians compared to point scores calculated by multivariate regression models. In addition, these models can be easily memorized at the point of care and could successfully classify patients into low- and high-risk of IM. The ANN models are more accurate, and could better assist clinicians in ruling out or ruling in IM than CART and FFT models. The main reason for this is that the ANN models are more flexible for measuring nonlinear relationships and would account for any implicit interactions between independent predictors. Therefore the ANN models are more suitable for use when the relationship between the predictors is complex.[132] The ANN models require minimal computing power. The artificial intelligence, both current and anticipated advancements, will play a key role to adopting innovative machine learning techniques in general practice.[133] Shiny is an R package that was broadly used for building an interactive web app.[134] Previous study has successfully developed a graphical interface for ANN that classifying 14 forest adapted species based on audio clips through Shiny package.[135] Extending Shiny app with ANN models in the healthcare settings could potentially assist clinicians to provide health solutions based on the information that are available to them.[136] Since the ANN models are not ready for current use, we would consider that the IM-Nolab and IM-Lab CART models are the most acceptable and preferable models that could be potentially implemented by clinicians in healthcare settings.

Our models performed well in the validation cohort for all three methods, and the probability of IM in the low-risk group ranged from 4.4% for the IM-Lab ANN model (LR: 0.26) to 8.8% for the IM-Nolab ANN model (LR: 12.89). All of the AUCs in the validation cohort were above 0.7, and the range of the AUCs for the IM-Lab models was between 0.86 and 0.95, which indicated our IM-Lab models have very good to excellent discrimination. An external validation of our models on other populations in different clinical settings would be desirable.

Strengths and Limitations

An important strength of this study is that we compared the prediction models using three different statistical methods. The IM-Nolab and IM-lab models performed well based on the results of the internal validation, especially the models incorporating hematologic parameters, which indicates the robustness of our models for the prediction of IM. However, this study has several limitations. First, response bias could be an issue since the clinical symptoms were reported by patients themselves in the portal, and patients could have overlooked their resolved or mild symptoms. Furthermore, the severity of the symptoms and clinical signs were not evaluated, which could have biased the association. Second, patients were recruited from the University Health Center, and most of our population was young adults under 24 years of age; therefore, further external validation is required with populations from other age groups or from other locations to determine the generalizability of our prediction model.

Ultimately, since the same prediction rule could result in different recommendations based on the pre-test probability of disease, our decision rules for IM should be linked to the prevalence data in the study population. The post-test probability for low- and high-risk groups of patients can be calculated according to the LRs estimated in our model and the prevalence of the population. For example, the IM-Lab ANN model would reduce the likelihood of IM to approximately 2% for the low-risk group given a prevalence of 10%, and would increase the likelihood of IM to 57% for the high-risk group.

Conclusion

We derived the IM-Lab and IM-Nolab models for the prediction of IM among university students using CART, FFT, and ANN methods, and have created potentially useful tools to help clinicians make a rapid diagnosis of IM. Each model was internally validated and demonstrated high value in predicting IM. External validation of our models on other population with different clinical settings is needed to further evaluate the performance of our models.

Chapter 8 Tables and Figures

Table 8.1 Individual predictors with the top 5 weighted accuracy value (WACC) in the derivation cohort.

Predictors	Sensitivity	Specificity	WACC
IM-Nolab Model			
Posterior cervical lymphadenopathy	0.5	0.84	0.67
Swollen lymph nodes	0.6	0.59	0.59
Nausea	0.34	0.84	0.59
Tonsillar exudate	0.34	0.84	0.59
Exudative pharyngitis	0.32	0.84	0.58
IM-Lab Model			
Atypical lymphocytes≥10%	0.83	0.94	0.89
Lymphocytes ≥4.0 × 10^9/L	0.7	0.97	0.84
Lymphocytes >40%	0.72	0.93	0.83
Posterior cervical lymphadenopathy	0.5	0.8	0.65
Myalgia	0.46	0.66	0.56

Table 8.2 Importance score of the independent predictors for artificial neural network (ANN) IM-Nolab model

Predictors	Importance	Normalized Importance
IM-Nolab model		
Posterior cervical adenopathy	0.164	100.00%
Fever	0.145	88.30%
Rash	0.107	65.30%
Tonsillar enlargement	0.096	58.60%
IM-Lab model		
Absolute lymphocyte counts	0.3	100.00%
Atypical lymphocyte percent	0.289	95.60%
Absolute neutrophil counts	0.061	20.30%
Lymphocytes percent	0.057	18.90%

Table 8.3 Classification accuracy and area under the curve (AUC) for each of the CART, FFT, and ANN models in the derivation and validation cohorts

	Derivation cohort			**Validation cohort**		
	IM/total (%)	LR	% in group	IM/total (%)	LR	% in group
CART model						
IM-Nolab						
Low risk	70/963 (7.3)	0.40	64.3	43/586 (7.3)	0.45	69.4
High risk	173/535 (32.3)	2.47	35.7	83/258 (32.2)	2.70	30.6
AUC	0.72			0.69		
IM-Lab						
Low risk	73/1210 (6.0)	0.33	86.6	36/712 (5.0)	0.30	84.4
High risk	170/288 (59.0)	7.44	13.4	90/132 (68.2)	12.2	15.6
AUC	0.92			0.93		
FFT model						
IM-Nolab						
Low risk	85/1027 (8.3)	0.44	68.6	51/620 (8.2)	0.51	73.5
High risk	158/471 (33.5)	2.48	31.4	75/224 (33.5)	2.87	26.5
AUC	0.74			0.71		
IM-Lab						
Low risk	52/1125 (4.6)	0.25	75.1	42/708 (5.9)	0.36	83.9
High risk	191/373 (51.2)	5.42	24.9	84/136 (61.8)	9.21	16.1
AUC	0.93			0.94		
ANN model						
IM-Nolab						
Low risk	125/1299 (9.6)	0.55	86.7	63/719 (8.8)	0.55	85.2
High risk	118/199 (59.3)	7.52	13.3	63/125 (50.4)	5.79	14.8

AUC	0.74			0.70		
IM-Lab						
Low risk	33/1196 (2.8)	0.15	73.2	31/707 (4.4)	0.26	83.8
High risk	210/302 (69.5)	11.79	26.8	95/137 (69.3)	12.89	16.2
AUC	0.94			0.97		

Note: *CART=classification and regression tree. FFT=fast and frugal tree. ANN=artificial neural network. LR=likelihood ratio*

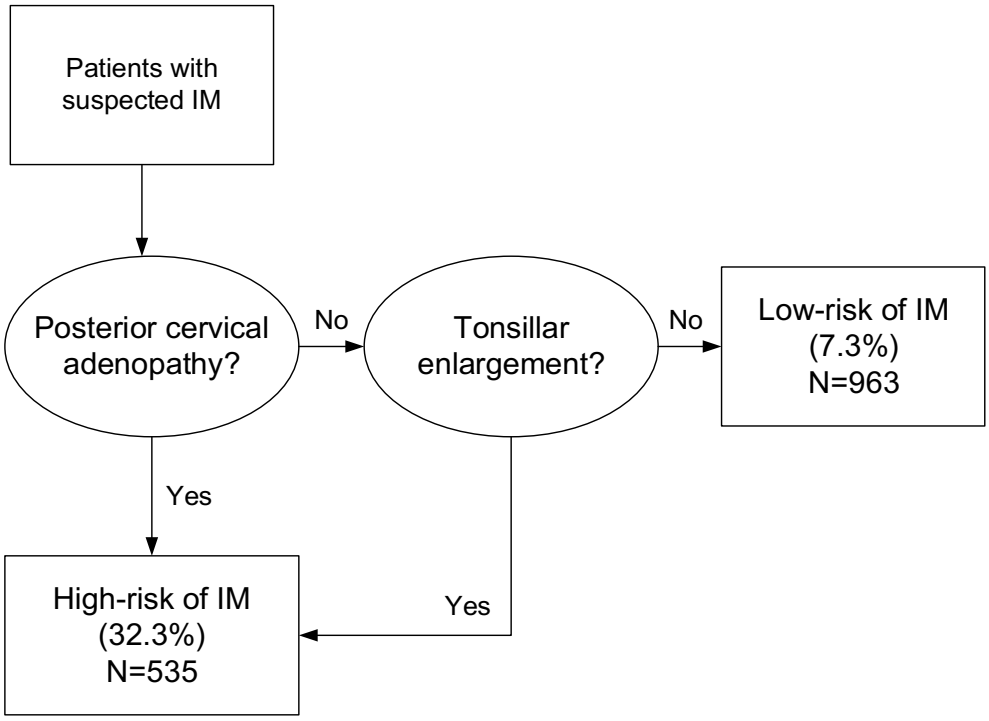


Figure 8.1 Classification and regression tree for IM-Nolab model in derivation cohort

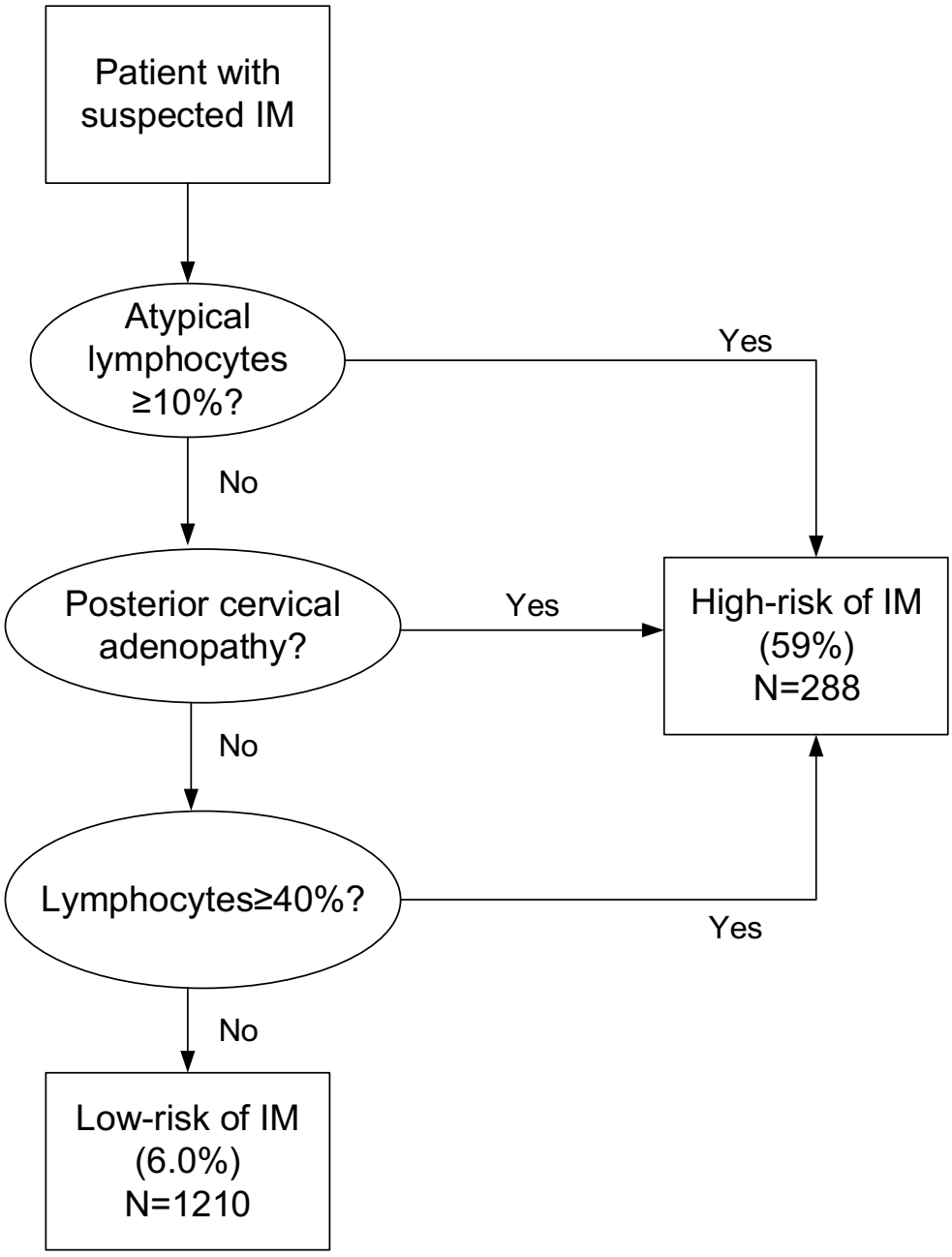


Figure 8.2 Classification and regression tree for IM-Lab model in derivation cohort

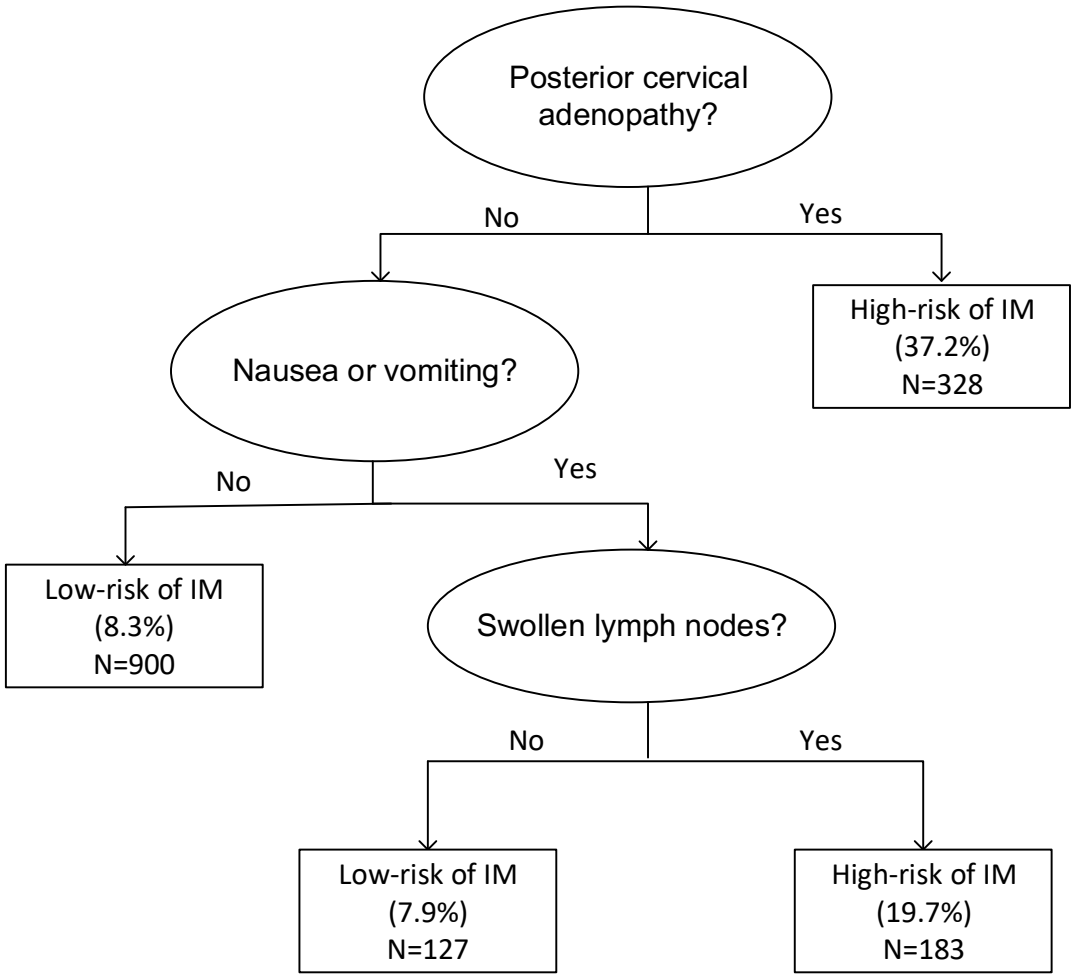


Figure 8.3 Fast and frugal tree for IM-Nolab model in derivation cohort

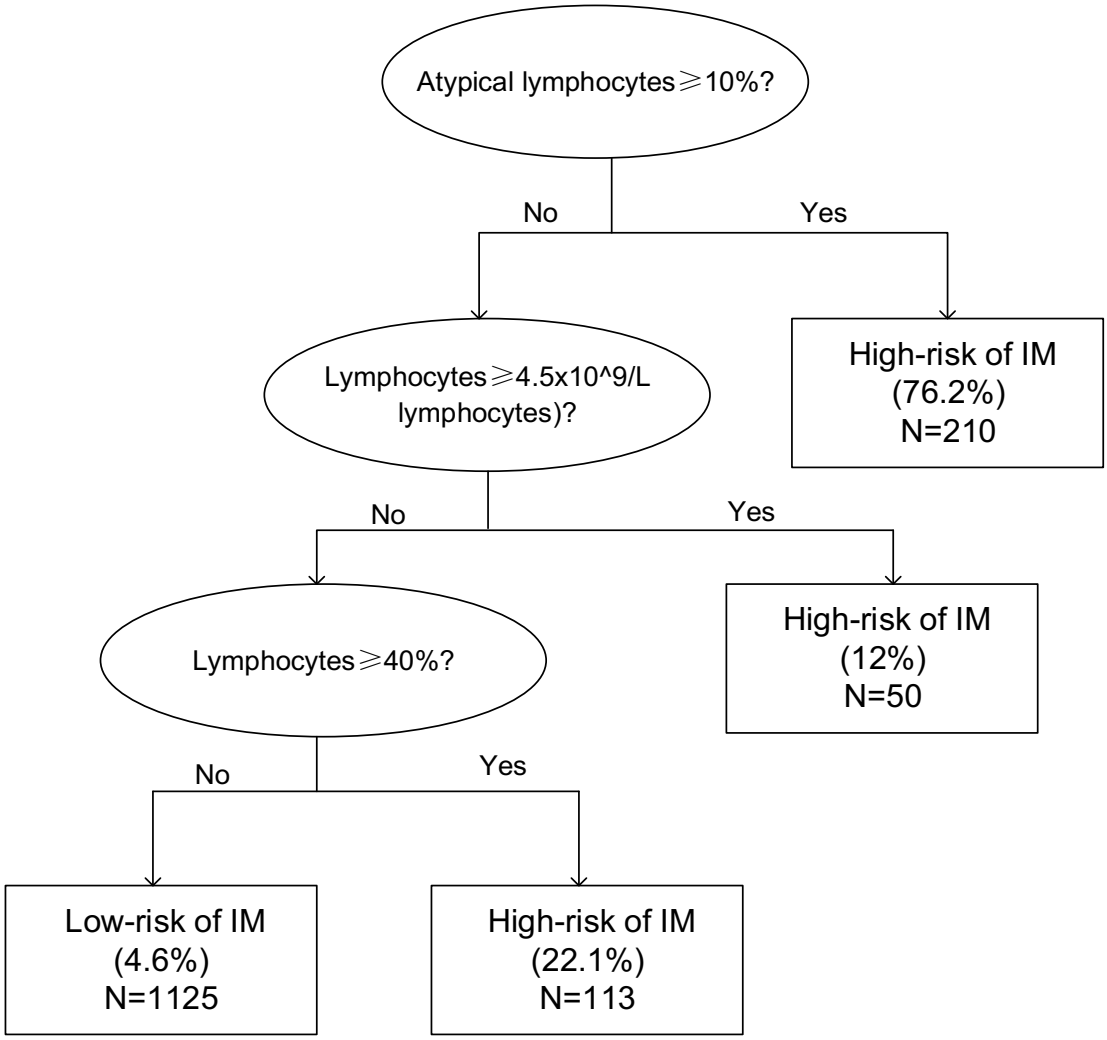
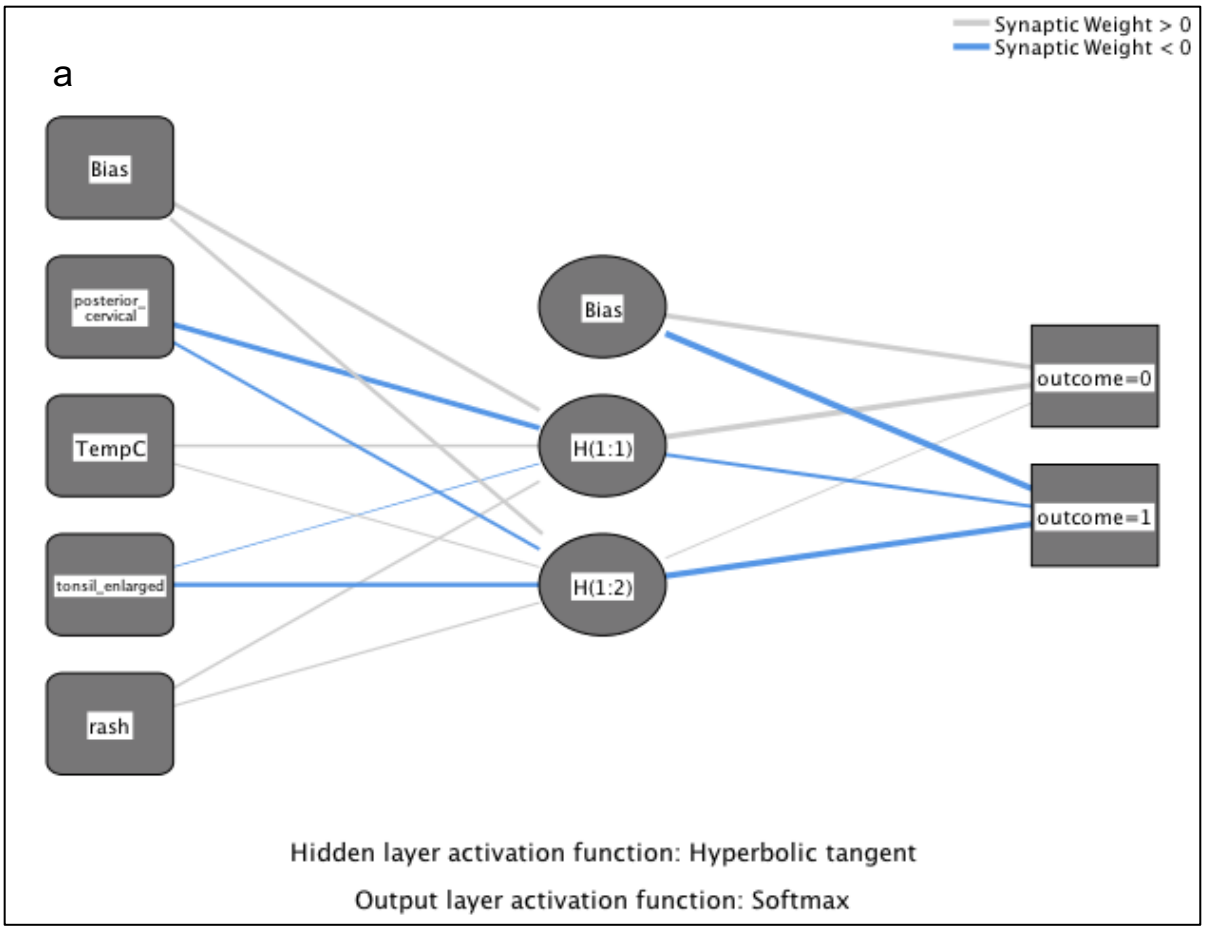
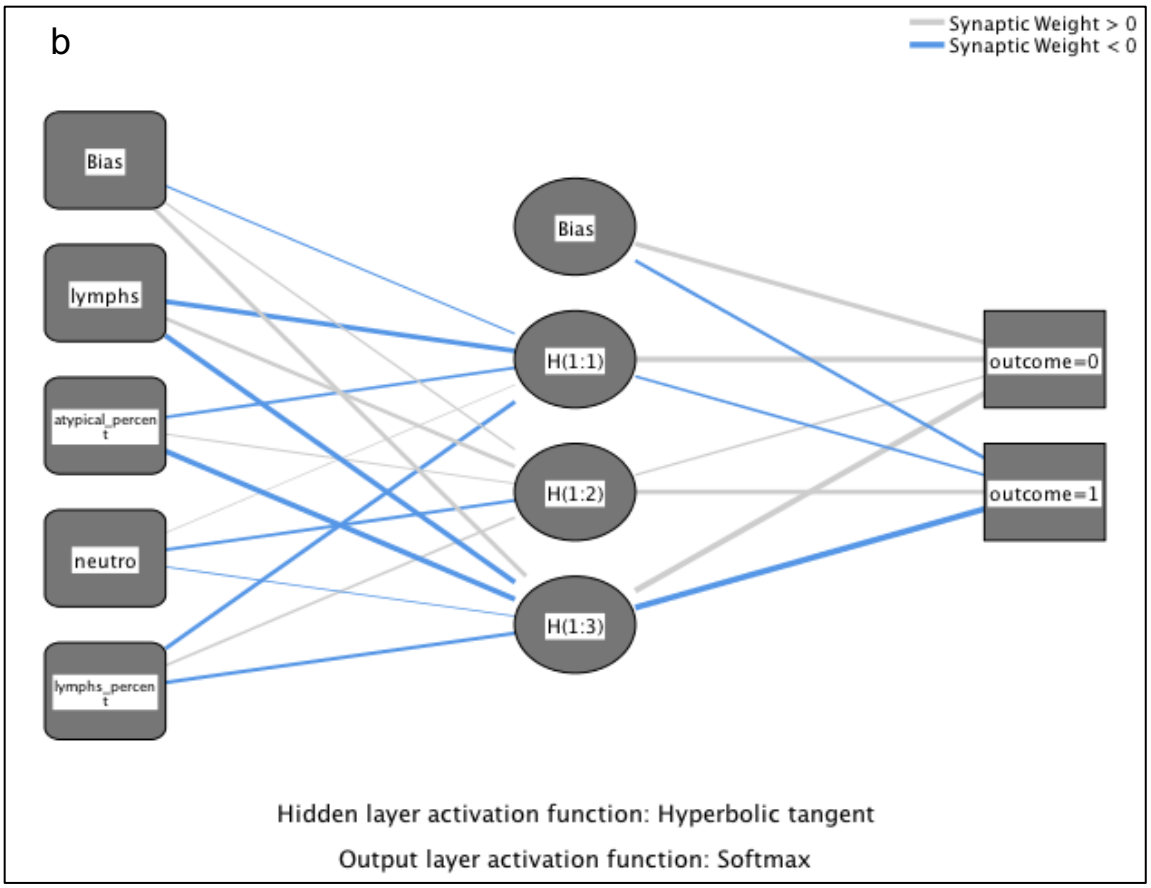


Figure 8.4 Fast and frugal tree for IM-Lab model in derivation cohort

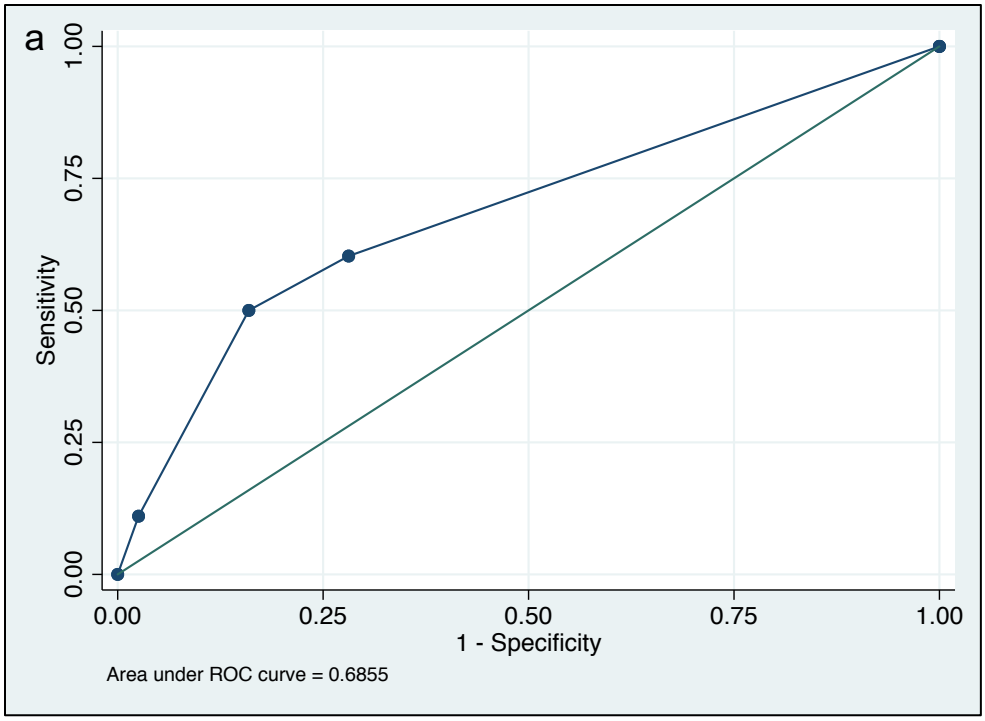


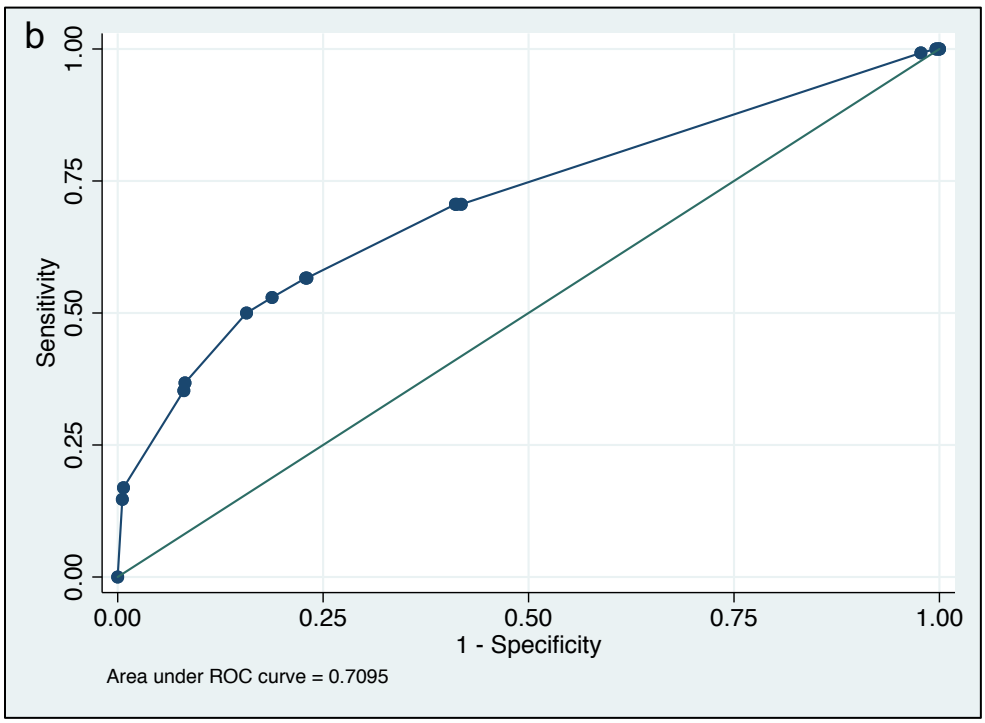
Note: posterior_cervical: posterior cervical lymphadenopathy, TempC: Body temperature ≥ 100.04 °F, tonsil_enlarged: Tonsillar enlargement, H(1:1) and H(1:2) are the parametric estimation in the hidden layer, outcome: 1=having IM, 0=not having IM.



Note: lymphs: absolute lymphocyte counts, atypical_percent: atypical lymphocyte percent, neutro: absolute neutrophil counts, lymphs_percent: lymphocytes percent; H(1:1), H(1:2) and H(1:3) are the parametric estimation in the hidden layer, outcome: 1=having IM, 0=not having IM.

Figure 8.5 Structure of artificial neural network (ANN): a) IM-Nolab model b) IM-Lab model


a
Sensitivity
0.00
0.25
0.50
0.75
1.00
0.00
0.25
0.50
0.75
1.00
1 - Specificity
Area under ROC curve = 0.6855


b
Sensitivity
0.00
0.25
0.50
0.75
1.00
0.00
0.25
0.50
0.75
1.00
1 - Specificity
Area under ROC curve = 0.7095

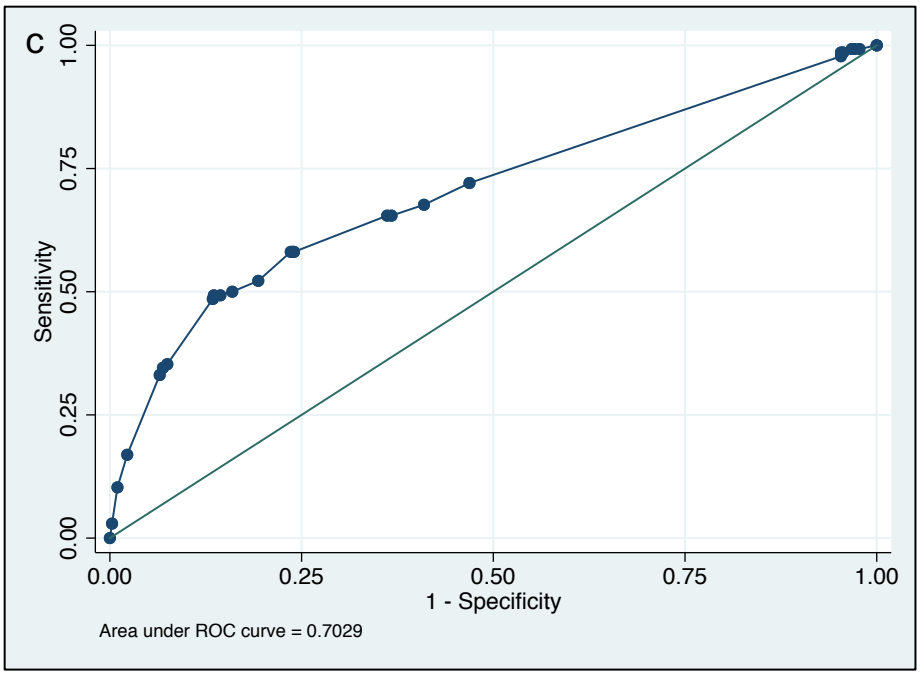
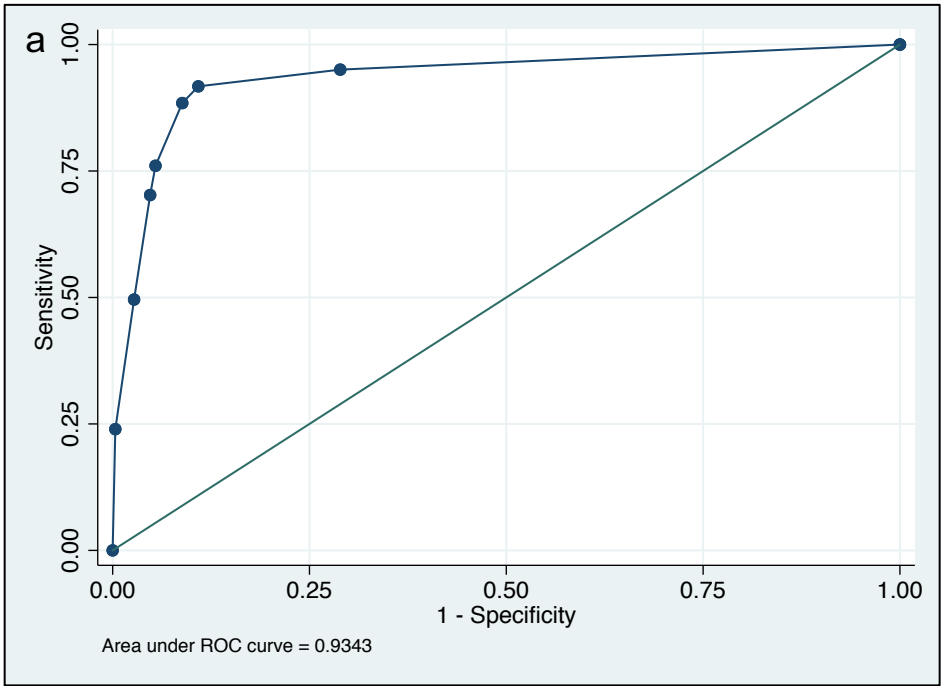
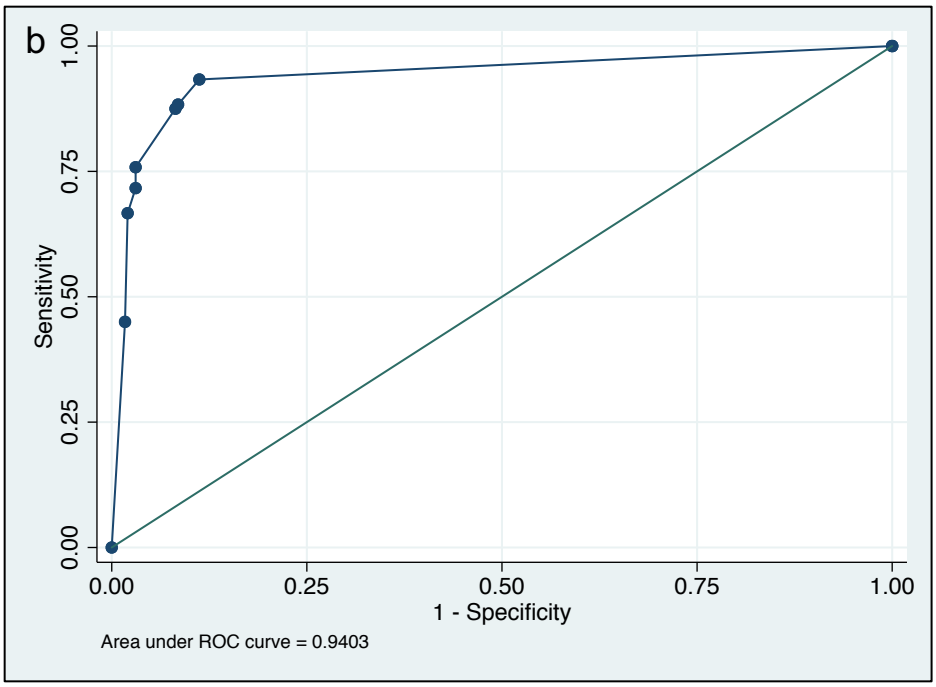


Figure 8.6 Receiver operating characteristic (ROC) curve of IM-Nolab model in the validation cohort: a) classification and regression tree; b) fast and frugal tree; c) artificial neural network





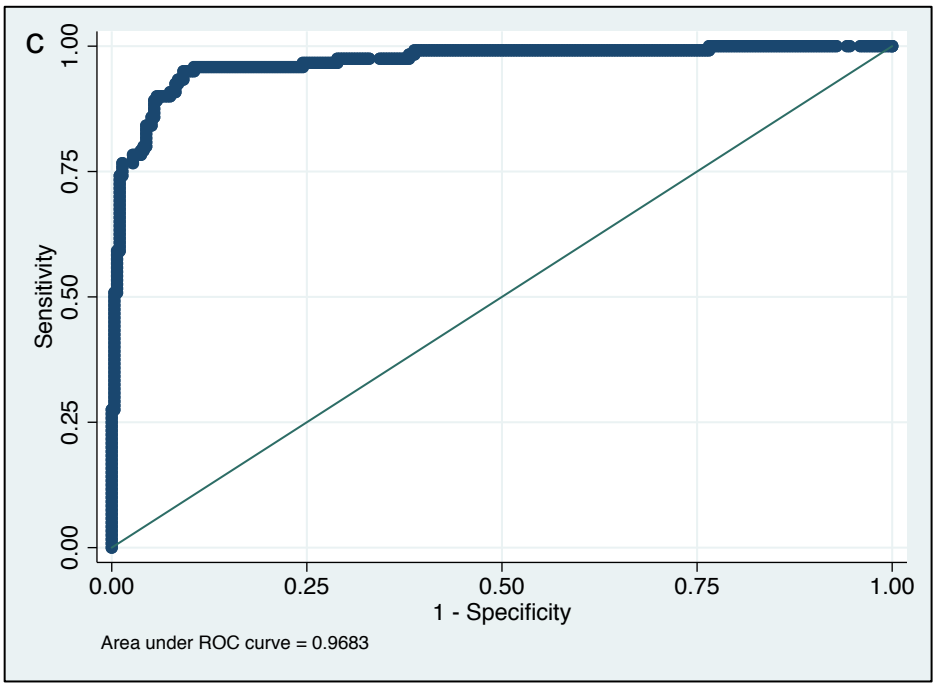


Figure 8.7 Receiver operating characteristic (ROC) curve of IM-Lab model in the validation cohort using: a) classification and regression tree; b) fast and frugal tree; c) artificial neural network

CHAPTER 7

DISSERTATION CONCLUSION

Although IM is a common condition among adolescents and young adult, there are limited well-designed prospective studies that address the diagnostic testing for IM. Previous studies have shown that the most common symptoms and signs of IM include sore throat, lymph node enlargement, fever, tonsillar enlargement, pharyngeal inflammation, transient palatal petechiae, and rashes.[7] However, the individual symptoms and signs are of limited value for the diagnosis of IM.[9,34] It is also found that using hematologic parameters are also helpful in diagnosing IM in clinically suspected patients.[34,41,45] So far, no study has been found to develop and validate a clinical prediction rule that combines symptoms, signs, and hematologic parameters to improve diagnosis of IM in patients with suspected IM. Also, the decision threshold for IM testing has not been studied or determined. If the test threshold for IM has been decided, a clinical prediction rule can be designed to classify patients as low-risk or high-risk for IM in order to help clinicians to make decisions on diagnostic testing.

Laboratory tests, such as the heterophile antibody tests and Epstein-Barr virus (EBV) specific antibody tests are promising, but each of these tests has their own inadequacy. Although the heterophile antibody tests are rapid, cost-effective, and simple compared to other serologic tests,[14] these tests are less sensitive than viral capsid antigen (VCA) tests, especially among younger children infected with EBV.[24] The VCA-IgM, VCA-IgG, and EBNA tests were developed for the detection of EBV-specific antibody responses.[3] However, the antibodies to VCA may not be produced at all or may only be produced for a short time; therefore, the results of the VCA tests may also be inaccurate.

The University Health Center (UHC) provides comprehensive healthcare for students at the University of Georgia (UGA). The UHC has electronic health records that maintain the details of the presence or absence of symptoms and signs, and the results of laboratory testing at every visit, therefore, it is appropriate data source to study IM.

Thus, we set out to: 1) Systematically review the literature and evaluate the diagnostic accuracy of individual symptoms, signs, hematologic parameters, and serologic tests for IM through meta-analysis; 2) determine the decision thresholds for ordering diagnostic tests for IM based on the cross-sectional survey of clinicians; 3) identify the symptoms, signs, and hematologic parameters that predict IM using the data collected from the UHC; and 4) use this information to develop and internally validate clinical prediction models for the diagnosis of IM by applying logistic regression analysis, classification and regression tree (CART), fast and frugal tree (FFT) and artificial neutral network (ANN) analysis.

Summary of Results

In Chapter 4, we conducted a systematic review and meta-analysis to determine the accuracy of symptoms, signs, and hematologic parameters in patients with suspected IM that used heterophile antibody test or viral capsid antigen tests as the reference standard. The likelihood of IM is appreciably increased by the presence of splenomegaly, palatal petechiae, posterior adenopathy, and axillary or inguinal adenopathy, while it is significantly decreased by any lymphadenopathy. Hematologic parameters are more useful in diagnosing IM, and an absolute lymphocyte count over 4×10^9/L, monocytes over 1×10^9/L, leukocytes over 5×10^9/L, as well as a higher percentage of lymphocytes and/or atypical lymphocytes are helpful for diagnosing IM. Since most of the clinical findings have limited diagnostic value in ruling out the disease when absent, physicians should not rely on the absence of any individual symptom or clinical sign for ruling out IM.

In Chapter 5, we conducted a systematic review and meta-analysis to determine the accuracy of serological tests in patients with suspected EBV infection that used immunofluorescent test as the reference standard test. Based on our findings, the most accurate tests to detect primary EBV infection include chemiluminescent microparticle immunoassay, enzyme-linked immunosorbent assay, and immunofiltration assay. We also found that the heterophile antibody tests had low sensitivity and were less accurate for ruling out IM when these tests are negative than EBV immunoassays. Both slide agglutination tests and solid-phase assays were much less sensitive in children. The chemiluminescent microparticle immunoassay is also relatively more accurate for detecting anti-VCA IgM or VCA IgG alone than other immunoassays. Therefore, the chemiluminescent microparticle immunoassay would be the most appropriate alternative to the immunofluorescent assay.

In Chapter 6, we estimated an overall test threshold for IM using the logistic regression as 9.5% (95% CI: 8.2% to 10.9%) based on a total of 819 clinical vignettes. We found that the test threshold for clinicians practicing greater than 10 years was significantly higher than for those practicing less or equal to 10 years (10.5% vs. 7.3%, p=0.02). No significant differences between specialties and practice sites were found with respect to the test threshold. Therefore, our estimated threshold was stable regarding the clinician's specialty and practice sites.

The results from the current threshold study can be used as guidance to develop a clinical prediction rule (CPR) for the management of IM. The CPR can classify patients at low and moderate risk of IM, corresponding to IM probabilities of less than 10% and greater than or equals to 10%, which are consistent with the test thresholds identified in our study. Our estimated threshold based on the realistic vignettes will not only increase the acceptability of the score development and classification, but also increase the efficiency in terms of the diagnostic testing in suspected IM cases.

In Chapter 7, we developed two prediction rules. The IM-Nolab score used only patients' symptoms and signs as independent predictors, and IM-Lab score added the hematologic

parameters as independent predictors to the model. The IM-NoLab risk score included posterior cervical adenopathy, rash, myalgias, headache, and tonsillar exudate and identified patients in the low-risk (8.8% IM) and high-risk (31.2%) groups in the validation dataset (AUC=0.76). The IM-Lab risk score included the atypical lymphocytes greater than 10%, the lymphocytes greater than 40%, tonsillar erythema, posterior cervical adenopathy, and swollen lymph nodes and identified low- and high-risk groups of 4% and 79.4% respectively in the validation group (AUC=0.94).

The IM-Nolab risk score could also be incorporated into a mobile application, where patients could do a self-assessment of their risk for IM and decide if further diagnostic testing is warranted. The IM-Lab score was designed to use minimal laboratory tests, and it provided more accurate results than the IM-Nolab. Both of the scores were well validated using internal validation approach. Upon external validation, these risk scores would be applicable for prioritizing diagnostic testing for IM in the outpatient setting.

In Chapter 8, we developed and internally validated decision support tools for IM-Nolab and IM-Lab models using three different innovative statistical approaches: classification tree (CART), fast and frugal tree (FFT), and artificial neural network (ANN). The IM-Nolab CART model has only two predictors; the IM-Lab CART model, IM-Nolab FFT model, and IM-Lab FFT model have three predictors; and the ANN models have five predictors. The probability of IM in the low- and high-risk groups in the validation cohort was 7.3% and 32.2% for IM-Nolab CART model (AUC=0.69); 5.9% and 61.8% for IM-Lab model (AUC=0.93); 8.2% and 33.5% for IM-Nolab FFT model (AUC=0.71); 5% and 68.2% for IM-Lab FFT model (AUC=0.94); 8.8% and 50.4% for IM-Nolab ANN model (AUC=0.70); and 4.4% and 69.3% for IM-Lab ANN model (AUC=0.97).

The derived IM-Lab and IM-Nolab models based on CART, FFT and ANN methods were internally validated, and the discrimination plots showed good discriminations for the IM-Nolab models and excellent discriminations for the IM-Lab models. The ANN models have poor

transparency and face validity for the decision rules, thus, they cannot be directly implemented by clinicians in the practice. We would consider that the IM-Nolab and IM-Lab CART models as the most acceptable models that could be implemented by clinicians in healthcare settings.

Implications and Future Research

Our study is the first study to estimate the decision threshold of IM testing. The test threshold estimated from our study was based on the limited numbers of simulated vignettes. In order to achieve a more accurate threshold, future study would observe the clinician's behavior with real patients in front of them and asking clinicians to estimate the probability of IM for each visit. This would allow the clinicians to account for the severity of symptoms, as well as the patients' attitude.

Based on our test threshold of IM, we successfully derived and internal validated CPRs for IM using a variety of statistical methods. However, our study population were recruited from the health center at the university, and most of our samples were young adults under 24 years of age. Future studies should externally validate our CPR with the populations from other age groups or from other locations in order to determine the generalizability of our prediction model. Also, our CPRs can result in different recommendations based on different pre-test probabilities of IM; therefore, the CPRs from our study should be linked to the prevalence data in the study population. The post-test probability for low- and high-risk groups of patients can be calculated according to the LRs estimated in our model and the prevalence of the population.

In addition, well-designed prospective studies are needed to investigate the role of the clinical findings and office-based tests among patients within a week of disease onset. There is also a need for studies to understand the natural history of IM and to identify the risk factors associated with the longer duration of illness among college students, which could better help clinicians to better understand the course of the disease.

REFERENCES

1. Andersson J. Clinical aspects on Epstein-Barr virus infection. *Scand J Infect Dis Suppl.* 1991;80:94-104.
2. Cohen JI. Epstein–Barr virus infection. *New England Journal of Medicine.* 2000;343(7):481-492.
3. De Paschale M, Clerici P. Serological diagnosis of Epstein-Barr virus infection: problems and solutions. *World journal of virology.* 2012;1(1):31.
4. HENKE CE, KURLAND LT, ELVEBACK LR. Infectious mononucleosis in Rochester, Minnesota, 1950 through 1969. *American journal of epidemiology.* 1973;98(6):483-490.
5. VAN CAUWENBERGE PB, VANDER MIJNSBRUGGE A-M. Pharyngitis: a survey of the microbiologic etiology. *The Pediatric infectious disease journal.* 1991;10(10):S39-42.
6. Candy B, Chalder T, Cleare AJ, Wessely S, White PD, Hotopf M. Recovery from infectious mononucleosis: a case for more than symptomatic therapy? A systematic review. *Br J Gen Pract.* 2002;52(483):844-851.
7. Hoagland RJ. Infectious mononucleosis. *Prim Care.* 1975;2(2):295-307.
8. White P, Thomas J, Amess J, et al. Incidence, risk and prognosis of acute and chronic fatigue syndromes and psychiatric disorders after glandular fever. *The British Journal of Psychiatry.* 1998;173(6):475-481.
9. Aronson MD, Komaroff AL, Pass TM, Ervin CT, Branch WT. Heterophil antibody in adults with sore throat: frequency and clinical presentation. *Annals of internal medicine.* 1982;96(4):505-508.
10. Paul JR, Bunnell W. The presence of heterophile antibodies in infectious mononucleosis. *The American Journal of the Medical Sciences.* 1932;183(1):90-103.

11. Davidsohn I. Serologic diagnosis of infectious mononucleosis. *Journal of the American Medical Association.* 1937;108(4):289-295.

12. Linderholm M, Boman J, Juto P, Linde A. Comparative evaluation of nine kits for rapid diagnosis of infectious mononucleosis and Epstein-Barr virus-specific serology. *Journal of clinical microbiology.* 1994;32(1):259-261.

13. Fleisher G, Collins M, Fager S. Limitations of available tests for diagnosis of infectious mononucleosis. *Journal of clinical microbiology.* 1983;17(4):619-624.

14. Ebell MH. Epstein-Barr virus infectious mononucleosis. *American family physician.* 2004;70:1279-1292.

15. Sprunt TP, Evans FA. Mononuclear leucocytosis in reaction to acute infections ("infectious mononucleosis"). *Bull Johns Hopkins Hosp.* 1920;31:410-417.

16. Evans A. Infectious mononucleosis and related syndromes. *The American journal of the medical sciences.* 1978;276(3):325-339.

17. Maki DG, Reich RM. Infectious mononucleosis in the athlete: diagnosis, complications, and management. *The American journal of sports medicine.* 1982;10(3):162-173.

18. Cohen J. Clinical aspects of Epstein-Barr virus infection. *Epstein-Barr Virus.* 2005;1:35-54.

19. Ebell MH. Epstein-Barr virus infectious mononucleosis. *Am Fam Physician.* 2004;70(7):1279-1287.

20. Gerber P, Walsh J, Rosenblum E, Purcell R. Association of EB-virus infection with the post-perfusion syndrome. *The Lancet.* 1969;293(7595):593-596.

21. Balfour HH, Jr., Sifakis F, Sliman JA, Knight JA, Schmeling DO, Thomas W. Age-Specific Prevalence of Epstein–Barr Virus Infection Among Individuals Aged 6–19 Years in the United States and Factors Affecting Its Acquisition. *The Journal of Infectious Diseases.* 2013;208(8):1286-1293.

22. Henle W, Henle G. Seroepidemiology of the virus. In: *The Epstein-Barr Virus.* Springer;

1979:61-78.

23. Grose C. The many faces of infectious mononucleosis: the spectrum of Epstein-Barr virus infection in children. *Pediatr Rev.* 1985;7(2):35-44.

24. Sumaya CV, Ench Y. Epstein-Barr virus infectious mononucleosis in children: I. Clinical and general laboratory findings. *Pediatrics.* 1985;75(6):1003-1010.

25. Odumade OA, Knight JA, Schmeling DO, Masopust D, Balfour HH, Hogquist KA. Primary Epstein-Barr virus infection does not erode preexisting CD8+ T cell memory in humans. *Journal of Experimental Medicine.* 2012;209(3):471-478.

26. Hallee TJ, Evans AS, Niederman JC, Brooks CM, Voegtly JH. Infectious mononucleosis at the United States Military Academy. A prospective study of a single class over four years. *The Yale journal of biology and medicine.* 1974;47(3):182.

27. Niederman J, Evans A, Subrahmanyan L, McCollum R. Prevalence, incidence and persistence of EB virus antibody in young adults. *New England Journal of Medicine.* 1970;282(7):361-365.

28. Sawyer R, Evans A, Niederman J, McCollum R. Prospective studies of a group of Yale University freshmen. I. Occurrence of infectious mononucleosis. *The Journal of infectious diseases.* 1971;123(3):263-270.

29. Crawford DH, Macsween KF, Higgins CD, et al. A cohort study among university students: identification of risk factors for Epstein-Barr virus seroconversion and infectious mononucleosis. *Clinical infectious diseases.* 2006;43(3):276-282.

30. Balfour Jr HH, Odumade OA, Schmeling DO, et al. Behavioral, virologic, and immunologic factors associated with acquisition and severity of primary Epstein–Barr virus infection in university students. *The Journal of infectious diseases.* 2012;207(1):80-88.

31. Luzuriaga K, Sullivan JL. Infectious Mononucleosis. *New England Journal of Medicine.* 2010;362(21):1993-2000.

32. Rea TD, Russo JE, Katon W, Ashley RL, Buchwald DS. Prospective study of the natural

history of infectious mononucleosis caused by Epstein-Barr virus. *J Am Board Fam Pract.* 2001;14(4):234-242.

33. Stevens JE, Bayrd ED, Heck FJ. Infectious mononucleosis: a study of 210 sporadic cases. *The American journal of medicine.* 1951;11(2):202-208.

34. Grotto I, Mimouni D, Huerta M, et al. Clinical and laboratory presentation of EBV positive infectious mononucleosis in young adults. *Epidemiology & Infection.* 2003;131(1):683-689.

35. Ebell MH, Call M, Shinholser J, Gardner J. Does This Patient Have Infectious Mononucleosis?: The Rational Clinical Examination Systematic ReviewDiagnosing Infectious MononucleosisDiagnosing Infectious Mononucleosis. *Jama.* 2016;315(14):1502-1509.

36. Schooley R. Epstein-Barr virus infections, including infectious mononucleosis. *Harrison's principles of internal medicine.* 1987:699-703.

37. Horwitz CA, Henle W, Henle G, et al. Clinical and laboratory evaluation of elderly patients with heterophil-antibody positive infectious mononucleosis: Report of seven patients, ages 40 to 78. *The American journal of medicine.* 1976;61(3):333-339.

38. Carter J, Edson R, Kennedy C. Infectious mononucleosis in the older patient. Paper presented at: Mayo Clinic Proceedings1978.

39. Maki DG, Reich RMJTAjosm. Infectious mononucleosis in the athlete: diagnosis, complications, and management. 1982;10(3):162-173.

40. Dalrymple W. Infectious mononucleosis: 2. Relation of bed rest and activity to prognosis. *Postgraduate medicine.* 1964;35(4):345-349.

41. Hoagland RJ. The Clinical Manifestations of Infectious Mononucleosis. A Report of Two Hundred Cases. *American Journal of Medical Sciences.* 1960;240(1):21-29.

42. Shapiro MF, Greenfield S. The complete blood count and leukocyte differential count. An approach to their rational application. *Common Diagnostic Tests: Use and Interpretation 2d ed Philadelphia: American College of Physicians.* 1990:183-203.

43. Wintrobe MM. Clinical hematology. *Academic Medicine.* 1962;37(1):78.

44. Brigden ML, Au S, Thompson S, Brigden S, Doyle P, Tsaparas Y. Infectious mononucleosis in an outpatient population: diagnostic utility of 2 automated hematology analyzers and the sensitivity and specificity of Hoagland's criteria in heterophile-positive patients. *Archives of pathology & laboratory medicine.* 1999;123(10):875-881.

45. Rice L, Jung M. Neutrophilic leukocytosis, neutropenia, monocytosis, and monocytopenia. In: *Hematology.* Elsevier; 2018:675-681.

46. Biggs TC, Hayes SM, Bird JH, Harries PG, Salib RJ. Use of the lymphocyte count as a diagnostic screen in adults with suspected Epstein-Barr virus infectious mononucleosis. *The Laryngoscope.* 2013;123(10):2401-2404.

47. Ebell MH, Afonso AM, Gonzales R, Stein J, Genton B, Senn N. Development and validation of a clinical decision rule for the diagnosis of influenza. *The Journal of the American Board of Family Medicine.* 2012;25(1):55-62.

48. Ebell MH. Strep throat. *American family physician.* 2003;68(5):937.

49. Ebell MH. Predicting pneumonia in adults with respiratory illness. *American family physician.* 2007;76(4):560.

50. Fine AM, Nizet V, Mandl KD. Large-scale validation of the Centor and McIsaac scores to predict group A streptococcal pharyngitis. *Archives of internal medicine.* 2012;172(11):847-852.

51. Chetham MM, Roberts KB. Infectious mononucleosis in adolescents. *Pediatric annals.* 1991;20(4):206-213.

52. Paul JR, Bunnell WW. Classics in infectious diseases. The presence of heterophile antibodies in infectious mononucleosis by John R. Paul and W. W. Bunnell. American Journal of the Medical Sciences, 1932. *Reviews of infectious diseases.* 1982;4(5):1062-1068.

53. Henle G, Henle W, Diehl V. Relation of Burkitt's tumor-associated herpes-ytpe virus to

infectious mononucleosis. *Proceedings of the National Academy of Sciences of the United States of America.* 1968;59(1):94.

54. Elgh F, Linderholm M. Evaluation of six commercially available kits using purified heterophile antigen for the rapid diagnosis of infectious mononucleosis compared with Epstein-Barr virus-specific serology. *Clinical and diagnostic virology.* 1996;7(1):17-21.

55. Bruu A-L, Hjetland R, Holter E, et al. Evaluation of 12 commercial tests for detection of Epstein-Barr virus-specific and heterophile antibodies. *Clin Diagn Lab Immunol.* 2000;7(3):451-456.

56. Svahn A, Magnusson M, Jägdahl L, Schloss L, Kahlmeter G, Linde A. Evaluation of three commercial enzyme-linked immunosorbent assays and two latex agglutination assays for diagnosis of primary Epstein-Barr virus infection. *Journal of clinical microbiology.* 1997;35(11):2728-2732.

57. Robertson P, Beynon S, Whybin R, et al. Measurement of EBV-IgG anti-VCA avidity aids the early and reliable diagnosis of primary EBV infection. *Journal of medical virology.* 2003;70(4):617-623.

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

59. McGinn TG, Guyatt GH, Wyer PC, et al. Users' guides to the medical literature: XXII: how to use articles about clinical decision rules. *Jama.* 2000;284(1):79-84.

60. Steyerberg EW. *Clinical prediction models.* Vol 381: Springer; 2009.

61. Ebell MH, Locatelli I, Senn N. A novel approach to the determination of clinical decision thresholds. *BMJ Evidence-Based Medicine.* 2015;20(2):41-47.

62. Katz BZ, Jason LA. Chronic fatigue syndrome following infections in adolescents. *Curr Opin Pediatr.* 2013;25(1):95-102.

63. HEATH JR CW, BRODSKY AL, POTOLSKY AI. Infectious mononucleosis in a general

population. *American journal of epidemiology.* 1972;95(1):46-52.

64. Davidson R. A survey of infectious mononucleosis in the North-East Regional Hospital Board area of Scotland, 1960–9. *Epidemiology & Infection.* 1970;68(3):393-400.

65. Evans AS. Infectious mononucleosis. Observations from a public health laboratory. *The Yale journal of biology and medicine.* 1961;34(3-4):261.

66. Hobson F, Lawson B, Wigfield M. Glandular fever: a field study. *British medical journal.* 1958;1(5075):845.

67. Ström J. Infectious Mononucleosis-is the Incidence Increasing? *Acta Medica Scandinavica.* 1960;168(1):35-39.

68. Belfrage S. Infectious mononucleosis. An epidemiological and clinical study. *Acta Medica Scandinavica.* 1962;171(5):531-541.

69. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of internal medicine.* 2011;155(8):529-536.

70. Reitsma J, Rutjes A, Whiting P, Vlassov V, Leeflang M. Assessing methodological quality. In: Deeks JJ, Bossuyt PM, Gatsonis C (editors), Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0. 0. The Cochrane Collaboration, 2009. *Available from: http://srdta cochrane org/.* 2009.

71. Doebler P. mada: Meta-Analysis of Diagnostic Accuracy. R package version 0.5. 8.[Internet]. 2017. In.

72. Team RC. R: A language and environment for statistical computing. *R Foundation for Statistical*

Computing. 2013.

73. Metz CE. Basic principles of ROC analysis. Paper presented at: Seminars in nuclear medicine1978.

74. Paramita DK, Fachiroh J, Haryana SM, Middeldorp JM. Two-step Epstein-Barr virus

immunoglobulin A enzyme-linked immunosorbent assay system for serological screening and confirmation of nasopharyngeal carcinoma. *Clinical and vaccine immunology.* 2009;16(5):706-711.

75. Van Houwelingen HC, Zwinderman KH, Stijnen T. A bivariate approach to meta-analysis. *Statistics in medicine.* 1993;12(24):2273-2284.

76. Ebell MH, Locatelli I, Mueller Y, Senn N, Morgan K. Diagnosis and treatment of community-acquired pneumonia in patients with acute cough: a quantitative study of decision thresholds in primary care. *Br J Gen Pract.* 2018;68(676):e765-e774.

77. University Health Center UoG. Accreditation. 2020; https://www.uhs.uga.edu/about/accreditation, May 1, 2020.

78. Tilton RC, Dias F, Ryan RW. Comparative evaluation of three commercial tests for detection of heterophile antibody in patients with infectious mononucleosis. *Journal of clinical microbiology.* 1988;26(2):275-278.

79. Lewis RJ. An introduction to classification and regression tree (CART) analysis. Paper presented at: Annual meeting of the society for academic emergency medicine in San Francisco, California2000.

80. Afonso AM, Ebell MH, Gonzales R, Stein J, Genton B, Senn N. The use of classification and regression trees to predict the likelihood of seasonal influenza. *Family practice.* 2012;29(6):671-677.

81. Phillips ND, Neth H, Woike JK, Gaissmaier W. FFTrees: A toolbox to create, visualize, and evaluate fast-and-frugal decision trees. *Judgment and Decision Making.* 2017;12(4):344-368.

82. Tu JV. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of clinical epidemiology.* 1996;49(11):1225-1231.

83. Liu J, Tang Z-H, Zeng F, Li Z, Zhou L. Artificial neural network models for prediction of

cardiovascular autonomic dysfunction in general Chinese population. *BMC medical informatics and decision making.* 2013;13(1):80.

84. Hutton B, Salanti G, Caldwell DM, et al. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. *Annals of internal medicine.* 2015;162(11):777-784.

85. Balfour HH, Jr., Holman CJ, Hokanson KM, et al. A prospective clinical study of Epstein-Barr virus and host interactions during acute infectious mononucleosis. *J Infect Dis.* 2005;192(9):1505-1512.

86. Sumaya CV, Ench Y. Epstein-Barr virus infectious mononucleosis in children. I. Clinical and general laboratory findings. *Pediatrics.* 1985;75(6):1003-1010.

87. Ventura KC, Hudnall SD. Hematologic differences in heterophile-positive and heterophile-negative infectious mononucleosis. *American journal of hematology.* 2004;76(4):315-318.

88. Ginsburg CM, Henle W, Henle G, Horwitz CA. Infectious mononucleosis in children. Evaluation of Epstein-Barr virus-specific serological data. *Jama.* 1977;237(8):781-785.

89. Chretien J, Esswein J, Holland W, McCauley C. Predictors of the duration of infectious mononucleosis. *Southern medical journal.* 1977;70(4):437-439.

90. Hossain A. Infectious mononucleosis and mononucleosis-like illnesses in children and adults in Saudi Arabia. *Journal of tropical pediatrics.* 1989;35(3):121-125.

91. Grotto I, Mimouni D, Huerta M, et al. Clinical and laboratory presentation of EBV positive infectious mononucleosis in young adults. *Epidemiology and infection.* 2003;131(1):683-689.

92. Llor C, Hernández M, Hernández S, Martínez T, Gómez FF. Validity of a point-of-care based on heterophile antibody detection for the diagnosis of infectious mononucleosis in primary care. *The European journal of general practice.* 2012;18(1):15-21.

93. Krabbe S, Hesse J, Uldall P. Primary Epstein-Barr virus infection in early childhood. *Archives of disease in childhood.* 1981;56(1):49-52.

94. Ho-Yen D, Martin KW. The relationship between atypical lymphocytosis and serological tests in infectious mononucleosis. *Journal of Infection.* 1981;3(4):324-331.

95. Gartzonika C, Vrioni G, Priavali E, Pappas G, Levidiotou S. Utility of real-time PCR in the diagnosis of primary Epstein-Barr virus infection. *J Med Microb Diagn.* 2012;1(118):2161-0703.1000118.

96. Lennon P, JP ON, Fenton JE, T OD. Challenging the use of the lymphocyte to white cell count ratio in the diagnosis of infectious mononucleosis by analysis of a large cohort of Monospot test results. *Clinical otolaryngology : official journal of ENT-UK ; official journal of Netherlands Society for Oto-Rhino-Laryngology & Cervico-Facial Surgery.* 2010;35(5):397-401.

97. Del CM, Pincus D. Incidence patterns of respiratory illness in Queensland estimated from sentinel general practice. *Australian family physician.* 1995;24(4):625-629, 632.

98. Linde A. Diagnosis and pathogenesis of infectious mononucleosis and other Epstein Barr virus associated diseases. *Rev Med Microbiol.* 1992;3:43-51.

99. Hess RD. Routine Epstein-Barr virus diagnostics from the laboratory perspective: still challenging after 35 years. *J Clin Microbiol.* 2004;42(8):3381-3387.

100. Lupo J, Germi R, Semenova T, Buisson M, Seigneurin JM, Morand P. Performance of two commercially available automated immunoassays for the determination of Epstein-Barr virus serological status. *Clinical and vaccine immunology : CVI.* 2012;19(6):929-934.

101. Henle G, Henle W. Immunofluorescence in cells derived from Burkitt's lymphoma. *Journal of bacteriology.* 1966;91(3):1248-1256.

102. Kleines M, Scheithauer S, Ritter K, Häusler M. Sensitivity of the Enzygnost anti-EBV/IgG for the determination of the Epstein-Barr virus immune status in pediatric patients. *Diagn Microbiol Infect Dis.* 2006;55(3):247-249.

103. Corrales I, Giménez E, Navarro D. Evaluation of the Architect Epstein-Barr Virus (EBV) viral capsid antigen (VCA) IgG, VCA IgM, and EBV nuclear antigen 1 IgG

chemiluminescent immunoassays for detection of EBV antibodies and categorization of EBV infection status using immunofluorescence assays as the reference method. *Clinical and vaccine immunology : CVI.* 2014;21(5):684-688.

104. Kasifoglu N, Oz S, Dinleyici EC, et al. Comparison of Methods Used for the Diagnosis of Epstein-Barr Virus Infections in Children. *Polish journal of microbiology.* 2018;67(1):81-88.

105. Ho DW, Field PR, Cunningham AL. Rapid diagnosis of acute Epstein-Barr virus infection by an indirect enzyme-linked immunosorbent assay for specific immunoglobulin M (IgM) antibody without rheumatoid factor and specific IgG interference. *J Clin Microbiol.* 1989;27(5):952-958.

106. Levin MJ, Weinstein MP, Sumaya CV, et al. The rapid diagnosis of infectious mononucleosis using an ELISA that detects IgM antibody to a peptide component of Epstein-Barr virus nuclear antigen. *Diagn Microbiol Infect Dis.* 1991;14(4):287-291.

107. Wiedbrauk DL, Bassin S. Evaluation of five enzyme immunoassays for detection of immunoglobulin M antibodies to Epstein-Barr virus viral capsid antigens. *J Clin Microbiol.* 1993;31(5):1339-1341.

108. Gómez MC, Nieto JA, Escribano MA. Evaluation of two slide agglutination tests and a novel immunochromatographic assay for rapid diagnosis of infectious mononucleosis. *Clinical and diagnostic laboratory immunology.* 2000;7(5):840-841.

109. Altuglu I, Bozkurt H, Samlioglu P, Zeytinoglu A. Evaluation of three different assays for the assessment of Epstein Barr Virus immunological status. *The new microbiologica.* 2007;30(4):393-398.

110. Fleisher GR, Collins M, Fager S. Limitations of available tests for diagnosis of infectious mononucleosis. *J Clin Microbiol.* 1983;17(4):619-624.

111. Smalley DL, Dunn CE, Hechinger MC. Enzyme immunoassay for antibodies to Epstein-Barr virus nuclear antigen. *Clinical laboratory science : journal of the American Society for Medical Technology.* 1990;3(6):397-398.

112. Fung MK, Mordarski KT, Bader SA, Gronowski AM. Evaluation of the Wampole Laboratories ELISA-based assay for Epstein-Barr virus serology. *Clinica chimica acta; international journal of clinical chemistry.* 2002;319(1):43-48.

113. Al Sidairi H, Binkhamis K, Jackson C, et al. Comparison of two automated instruments for Epstein-Barr virus serology in a large adult hospital and implementation of an Epstein-Barr virus nuclear antigen-based testing algorithm. *Journal of medical microbiology.* 2017;66(11):1628-1634.

114. Grandjean Lapierre S, Vallières E, Rabaamad L, Labrecque M, Chartrand C, Renaud C. Evaluation of the abbot Architect(™) epstein-barr virus viral capsid antigen IgM, viral capsid antigen IgG and nuclear antigen IgG assays in a pediatric and adult population. *Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology.* 2016;81:1-5.

115. Chan KH, Luo RX, Chen HL, Ng MH, Seto WH, Peiris JS. Development and evaluation of an Epstein-Barr virus (EBV) immunoglobulin M enzyme-linked immunosorbent assay based on the 18-kilodalton matrix protein for diagnosis of primary EBV infection. *J Clin Microbiol.* 1998;36(11):3359-3361.

116. de Ory F, Guisasola ME, Sanz JC, García-Bermejo I. Evaluation of four commercial systems for the diagnosis of Epstein-Barr virus primary infections. *Clinical and vaccine immunology : CVI.* 2011;18(3):444-448.

117. Gerber MA, Shapiro ED, Ryan RW, Bell GL. Evaluations of enzyme-linked immunosorbent assay procedure for determining specific Epstein-Barr virus serology and of rapid test kits for diagnosis for infectious mononucleosis. *Journal of clinical microbiology.* 1996;34(12):3240-3241.

118. Feng Z, Li Z, Sui B, Xu G, Xia T. Serological diagnosis of infectious mononucleosis by chemiluminescent immunoassay using capsid antigen p18 of Epstein-Barr virus. *Clinica chimica acta.* 2005;354(1-2):77-82.

119. Jiang SY, Yang JW, Shao JB, Liao XL, Lu ZH, Jiang H. Real-time polymerase chain reaction for diagnosing infectious mononucleosis in pediatric patients: A systematic review and meta-analysis. *Journal of medical virology.* 2016;88(5):871-876.

120. Bailey R. Diagnosis and treatment of infectious mononucleosis. *American family physician.* 1994;49(4):879-888.

121. Pauker SG, Kassirer JP. The threshold approach to clinical decision making. *New England Journal of Medicine.* 1980;302(20):1109-1117.

122. Hall KH. Reviewing intuitive decision-making and uncertainty: the implications for medical education. *Medical education.* 2002;36(3):216-224.

123. Ramagopalan SV, Giovannoni G, Yeates DG, Seagroatt V, Goldacre MJ. Sex ratio of infectious mononucleosis and possible relevance to multiple sclerosis. *Multiple Sclerosis Journal.* 2013;19(3):359-361.

124. Montalto NJ. An office-based approach to influenza: clinical diagnosis and laboratory testing. *American family physician.* 2003;67(1):111-118.

125. Ebell MH, Afonso AM, Geocadin RG. Prediction of survival to discharge following cardiopulmonary resuscitation using classification and regression trees. *Critical care medicine.* 2013;41(12):2688-2697.

126. Ebell MH, Hansen JG. Proposed clinical decision rules to diagnose acute rhinosinusitis among adults in primary care. *The Annals of Family Medicine.* 2017;15(4):347-354.

127. Marshall RJ. The use of classification and regression trees in clinical epidemiology. *Journal of clinical epidemiology.* 2001;54(6):603-609.

128. Hinton GE. How neural networks learn from experience. *Scientific American.* 1992;267(3):144-151.

129. Patil S, Henry JW, Rubenfire M, Stein PD. Neural network in the clinical diagnosis of acute pulmonary embolism. *Chest.* 1993;104(6):1685-1689.

130. Baxt WG. Use of an artificial neural network for the diagnosis of myocardial infarction.

Annals of internal medicine. 1991;115(11):843-848.

131. Ebell MH. Artificial neural networks for predicting failure to survive following in-hospital cardiopulmonary resuscitation. *Journal of family practice.* 1993;36:297-297.

132. Greenwood D. An overview of neural networks. *Behavioral science.* 1991;36(1):1-33.

133. Shahid N, Rappon T, Berta W. Applications of artificial neural networks in health care organizational decision-making: A scoping review. *PloS one.* 2019;14(2):e0212356.

134. Beeley C. *Web application development with R using Shiny.* Packt Publishing Ltd; 2013.

135. Ruff ZJ, Lesmeister DB, Appel CL, Sullivan CM. A Convolutional Neural Network and R-Shiny App for Automated Identification and Classification of Animal Sounds. *bioRxiv.* 2020.

136. Patel VL, Shortliffe EH, Stefanelli M, et al. The coming of age of artificial intelligence in medicine. *Artificial intelligence in medicine.* 2009;46(1):5-17.

APPENDICES

Appendix A. Definition Questions Based on the QUADAS-2 Instruments

Patient selection
1. Was a consecutive sample or non-random sample of patients enrolled?
Y: Study enrolled consecutive patients or a non-random sample of patients that clinically suspected with infectious mononucleosis from inpatient or outpatient setting, or from bank of sera compatible with primary EBV infection.
N: A random sample or non-consecutive sample was used.
U: Uncertain

2. Was the study designed to avoid a case-control design (Y/N/U)?
Y: The study population was drawn from a prospective study (accuracy of symptoms, signs, or hematologic parameters) or from a prospective cohort or retrospective cohort (accuracy of diagnostic tests) that included patients with suspected infectious mononucleosis or suspected EBV infection
N: The study population consisted of patients with known EBV disease and separate groups with other confirmed infections (e.g. RA, toxoplasmosis) and/or healthy controls
U: Uncertain

3. Did the study design avoid inappropriate exclusion criteria (Y/N/U)?
Y: There were no inappropriate exclusion criteria, such as excluding those with uncertain findings.
N: The study used inappropriate exclusion criteria.
U: Uncertain

4. Patient Selection Risk of Bias: What is the likelihood that patient selection could have introduced bias (L/H/U)?
L: Low likelihood of bias due to patient selection or enrollment (“Yes” to 1, 2 and 3)
H: High likelihood of bias due to patient selection (“No” to 1, 2 or 3)
U: Unable to judge degree of bias.

5. Concerns About Patient Selection Applicability: Are there concerns that included patients and setting do not match the review question?
L: Low risk of bias, the patients or settings are from the inpatient or outpatient setting and samples are clinically suspected with IM or confirmed with IM by the reference standard test
H: High risk of bias, the patients or settings do not match the review question, for example, the study focused on a group of patients with IM that caused by the infection other than EBV, or selected patients had other acute illness consistent with EBV related infectious mononucleosis, transplant recipients, HIV subjects, or immunodeficiency subjects
U: Uncertain

Index Test
6. Were index test results interpreted without knowledge of reference standard?
Y: Yes. Clearly stated that index test was performed without knowledge of reference standard
N: No (including when index and reference std were performed by same observer, although blinding was not addressed)
U: Uncertain

7. If a threshold or cutoff was used to define an abnormal index test, was it pre-specified?
Y: The threshold was pre-specified or there was no threshold mentioned

N: The threshold was established post-hoc
U: A threshold was used but it is not clear when it was specified

8. Index Test Risk of Bias: What is the likelihood that conduct of the index test could have introduced bias (L, H, U)?
L: Low likelihood of bias due to failure to mask reference std (“Yes” or “Uncertain to 6, and “Yes” to 7)
H: High likelihood of bias due to failure to mask to reference standard (“No” to 6 or “No” or “Uncertain” to 7)
U: Uncertain

9. Concerns Regarding Index Test Applicability: Are there concerns that the index test differs from those specified in the review question?
L: Low likelihood: the index test in this study is a sign, symptom, hematologic parameter or serological diagnostic test
H: High likelihood: the index test in this study may not belongs to any sign, symptom, hematologic parameter or serological diagnostic test
U: Uncertain

Reference Standard Test
10. Is the reference standard likely to correctly classify patients as having infectious mononucleosis or acute EBV infection?
Y: Yes, used the commercial kit of EBV-specific test or heterophile antibody test as the reference standard test for the accuracy of symptoms, signs or hematologic parameters.
N: No, used another reference test.
U: Uncertain

11. Was the reference standard interpreted without knowledge of the index test?
Y: Yes, reference standard interpretation was masked to index test results
N: No, reference standard interpretation not masked to index test results
U: Uncertain, no mention is made of masking the person performing the reference standard

12. Reference Standard Risk of Bias: Could conduct or interpretation of the reference standard could have introduced bias?
L: Low likelihood of bias due to the reference standard (“Yes” to 10, “Yes” or “Uncertain” to 11)
H: High likelihood of bias due to inadequate reference standard (“No” or “Uncertain” to 10, “No”.to 11)
U: Uncertain

13. Concerns Regarding Applicability of the Reference Standard: are there concerns that the target conditions defined by the reference standard do not match the review question?
L: Low likelihood of bias, i.e. the reference standard was intended to detect infectious mononucleosis or acute EBV infection
H: High likelihood of bias, i.e. the reference standard was not intended to detect infectious mononucleosis or acute EBV infection
U: Uncertain

Patient Flow and Timing
14. Did all patients receive a reference standard?
Y: Yes, all patients received some sort of reference standard (no partial verification bias)
N: No, some patients did not receive any reference standard (partial verification bias)
U: Uncertain

15. Did all patients receive the same reference standard?
Y: Yes, all used the same reference standard (no differential verification bias)
N: No, the reference standard varied depending on the results of the index test (differential verification bias)
U: Uncertain

16. Were all patients included in the analysis?
Y: All patients were properly accounted for in the analysis
N: Some patients were not accounted for or dropped for unclear reasons
U: Uncertain

17. Patient Flow Risk of Bias: Could patient flow have introduced bias?
L: Low likelihood of bias based on absence of partial verification bias and good follow-up (“Y” on 14 and15, “Yes” or “Uncertain” for 16)
H: High likelihood of bias based on partial verification bias or poor follow-up (“No” to 14 or 15, or significant number of patients lost to follow-up in 16)
U: Uncertain

Overall risk of bias
Note: Overall risk of bias was low (L) if all domains were at low risk of bias; moderate (M) if one domain was at high risk of bias (including any “U”); and high (H) if two or more domains were at high risk of bias. Y = yes, N = no, U = uncertain.

Appendix B. Informed Consent Form

Name of Study: Study of Clinical Decision-Making About Patients with Suspected Infectious Mononucleosis

INFORMED CONSENT

This study is being conducted by Mark Ebell, MD, MS and Xinyan Cai, MPH from the College of Public Health at the University of Georgia in Athens, Georgia. The purpose of this study is to better understand the decision thresholds that clinicians use when making decisions about infectious mononucleosis. This information will be used to develop and improve decision-making regarding infectious mononucleosis.

STUDY DESCRIPTION

You will be asked to complete a brief survey that will take approximately 5 to 6 minutes. We will first ask for demographic information including years in practice, practice site and specialty. You will then be asked to evaluate the preferred management option in 7 scenarios of patients with sore throat.

CONFIDENTIALITY

We will not ask for your name or any personal identifiers. This research involves the transmission of data over the Internet. Every reasonable effort has been taken to ensure the effective use of available technology; however, confidentiality during online communication cannot be guaranteed. The information from the online survey will be destroyed after this study is completed and will not be used or distributed for future research.

CONTACT INFORMATION

If you have concerns or any further questions about the research, now or during the course of the study, please contact Mark H Ebell, MD, MS, Department of Epidemiology, University of Georgia (email: ebell@uga.edu, phone 706-247-4953). Additional questions or problems regarding your rights as a research participant should be addressed to the Chairperson, Institutional Review Board, University of Georgia (Telephone (706) 542-3199; E-mail Address IRB@uga.edu).

By responding to this survey, I acknowledge that I have read this information and agree to take part in this study titled "Understanding thresholds for diagnosis of infectious mononucleosis." I understand that my participation is voluntary. I can refuse to participate or stop taking part at any time without giving any reason, and without penalty. I can ask to have all of the information that I provide returned to me, removed from the research records, or destroyed.

Appendix C. Example of Online Survey

Please provide demographic and practice information:

Specialty: □ Family Medicine □Internal Medicine □ Physician Assistant
□ Nurse Practitioner □ Other specialty:_____

Years in practice: _____

Practice site: □ Primary care □ Urgent care □ Emergency Department
□ Other practice site: _____

Are you working at a student health clinic? □ Yes □ No

Instruction: Please answer all questions as if the patients were presenting to your office or clinic. For each scenario, first enter your estimated probability that the patient has infectious mono (from 0 to 100%). Then, choose the preferred diagnostic option (A or B).

Please assume that the "Monospot" test for infectious mononucleosis that your clinic uses is 80% sensitive in the first 7 days of infection and 95% sensitive after 7 days. This is the only test available to you for diagnosing mononucleosis. For each scenario that this is taking place in 2019 and COVID-19 is not a potential cause of the patient's symptoms.

Scenario 1: A 30-year old female comes to your office. She reports a ten-day history of sore throat accompanied by a skin rash. On examination, her tonsils are normal, and she has no posterior cervical adenopathy. She did not visit any doctor or order any test since her symptom onset. She denies shortness of breath, coughing, sore muscles or joints, sleeping too much, feeling nausea or feverish, headache, and is otherwise healthy with no serious comorbidities. Based on a validated clinical prediction rule, her probability of IM is about 1%.

□ A. You feel that the infectious mononucleosis is unlikely, and you will not order any test for mono.
□ B. You feel that more information is needed, and you will order a diagnostic test (for example: heterophile serology test ["Monospot"], viral capsid antigen, or EBV nucleic acid test) for infectious mononucleosis.

Scenario 2: A 30-year old female comes to your office. She reports a 6-day history of sore throat and coughing accompanied by a skin rash. On examination, her tonsils are normal, and she has no posterior cervical adenopathy. She did not visit any doctor or order any test since her symptom onset. She denies shortness of breath, sore muscles or joints, sleeping too much, feeling nausea, feverish or headache, and is otherwise healthy with no serious comorbidities. Based on a validated clinical prediction rule, her probability of IM is about 4%.

 A. You feel that the infectious mononucleosis is unlikely, and you will not order any test for mono.
 B. You feel that more information is needed, and you will order a diagnostic test (for example: heterophile serology test [“Monospot”], viral capsid antigen, or EBV nucleic acid test) for infectious mononucleosis.

Scenario 3: A 24-year old female comes to your office. She reports a 6-day history of sore throat, coughing, sore muscles and joints, accompanied by a skin rash. On examination, her tonsils are normal, and she has no posterior cervical adenopathy. Her spleen is in normal size. She came to see the doctor 3 days after the sore throat began and had a negative strep test at that time. However, her sore throat persists. She denies shortness of breath, sleeping too much, feeling nausea, feverish or headache, and is otherwise healthy with no serious comorbidities. Based on a validated clinical prediction rule, her probability of IM is about 7%.

 A. You feel that the infectious mononucleosis is unlikely, and you will not order any test for mono.
 B. You feel that more information is needed, and you will order a diagnostic test (for example: heterophile serology test [“Monospot”], viral capsid antigen, or EBV nucleic acid test) for infectious mononucleosis.

Scenario 4: An 18-year old female student comes to your office. She reports a ten-day history of sore throat and coughing. She started experiencing sore muscles and joints, feeling nausea, sleeping too much accompanied by a skin rash a week ago. On examination, her tonsils are normal, and she has no posterior cervical adenopathy. She came to see the doctor 3 days after the sore throat began and had a negative strep test at that time, but her symptoms have persisted. She denies shortness of breath, feverish or headache, and is otherwise healthy with no serious comorbidities. Based on a validated clinical prediction rule, her probability of IM is about 12%.

 A. You feel that the infectious mononucleosis is unlikely, and you will not order any test for mono.
 B. You feel that more information is needed, and you will order a diagnostic test (for example: heterophile serology test [“Monospot”], viral capsid antigen, or EBV nucleic acid test) for infectious mononucleosis.

Scenario 5: An 18-year old female student comes to your office. She reports a ten-day history of sore throat and coughing. She started experiencing sore muscles and joints, sleeping too much, feeling nausea, feverish and headache accompanied by a skin rash a week ago. On examination, her tonsils are normal, and she has no posterior cervical adenopathy. She came to see the doctor 3 days after the sore throat began and had a negative strep test at that time, but her symptoms have persisted. She denies shortness of breath and is otherwise healthy with no serious comorbidities. Based on a validated clinical prediction rule, her probability of IM is about 18%.

□ A. You feel that the infectious mononucleosis is unlikely, and you will not order any test for mono.
□ B. You feel that more information is needed, and you will order a diagnostic test (for example: heterophile serology test [“Monospot”], viral capsid antigen, or EBV nucleic acid test) for infectious mononucleosis.

Scenario 6: An 18-year old female student comes to your office. She reports a ten-day history of sore throat and coughing. She started experiencing sore muscles and joints, sleeping too much, feeling nausea, feverish and headache accompanied by a skin rash a week ago. On examination, her tonsils are enlarged with exudate, and she has no posterior cervical adenopathy. She came to see the doctor 3 days after the sore throat began and had a negative strep test at that time, but her symptoms have persisted. She denies shortness of breath and is otherwise healthy with no serious comorbidities. Based on a validated clinical prediction rule, her probability of IM is about 25%.

□ A. You feel that the infectious mononucleosis is unlikely, and you will not order any test for mono.
□ B. You feel that more information is needed, and you will order a diagnostic test (for example: heterophile serology test [“Monospot”], viral capsid antigen, or EBV nucleic acid test) for infectious mononucleosis.

Scenario 7: An 18-year old female student comes to your office. She reports a ten-day history of sore throat and coughing. She started experiencing sore muscles and joints, sleeping too much, feeling nausea, feverish and headache accompanied by a skin rash a week ago. On examination, her tonsils are enlarged with exudate, and presented with posterior cervical adenopathy. She came to see the doctor 3 days after the sore throat began and had a negative strep test at that time, but her symptoms have persisted. She denies shortness of breath and is otherwise healthy with no serious comorbidities. Based on a validated clinical prediction rule, her probability of IM is about 30%.

□ A. You feel that the infectious mononucleosis is unlikely, and you will not order any test for mono.
□ B. You feel that more information is needed, and you will order a diagnostic test (for example: heterophile serology test [“Monospot”], viral capsid antigen, or EBV nucleic acid test) for infectious mononucleosis.

Summary table of clinical vignettes

	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7
Age	30	30	24	18	18	18	18
Days of symptoms	10	6	6	10	10	10	10
Sore throat	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Rash	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cough		Yes	Yes	Yes	Yes	Yes	Yes
Sore muscles			Yes	Yes	Yes	Yes	Yes
Sore joints			Yes	Yes	Yes	Yes	Yes

Sleeping too much				Yes	Yes	Yes	Yes
Nausea				Yes	Yes	Yes	Yes
Fever					Yes	Yes	Yes
Headache					Yes	Yes	Yes
Enlarged tonsils with exudate						Yes	Yes
Posterior cervical adenopathy							Yes
Probability of infectious mono	1%	4%	7%	12%	18%	25%	30%

Appendix D. Summary Estimates of the Index Test for Each Individual Study [a]

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	LR+ (95% CI)	LR- (95% CI)	DOR
ELISA									
By serological Profile (Positive IgM and/or IgG, and negative EBNA IgG)									
All ages									
de Ory, 2011	69	2	8	40	0.90	0.95	18.82	0.11	
Koçoğlu, 2014	5	11	12	68	0.29	0.86	2.11	0.82	
Svahn, 1997									
Commercial kit 1	40	15	0	124	1.00	0.89	9.27	0.00	
Commercial kit 2	40	10	0	153	1.00	0.94	16.30	0.00	
Commercial kit 3	38	0	2	163	0.95	1.00	95.00	0.05	
Samlley, 1990	80	1	4	9	0.95	0.90	9.52	0.05	
Levin, 1991	76	16	1	205	0.99	0.93	13.63	0.01	
Devanthery, 2010	65	2	29	291	0.69	0.99	101.30	0.31	
Chan, 1998	22	3	1	81	0.96	0.96	26.78	0.05	
Fung, 2002	13	15	0	136	1.00	0.90	10.07	0.00	
Ho, 1989	341	4	12	1315	0.97	1.00	318.54	0.03	
Wiedbrauk, 1993									
Commercial kit 1	49	0	4	72	0.92	1.00	92.00	0.08	
Commercial kit 2	28	0	25	73	0.53	1.00	53.00	0.47	
Commercial kit 3	50	0	6	71	0.89	1.00	89.00	0.11	
Commercial kit 4	38	10	11	61	0.78	0.86	5.51	0.26	
Commercial kit 5	43	20	3	47	0.93	0.70	3.13	0.09	
Summary					0.90 (0.82, 0.94)	0.95 (0.91, 0.97)	19.90 (9.68, 37.20)	0.12 (0.06, 0.19)	192.0 (62.5, 462.0)
Children									
Kasifoglu, 2018	23	8	2	145	0.92	0.95	17.60	0.08	

Summary					0.92 (0.73, 0.98)	0.95 (0.90, 0.97)	18.60 (9.04, 35.80)	0.10 (0.02, 0.27)	302.0 (44.5, 990.0)
VCA IgM assay									
Altuglu, 2007									
Commercial kit 1	7	1	0	35	1.00	0.97	36.00	0.00	
Commercial kit 2	5	0	2	36	0.71	1.00	71.00	0.29	
Kasifoglu, 2018	20	7	4	147	0.83	0.95	18.33	0.17	
de Ory, 2011	49	2	24	46	0.67	0.96	16.11	0.34	
Koçoğlu, 2014	4	20	8	69	0.33	0.78	1.48	0.86	
Fung, 2002	30	15	1	125	0.97	0.89	9.03	0.04	
Summary					0.75 (0.53, 0.89)	0.92 (0.84, 0.96)	9.60 (4.14, 17.90)	0.29 (0.12, 0.53)	42.9 (8.3, 130.0)
VCA IgG assay									
Altuglu, 2007									
Commercial kit 1	32	2	5	6	0.86	0.75	3.46	0.18	
Commercial kit 2	28	0	9	8	0.76	1.00	76.00	0.24	
Kasifoglu, 2018	145	1	6	26	0.96	0.96	25.93	0.04	
de Ory, 2011	90	0	11	18	0.89	1.00	89.00	0.11	
Koçoğlu, 2014	75	6	3	17	0.96	0.74	3.69	0.05	
Fung, 2002	90	11	14	49	0.87	0.82	4.72	0.16	
Summary					0.89 (0.82, 0.94)	0.82 (0.74, 0.88)	4.92 (3.40, 7.20)	0.14 (0.08, 0.23)	39.5 (18.1, 77.7)
CLIA									
By serological Profile (Positive IgM and/or IgG, and negative EBNA IgG)									
de Ory, 2011									
Commercial kit 1	75	2	5	35	0.94	0.95	17.34	0.07	
Commercial kit 2	73	1	9	42	0.89	0.98	38.28	0.11	
Corrales, 2014	117	30	1	220	0.99	0.88	8.26	0.01	
Summary					0.95 (0.83, 0.98)	0.94 (0.86, 0.98)	17.80 (6.85, 39.80)	0.06 (0.01, 0.17)	367.0 (108.0, 967.0)
VCA IgM assay									

Al Sidairi, 2017	17	0	1	47	0.94	1.00	94.00	0.06	
de Ory, 2011									
Commercial kit 1	71	2	6	46	0.92	0.96	22.13	0.08	
Commercial kit 2	60	2	17	46	0.78	0.96	18.70	0.23	
Corrales, 2014	143	1	14	190	0.91	0.99	173.97	0.09	
Summary					0.88 (0.79, 0.93)	0.97 (0.93, 0.99)	38.90 (12.30, 94.80)	0.13 (0.07, 0.21)	356.0 (66.7, 1030.0)
VCA IgG assay									
Al Sidairi, 2017	35	0	1	29	0.97	1.00	97.00	0.03	
de Ory, 2011									
Commercial kit 1	29	28	3	65	0.91	0.70	3.01	0.13	
Commercial kit 2	30	15	2	78	0.94	0.84	5.81	0.07	
Corrales, 2014	257	14	2	88	0.99	0.86	7.23	0.01	
Lapierre, 2016	103	1	2	38	0.98	0.97	38.26	0.02	
Summary					0.96 (0.90, 0.99)	0.92 (0.85, 0.95)	12.20 (6.83, 20.40)	0.05 (0.02, 0.12)	324.0 (105.0, 728.0)
ICA									
By serological Profile (Positive IgM and/or IgG, and negative EBNA IgG)									
Koçoğlu, 2014	7	2	6	67	0.54	0.97	18.58	0.48	
Gomez, 2000									
All age group	46	1	52	251	0.47	1.00	118.29	0.53	
Children	14	1	32	63	0.30	0.98	19.48	0.71	
Summary					0.41 (0.30, 0.52)	0.98 (0.95, 0.99)	35.20 (7.67, 113.00)	0.60 (0.48, 0.71)	61.4 (11.3, 210.0)
VCA IgM assay									
Koçoğlu, 2014	9	10	3	79	0.75	0.89	6.68	0.28	
Summary					0.75 (0.45, 0.92)	0.89 (0.80, 0.94)	6.81 (3.14, 12.5)	0.30 (0.09, 0.62)	31.2 (5.49, 98)
VCA IgG assay									
Koçoğlu, 2014	66	4	12	19	0.85	0.83	4.87	0.19	
Summary					0.85 (0.75, 0.91)	0.83 (0.62, 0.93)	5.51 (2.19, 13.2)	0.2 (0.11, 0.33)	31.9 (7.7, 97.5)

IB									
By serological Profile (Positive IgM and/or IgG, and negative EBNA IgG)									
Kasifoglu, 2018	23	18	2	135	0.92	0.88	7.82	0.09	
Koçoğlu, 2014	10	1	7	68	0.59	0.99	40.59	0.42	
de Ory, 2014	66	13	4	34	0.94	0.72	3.41	0.08	
Summary					0.86 (0.56, 0.97)	0.91 (0.62, 0.98)	12.8 (2.59, 40.4)	0.18 (0.05, 0.45)	70.2 (25, 149)
VCA IgM assay									
Kasifoglu, 2018	20	15	4	139	0.83	0.90	8.56	0.18	
Koçoğlu, 2014	11	3	1	86	0.92	0.97	27.19	0.09	
Summary					0.83-0.92	0.90-0.97	8.56-27.19	0.09-0.18	46.33-315.33
VCA IgG assay									
Kasifoglu, 2018	150	6	1	21	0.99	0.78	4.47	0.01	
Koçoğlu, 2014	76	9	2	14	0.97	0.61	2.49	0.04	
De Ory, 2014	98	5	0	14	1.00	0.74	3.80	0.00	
Summary					0.99 (0.95, 1)	0.7 (0.56, 0.81)	3.4 (2.26, 5.32)	0.03 (0.006, 0.08)	205 (32.4, 704)
MFI									
By serological Profile (Positive IgM and/or IgG, and negative EBNA IgG)									
Devanthery, 2010	69	8	25	285	0.73	0.97	26.88	0.27	
Summary					0.73 (0.64, 0.81)	0.97 (0.95, 0.98)	28.50 (14.20, 53.40)	0.28 (0.20, 0.38)	107.0 (43.4, 221.0)
IMFA									
By serological Profile (Positive IgM and/or IgG, and negative EBNA IgG)									
de Ory, 2011	76	4	6	39	0.93	0.91	9.96	0.08	
Summary					0.93 (0.85, 0.97)	0.91 (0.78, 0.96)	11.40 (4.25, 25.50)	0.09 (0.04, 0.18)	160.0 (34.0, 489.0)
VCA IgM assay									
de Ory, 2011	57	0	20	48	0.74	1.00	74.00	0.26	
Summary					0.74 (0.63, 0.82)	0.99 (0.86, 1)	169 (5, 984)	0.28 (0.19, 0.39)	670 (15.1, 4030)

VCA IgG assay									
de Ory, 2011	85	0	22	18	0.79	1.00	79.00	0.21	
Summary					0.79 (0.71, 0.86)	0.97 (0.69, 0.99)	74.8 (2.56, 435)	0.23 (0.15, 0.33)	372 (8.5, 1890)
ELFA									
By serological Profile (Positive IgM and/or IgG, and negative EBNA IgG)									
Koçoğlu, 2014	7	1	9	75	0.44	0.99	33.25	0.57	58.33333333
De Ory, 2012	22	2	3	16	0.88	0.89	7.92	0.14	56.57142857
Summary					0.44-0.88	0.89-0.99	7.92-33.25	0.14-0.57	56.6-58.3
VCA IgM assay									
Koçoğlu, 2014	8	4	4	85	0.67	0.96	14.83	0.35	42.37142857
De Ory, 2012	22	3	6	19	0.79	0.86	5.76	0.25	23.04
Summary					0.67-0.79	0.86-0.96	5.76-14.83	0.25-0.35	23.0-42.4
VCA IgG assay									
Koçoğlu, 2014	70	3	8	20	0.90	0.87	6.88	0.12	57.33333333
De Ory, 2012	32	0	3	8	0.91	1.00	91.00	0.09	1011.111111
Summary					0.90-0.91	0.87-1.00	6.88-91.00	0.09-0.12	57-1011
Heterophile antibody test									
Slide agglutination test									
All ages									
Elgh, 1996									
Commercial kit 1		48	2	5	45	0.91	0.96	21.28	0.10
Commercial kit 2		45	0	8	47	0.85	1.00	85.00	0.15
Commercial kit 3		45	0	8	47	0.85	1.00	85.00	0.15
Svahn, 1997									
Commercial kit 1		31	0	9	163	0.78	1.00	78.00	0.23
Commercial kit 2		34	2	6	161	0.85	0.99	69.27	0.15
Tamaro, 2009		13	13	8	133	0.62	0.91	6.95	0.42
Fleisher, 1983		107	4	17	372	0.86	0.99	81.11	0.14
Chan, 1998		9	2	14	82	0.39	0.98	16.43	0.62

Gomez, 2000									
Commercial kit 1		57	27	41	225	0.58	0.89	5.43	0.47
Commercial kit 2		38	3	60	249	0.39	0.99	32.57	0.62
Linderholm, 1994									
Commercial kit 1		33	10	13	52	0.72	0.84	4.45	0.34
Commercial kit 2		35	5	11	57	0.76	0.92	9.43	0.26
Commercial kit 3		35	3	11	59	0.76	0.95	15.72	0.25
Commercial kit 4		39	9	7	53	0.85	0.85	5.84	0.18
Commercial kit 5		35	1	11	61	0.76	0.98	47.17	0.24
Summary					0.71 (0.61, 0.80)	0.96 (0.92, 0.98)	17.30 (8.69, 32.80)	0.31 (0.21, 0.41)	59.1 (23.8, 125.0)
Children									
Gomez, 2000									
Commercial kit 1		15	2	31	62	0.33	0.97	10.43	0.70
Commercial kit 2		7	0	39	64	0.15	1.00	15.00	0.85
Linderholm, 1994									
Commercial kit 1		3	1	5	11	0.38	0.92	4.50	0.68
Commercial kit 2		3	0	5	12	0.38	1.00	38.00	0.63
Commercial kit 3		4	0	4	12	0.50	1.00	50.00	0.50
Commercial kit 4		4	1	4	11	0.50	0.92	6.00	0.55
Commercial kit 5		3	0	5	12	0.38	1.00	38.00	0.63
Summary					0.35 (0.24, 0.47)	0.95 (0.89, 0.98)	7.39 (3.29, 14.60)	0.69 (0.57, 0.80)	11.0 (4.4, 22.4)
Solid-phase assay									
All ages									
Elgh, 1996									
Commercial kit 1		43	0	10	47	0.81	1.00	81.00	0.19
Commercial kit 2		37	1	16	45	0.70	0.98	32.11	0.31
Commercial kit 3		48	1	4	46	0.92	0.98	43.38	0.08
Linderholm, 1994									
Commercial kit 1		29	3	17	59	0.63	0.95	13.03	0.39

Commercial kit 2		30	0	16	62	0.65	1.00	65.00	0.35
Commercial kit 3		32	2	14	60	0.70	0.97	21.57	0.31
Commercial kit 4		33	3	13	59	0.72	0.95	14.83	0.30
Summary					0.67 (0.60, 0.73)	0.95 (0.92, 0.97)	15.00 (7.86, 25.80)	0.35 (0.28, 0.42)	44.1 (20.9, 80.7)
Children									
Linderholm, 1994									
Commercial kit 1		2	0	6	12	0.25	1.00	25.00	0.75
Commercial kit 2		2	0	6	12	0.25	1.00	25.00	0.75
Commercial kit 3		2	0	6	12	0.25	1.00	25.00	0.75
Commercial kit 4		4	1	4	11	0.50	0.92	6.00	0.55
					0.34 (0.20, 0.51)	0.93 (0.81, 0.98)	6.13 (1.80, 17.30)	0.71 (0.52, 0.88)	9.21 (2.07, 28.40)

[a] ***Note****: VCA: viral capsid antigen, EBNA: Epstein Barr nuclear antigen, IFA=Immunofluorescence assay, EIA= enzyme immunosorbent assay, ELISA: enzyme-linked immunosorbent assay, CLIA: chemiluminescent microparticle immunoassay, MFI: multiplex follow immunoassay, IB: immunoblot based assay, IMFA: immunofiltration assay, ICA: immunochromatographic assay, ELFA: enzyme-linked fluorescent assay*