

**THE EFFICACY OF HISTORY AND PHYSICAL EXAM FOR DIAGNOSIS**

**BY**

**GRAHAM MADDAFORD, B.SC., PA-S, MPAS (CANDIDATE)**

Supervisor: Dr. Aditya Sharma, MD, FRCPC

A capstone project submitted to the Faculty of Graduate Studies of The University of Manitoba

in partial fulfillment of the requirements for the degree of

**MASTER OF PHYSICIAN ASSISTANT STUDIES**

Physician Assistant Studies, University of Manitoba, Winnipeg

May 31, 2019

## TABLE OF CONTENTS

|  |           |
|--|-----------|
| <b>1. ABSTRACT</b>                                     | <b>3</b>  |
| <b>2. INTRODUCTION</b>                                 | <b>4</b>  |
| <b>2.1 BACKGROUND</b>                                  | <b>5</b>  |
| <b>2.1.1 PHYSICAL EXAM ORIGIN</b>                      | <b>5</b>  |
| <b>2.1.2 PHYSICAL EXAM MODERN</b>                      | <b>6</b>  |
| <b>2.1.3 INTEROBSERVER RELIABILITY</b>                 | <b>6</b>  |
| <b>2.1.4 SENSITIVITY</b>                               | <b>7</b>  |
| <b>2.1.5 SPECIFICITY</b>                               | <b>7</b>  |
| <b>2.1.6 LIKELIHOOD RATIOS</b>                         | <b>7</b>  |
| <b>2.1.7 LACHMAN TEST</b>                              | <b>8</b>  |
| <b>2.1.8 S3 HEART SOUND</b>                            | <b>9</b>  |
| <b>2.1.9 THE ALVARADO SCORE</b>                        | <b>10</b> |
| <b>3. METHODS</b>                                      | <b>12</b> |
| <b>4. RESULTS</b>                                      | <b>13</b> |
| <b>4.1 LACHMAN TEST</b>                                | <b>13</b> |
| <b>4.2 THIRD HEART SOUNDS</b>                          | <b>14</b> |
| <b>4.3 ALVARADO SCORE</b>                              | <b>15</b> |
| <b>4.3.1 ALVARADO SCORE VS COMPUTERIZED TOMOGRAPHY</b> | <b>15</b> |
| <b>4.3.2 ALVARADO SCORE VS ULTRASONOGRAPHY</b>         | <b>16</b> |
| <b>5. DISCUSSION</b>                                   | <b>17</b> |
| <b>5.1 LACHMAN TEST DISCUSSION</b>                     | <b>18</b> |
| <b>5.2 S3 HEART SOUND DISCUSSION</b>                   | <b>19</b> |
| <b>5.3 THE ALVARADO SCORE DISCUSSION</b>               | <b>20</b> |
| <b>5.4 LIMITATIONS</b>                                 | <b>22</b> |
| <b>5.5 FUTURE ASPIRATIONS</b>                          | <b>22</b> |
| <b>6. CONCLUSION</b>                                   | <b>23</b> |
| <b>7. ACKNOWLEDGMENTS</b>                              | <b>24</b> |
| <b>8. APPENDIX</b>                                     | <b>25</b> |
| <b>9. BIBLIOGRAPHY</b>                                 | <b>26</b> |

## FIGURES

FIGURE 1. COMPARISON OF NORMAL ACL ANATOMY TO ACL DISRUPTION

FIGURE 2. THE ALVARADO SCORE

## **1. Abstract**

### **Purpose**

This literature review looks to analyze the reliability of physical exams, along with signs and symptoms when determining the diagnosis.

### **Methods**

To answer this question a literature review was performed looking at the reliability of the Lachman test, the S3 heart sound, and the Alvarado score for aiding in determining the diagnosis

### **Results**

The Lachman test was good for ruling in ACL disruption with a specificity from (0.91- 0.97) and poorly for ruling out ACL disruption with a sensitivity ranging from (0.70-0.87) The presence of an S3 heart sound is a highly specific test for the presence of heart failure with a specificity ranging from 0.88(18) to 0.99(9) and a patient with an S3 there is an LR of 11 that the patient is experiencing heart failure. In terms of diagnostic accuracy, the Alvarado score cut of a point off 5 was excellent for ruling out appendicitis with sensitivities reaching 99% overall with each subgroup being 96% men, 99% woman, and 99% in children.

### **Conclusion**

It was not shown that physical exam findings along with clinical suspicion cannot replace advanced imaging testing and specific blood work testing. Physical exam findings for clinical suspicion should remain in standard practice when examining patients and should be considered before ordering multiple tests as physical exams can provide a lot of information before the testing results come back.

## 2.Introduction

The medical diagnosis is one of the many goals for any aspiring and seasoned medical professional. This is the process of determining the disease or condition by solely examining the patient's signs, symptoms, physical exam findings, and investigations. Historically this has been obtained by performing an extensive history and physical exam to determine the diagnosis.

Peterson et al observed that in 1992 that 86% of cases obtained the correct diagnosis by history and physical alone. (1) In modern medicine, physicians, physician assistants, and other medical professionals have a lot more in their toolbelts, including access to investigations, lab results and now modern technology including advanced imaging and genetic sequencing.

The issue with physical exam and signs is that they can be subjective with some tests lacking in sensitivity and specificity. Technology integration should be actively encouraged to keep physical exam up to date with modern medicine. The more recent trend with medicine is the interest in treating the lab values, rather than the actual patient. In practice there is a tendency to quickly review the labs from last night, look at the most recent nursing note. If there were no concerns with the patient overnight, there are many busy students who fail to examine their patients. (2) The trend has moved from listening and feel, to a more test and retest. Garibaldi et al lists a few factors that could have led to a decline in physical examination skills in recent years. One hypothesis is that in modern hospitals, graduate medical trainees spend as little as 12% of their time in direct contact with patients and their families. (2) This decreases the opportunities for practice and reduces the number of practitioners who are confident in their ability to teach these examination skills. There is also a question of the relevance of physical exam in the age of technology, but medical personnel fails to recognize that many maneuvers are

just as reliable as gold standard technology-based tests. (2) For example as looking at signs of heart failure such as a positive abdominal jugular test with a kappa of 0.92 compared to interstitial edema ( by chest radiograph) with a Kappa of 0.83 as described by Garibaldi et al. (2) (See table 1 in appendix)

In current medicine with new incoming diagnostic tests and investigations can be readily available in large hospitals. This poses a minute problem in a large hospital setting with large amounts of funding and resources available. The problem with relying on these to make the diagnosis is in underserved areas. A study published by Statistics Canada 2011 estimates a rural population of approximately 6.3 million Canadians(3). This accounts for a population that may not have the resources or access to expensive and high-tech diagnostics. Rural health care units (including northern nursing stations, small hospitals, and clinics) resources can be scarce therefore a good history and physical exam by the clinician of the area may be the only method of determining the patients' diagnosis.

Now the question arises, is this because the history and physical exam alone is not sensitive and specific enough to lead to diagnosis? Is there too much variability from physician to physician's ability to make the diagnosis?

This review aims to determine if physical exams are as specific and sensitive as gold standard imaging in leading to an accurate diagnosis

## **2.1 Background**

### **2.1.1 Physical exam origin**

The physical diagnosis had its origins in Grecian medicine. They took a careful history and practiced direct auscultation. They largely used observation, with careful documentation that

was shown to similar to that of the modern-day clinical interview. One of the most famous was Hippocrates (460-370 BC) lived during the golden age of Greece. His clinical cases that were collected not 1700 years later, demonstrated a high level of medicine including a careful history, inspection, palpation, direct auscultation with further examination of the sputum and urine. (4) This method of clinical examination takes many of the similar steps that are performed in the modern clinical interview.

### **2.1.2 Physical exam modern**

The modern clinical exam involves but not inclusive to take a focused history, including asking information about the timing of current symptoms, family history, previous medical history, current medications, allergies, examining the patient by inspection, auscultation, percussion, palpation along with obtaining a set of vital signs. (5) Clinical exams can include a basic set of blood work looking at electrolytes, creatinine, lipid profile, liver enzymes, and complete blood count look at organ function at a chemical level. Further tests are usually ordered once a differential diagnosis is established to formulate the final diagnosis. (6)

### **2.1.3 Interobserver reliability**

Reliability of physical exams is commonly measured by a calculated kappa score. A kappa score of 0 means that agreement between two observers happens by chance alone. A kappa of 1 indicates perfect agreement. The kappa can range from -1 to +1, but it has its limitations. In general, a kappa score greater than 0.4 is considered reasonable for a diagnostic test. (2) Many physical examination maneuvers have kappa scores between 0.4 and 0.75, which indicates intermediate to good reliability. Even after a diagnosis is made, the physical examination is important in following the disease's trajectory and severity. (2)

#### **2.1.4 Sensitivity**

The sensitivity of a test is defined as the proportion of people with the disease who will have a positive result. For example, a highly sensitive test is one that correctly identifies patients with a disease. If a test has a sensitivity of 100% it will identify all patients who have the disease.

(7) Tests that are highly sensitive can be useful for ruling out disease if a person has a negative result. For example, women have pap smears every three years once they are sexually active, a negative result on a pap smear, increases the likelihood that they do not have cervical cancer.

#### **2.1.5 Specificity**

The specificity of a test is the proportion of people without the disease who will have a negative result. This meaning that specificity of a test refers to how well a test identifies patients who do not have a disease. (7) A test with 100% sensitivity will identify every patient who does not have the disease. Tests with high specificity (a high true negative rate) are most useful when the result is positive. This can be useful to rule in a patient who has a certain disease.

#### **2.1.6 Likelihood Ratios**

Likelihood ratios are used to better understand how test performance affects clinical decisions. These offer an approach to assessing test performance that is unaffected by the rate of the condition being assessed in the population which also incorporates the found sensitivity and specificity into a more clinical useful value. (7) A positive likelihood ratio indicates the effect of a positive examination finding on the probability that the condition exists. While a negative likelihood ratio looks at the effect of a negative examination on the probability of that conditions

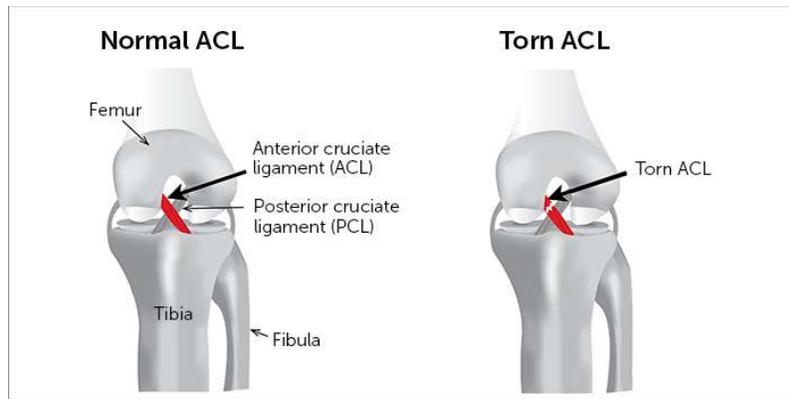
exists. If a result is negative for a diagnostic test, it is suggestive that the chance of the condition existing is very low. (7)

### **2.1.7 Lachman test**

The primary function of the Anterior cruciate ligament (ACL) is to control anterior translation of the tibia as well as Varus and valgus forces on the knee joint. The ACL originates from the posteromedial aspect of the lateral femoral condyle. The ACL is the most commonly injured knee ligament. In the United States, there is an annual incidence in the general population of approximately 1 in 3500. Around half of these requiring surgical reconstruction, the estimated cost for these injuries is almost a billion dollars per year. (8) ACL rupture can leave patients with knee joint instability during even during daily functional activities. Surgical intervention for the ACL rupture is the preferred treatment in patients that remain active and have presenting instability.

Thus, this leads to the importance to accurately being able to perform the most accurate clinical exam tests available to aid in the diagnosis of an ACL tear. This is to accurately allow medical professionals to screen for ACL injuries that require a further referral, advanced sets of imaging with possible surgical intervention as therapy. There are three clinical tests that are commonly used to look for ACL disruption including the pivot test, anterior drawer test, and Lachman test. This review looking primarily at the sensitivity and specificity of the Lachman Test for aiding in the diagnosis of ACL disruption.

**Figure 1. Comparison of Normal ACL Anatomy to ACL Disruption**



<https://www.google.com/search?q=acl&client=firefox-b->

[d&source=lnms&tbn=isch&sa=X&ved=0ahUKEwic88uo857iAhWD11QKHQXDB0Q\\_AUIDigB&biw=488&bih=452#imgrc=qf1799kalzrM](https://www.google.com/search?q=acl&client=firefox-b-d&source=lnms&tbn=isch&sa=X&ved=0ahUKEwic88uo857iAhWD11QKHQXDB0Q_AUIDigB&biw=488&bih=452#imgrc=qf1799kalzrM)

M:

The Lachman test is performed with the patient lying supine and with the involved extremity on the side of the examiner. The femur is stabilized with 1 hand, with the patient's knee joint in 20° to 30° of flexion. The examiner's other hand is applied to the posterior aspect of the proximal tibia. An anteriorly directed force is applied to displace the tibia. Increased anterior tibial translation with a soft endpoint compared to the other side constitutes a positive test, indicating disruption of the ACL.(8)

### **2.1.8 S3 Heart sound**

Heart failure is the pathophysiological state in which an abnormality of cardiac function is responsible for failure of the heart to pump blood at a rate required by metabolizing tissues, or to do so requires an elevated filling pressure to meet the demands. An alternative way to define

left sided heart failure is there is a decrease in left ventricular ejection fraction or an increased filling pressure. Left sided heart failure can be divided into one of three groups. A patient can have either decreased ejection fraction with normal filling pressure, decreased ejection fraction with increased filling pressure, or lastly normal ejection fraction with increased filling pressure. Ejection fraction can be measured with standard methods including echocardiograms, radionuclide angiography, Magnetic resonance imaging or cardiac catheterization. A decrease in ejection fraction shows there is an issue with systolic dysfunction. If the heart is failing to meet demands, it adapts by increasing left ventricular filling pressure to optimize cardiac output. Increased filling pressure also correlates with increased symptoms such as dyspnea and edema.

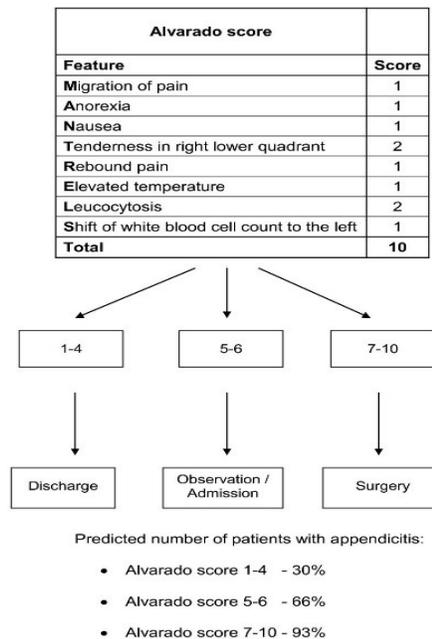
The left ventricular third heart sound or sometimes referred to as ventricular gallop occurs during ventricular vibration with rapid diastolic filling. This vibration may occur when either filling pressure is increased, or ventricular compliance is reduced. The third heart sound is low pitched and may be faint or intermittent. It should be sought with the bell of the stethoscope over the apical impulse. Listening while the patient is in the 45-degree left lateral decubitus position doubles the yield. The third heart sound may be confused with other diastolic sounds such as an opening snap, an abnormally split-second heart sound, or even a fourth heart sound if the patient is tachycardic. The third heart sound is, with rare exception, the only mid-diastolic sound. It occurs approximately 150 milliseconds (ms) after the second heart sound, or 5 times longer than the normal split of the second heart sound. (9)

### **2.1.9 Alvarado score**

Acute appendicitis is the most common cause of an acute abdomen requiring surgery, with a lifetime risk of about 7%.(10) Appendicitis can have symptoms that mimic over

conditions, leading to a certain challenge when coming up with the diagnosis. In 1986, Alvarado constructed a 10-point clinical scoring system, also known by the acronym MANTRELS, for the diagnosis of acute appendicitis as based on symptoms, signs and simple diagnostic tests in patients presenting with suspected acute appendicitis.(11) The Alvarado score looks at risk stratification in patients who are presenting with abdominal pain and linking the probability of appendicitis to recommendations including discharge, if the patient needs to be observed or if surgical intervention is indicated. This score can be used to aid in clinical judgment if further investigations including ultrasound or CT scans are needed. This clinical tool can be useful in areas of the world that have limited resources, a large time lag or lacking availability of imaging procedures.

**Figure 2. The Alvarado Score**



(12)

Figure originally used by Alvarado et al, 1986, A Practical Score for the Early Diagnosis of Acute Appendicitis

### 3.Methods

This paper included three small scale literature reviews for the Lachman test, S3 heart sounds, and Alvarado score. Several procedures were followed to ensure high-quality review of the literature on physical exam/scores specificity, sensitivity and interprofessional reliability. Three databases were searched including PubMed, Google Scholar, University of Manitoba Libraries. Secondly, the reference section for each article found was searched in order to find additional articles. Initially, a comprehensive search of peer-reviewed journals, but not including case reports, or peer reviews was completed based on a wide range of key terms and selection criteria are outlined in the table below

**Table 2. Selection criteria, key search terms, and number of studies meeting selection criteria**

|                    | Lachman test  | S3 heart sounds   | Alvarado score   |
|--------------------|---|---|--|
| Selection Criteria | <p>English transcript</p> <p>Published in the last 20 years</p> <p>Studies performed in developed countries</p> <p>Randomized control trial/ meta-analysis/ systematic review</p> | <p>English Transcript</p> <p>Published in the last 30 years</p> <p>Studies performed in developed countries</p> <p>Randomized control trial/ meta-analysis /systematic review</p>   | <p>English Transcript</p> <p>Published in the last 20 years</p> <p>Studies performed in developed countries</p> <p>Randomized control trial/ meta-analysis systematic review</p> |
| Key search terms   | <p>Clinical exam;</p> <p>Lachman’s test, inter-user reliability, Kappa, reliability, ACL tear diagnosis, ACL disruption; Arthroscopy; Gold standard; maneuver</p>                 | <p>Left ventricular failure; ventricular gallop; inter-user reliability; heart failure; sensitivity; specificity; hemodynamic, auscultation; S3; fluid overload; gold standard;</p> | <p>Acute appendicitis; Alvarado score; surgical appendicitis; inter-user reliability; sensitivity; specificity, validity, RLQ pain,</p>  |

|       |         |         |          |
|-------|---------|---------|----------|
| Yield | 7 of 47 | 6 of 25 | 14 of 40 |
|-------|---------|---------|----------|

It should be noted that a meta-analysis was not conducted because the focus of the studies, method of data analysis, subject area and types of physical exam finding varied considerably. The overall lack of assessment precision would make a meta-analysis essentially meaningless.

## 4. Results

### 4.1 Lachman Test

The literature was reviewed to analyze the diagnostic accuracy of the Lachman test for diagnosing ACL both partial or complete ruptures. Multiple sources were reviewed including pooled data from meta-analyses by Benjaminse et al(8), Jackson J et al(13), Mulligan et al(14) and Scholten et al(15), see table below.

**Table 3. Lachman Test sensitivity and specificity**

| Study              | Study type    | Number of patients | Sensitivity with CI | Specificity with CI |
|--------------------|---------------|--------------------|---------------------|---------------------|
| Benjaminse A et al | Meta analysis | 2276               | 0.85 (0.83-0.87)    | 0.94 (0.92-0.95)    |
| Jackson J et al    | Meta analysis | NA                 | .87 (0.76-0.98)     | 0.93 (0.89-0.96)    |
| Mulligan et al     | prospective   | 52                 | 0.70 (0.49-0.84)    | 0.97 (0.83-0.99)    |
| Scholten et al     | metanalysis   | 969                | 0.86 (0.76-0.92)    | 0.91 (76-0.92)      |

There have even been studies comparing Physical examination and MRI findings for diagnosing ACL tears. It was shown by Rose et al(16) that there was no significant difference between the accuracy of clinical examination and MRI in the diagnosis of ACL tears. Similar studies including a systematic review by Jackson J. et al(13) that clinical examination of ACL disruption has a sensitivity of 0.74 (CI 95% 0.60-0.88) and specificity )0.95 CI (0.92-0.98)

compared to MRI with a sensitivity of 0.87 (0.83- 0.91) and specificity of 0.91 ( 0.88-0.91). Interobserver reliability was only looked at in one study by Mulligan et al showing that the Lachman test had interrater reliability of k= 1.00 showing perfect reliability between the examiners.(14)

#### 4.2 Third heart sound

The literature was reviewed to analyze the diagnostic accuracy of the third heart sound (ventricular filling gallop or S3) for left ventricular dysfunction. Multiple sources were reviewed including pooled data from the meta-analysis by Simel et al(9), prospective trial by Harlen et al(17) and retrospective trial by Patel et al(18) shown in the table below

**Table 4. Third Heart Sound Accuracy for Left Ventricular Dysfunction**

| Study        | Study type        | Sensitivity                            | Specificity                            |
|--------------|-------------------|--|--|
| Simel et al  | Meta analysis     | 0.13                                   | 0.99                                   |
| Harlen et al | Prospective trial | 0.31                                   | 0.95                                   |
| Patel et al  | Retrospective     | 0.51 <sub>1</sub><br>0.78 <sub>2</sub> | 0.90 <sub>1</sub><br>0.88 <sub>2</sub> |

<sub>1</sub> abnormal systolic dysfunction where EF was less than 50%

<sub>2</sub> abnormal systolic dysfunction where EF was less than 30%

The presence of an S3 was highly predictive of abnormal ejection fraction, though the absence of an S3 is not uncommon in a patient with mild impaired EF. It was also shown by Minami et al the presence of an S3 was independently associated with an increase of in-hospital all-cause death [adjusted odds ratio(OR), 1.69; 95% confidence interval (CI), 1.19–2.41; p = 0.003] and cardiac death(adjusted OR, 1.66; 95% CI, 1.08–2.54; p = 0.020).(19) Among the congestive physical findings related to heart failure (S3, rales, jugular venous distension and peripheral edema) it was determined by Simel et al that the presence of a third heart sound (ventricular filling gallop) increased the likelihood of heart failure the most (LR, 11; 95% CI,

4.9-25)(9) Interobserver reliability of examination of a third heart sound was studied by Lok et al looking at reliability of detecting between Two cardiologists, one general internist, three senior and two junior postgraduate internal medicine trainees.(20) They found that detecting an S3 was  $K = 0.18$  (95% CI, 0.13 to 0.24) concluding there was no apparent trend in the accuracy or interobserver agreement with regard to the level of observer experience.(20)

### **4.3 Alvarado Score for Acute Appendicitis**

The literature review to analyze the diagnostic accuracy of using the Alvarado score for diagnosing acute appendicitis. A systematic review done by Ohle et al looked at 42 different studies looking at the diagnostic accuracy of the score at two cut off points score of 5 (1 to 4 vs. 5 to 10) and score of 7 (1 to 6 vs 7 to 10). The analysis focused on three subgroups men women and children.(11) It was found that in terms of diagnostic accuracy, the cut of point of 5 was excellent for ruling out appendicitis with sensitivities reaching 99% overall with each subgroups being 96% men, 99% woman, and 99% in children. When they looked at ruling in appendicitis and progressing to surgery using the cut-off point of 7, the score did not perform as well showing a specificity overall 81%, with each subgroup including 57% men, 73% in women and 76% in children. (11)

#### **4.3.1 Alvarado vs Computerized Tomography (CT)**

A prospective study by Ashraf et al looked to compare the performance of the Alvarado score and CT scan in the diagnosis of acute appendicitis.(21) Histopathology diagnosis was used as the gold standard against each of these methods for detection. The score of seven or more was considered positive and then received follow up imaging and treatment. The overall sensitivity and specificity for both genders was 85% and 60% respectively. In male's sensitivity were 87.5% and specificity of 84.1%.(21) while in females it was significantly lower with a sensitivity

of 79.7% and specificity of 38%.(21) The CT scan of the abdomen, however, performed more consistently with an overall sensitivity of 94.2% and specificity of 90%.(21) Performing equally well in both male and female patients. Thus showing that CT is more accurate for detecting acute appendicitis especially in female patients.

#### **4.3.2 Alvarado vs Ultrasonography (U/S)**

Another study by Nishikant et al 2015 compared the modified Alvarado score to ultrasonography in patients with acute appendicitis. This sensitivity and specificity of both were correlated with histopathological findings.(22) They used 7 as the cut off mark for the modified Alvarado score and found that it was shown to have a sensitivity of 98.44% and specificity of 94.4% for positive acute appendicitis with a false positive rate of 4.43%.(22) Modified Alvarado score <7 was observed in patients among (42.19%) were histopathological proved acute appendicitis and (57.81%) patients were histopathological negatives. Whereas Ultrasonography had a sensitivity of 98.33% and specificity of 90%.(22) Among the patients with a Modified Alvarado Score > 7, (96.2%) were ultrasonographically diagnosed as acute appendicitis and 6(3.8%) patients were ultrasonography negative.(22) Among these patients whose Modified Alvarado Score <7, (50.52%) were ultrasonographical diagnosed as acute appendicitis and (57.8%) patients were histopathologically negative. (22)

### **5. Discussion**

This literature review attempted to bring forward additional information on the diagnostic accuracy of physical examination skills. This study's aim was to identify certain physical exams or scores that can aid clinical decision making in areas that are resource scarce or do not have access to high-level imaging for diagnosis.

## 5.1 Lachman test

Upon review of the literature it was found that the Lachman test had overall specificity ranging from (0.91- 0.97) and sensitivity ranging from (0.70-0.87)(8)((13)(14) (15). Meaning that this test was good for ruling in an anterior cruciate ligament disruption but does not perform as well for ruling out ACL disruption. This shows that performing this test along with a good history can guide a clinician of ruling in an ACL disruption accurately.

We can speculate that a lot of clinicians today would rely on the MRI for further confirmation of this diagnosis. Things we have to consider is the wait times to receive an MRI, right now in Manitoba the average wait time for a nonemergent MRI is 16 weeks(23). This clinical test does not replace the accuracy of MRI as shown before ACL disruption detection by MRI has a sensitivity of 0.87 (0.83- 0.91) and specificity of 0.91 ( 0.88-0.91)(16), but can be considered before ordering an elective MRI to help reduce the expenses for low post-test yields.

Some of the limitations of these reported values is when looking at interobserver reliability, there was only one study that was included and met criteria. The report of kappa of 1 (perfect agreement) by Mulligan et al is concluded only from 2 separate physicians. Since this was a small-scale study, more data would be needed in order to determine a more accurate kappa value. Upon further analysis of the studies included in the meta-analysis by Benjaminse et al(8) and Jackson J et al(13), not all studies looked at specificity, so the reported value may be falsely higher or lower than the actual normal due to lack of data diversity to support it.

Overall the Lachman test has good sensitivity with high specificity and is a reliable test to perform when suspecting an ACL disruption and should be considered prior to order further investigations.

## **5.2 S3 heart sounds**

After reviewing the literature, it was found the presence of an S3 heart sound is a highly specific test for the presence of heart failure with a specificity ranging from 0.88(18) to 0.99(9). It was also shown by Simel et al that among the congestive physical findings related to heart failure (S3, rales, jugular venous distension and peripheral edema), an S3 had the highest LR of 11 that the patient is experiencing heart failure.(9) Thus, this exam can be used as a good predictor if is experiencing heart failure. It was also shown by Minami et al the presence of an S3 was independently associated with an increase of in-hospital all-cause death [adjusted odds ratio(OR), 1.69 and cardiac death(adjusted OR, 1.66)(19).

Early detection may lead to faster treatment while entering the hospital and can help decrease in cardiac-related death. This test, however, is not sensitive meaning that if a patient does not have this finding it does not mean that they are not in heart failure. Heart failure can be caused by many different etiology's such as ventricle wall abnormalities, increased filling pressure, and valvular disease. This exam finding can be used to help rule in future investigations such as echocardiograms, cardiology consultations, help decrease improper use of resources and decrease wait time for further investigations.(9) For example, last year in Manitoba thousands of patients who are considered candidates for "elective" tests are waiting anywhere from 10 to 15 months, with a median wait of 12 months. (24)

There are multiple arguments behind the reasoning of this, but one could also argue its that we as medical professionals are becoming less and less in tune with physical exam findings.

Another reason could be that with a large aging population the demand for increased due to numbers. The inter-user reliability of detecting an S3 has not been extensively studied, but was looked at by Lok et al 1998 showing between multiple professional at different levels of education detecting an S3 was no better than chance alone.(20)

One limitation about this study is that it was only performed in 1 hospital, to obtain a better idea multiple sites with multiple medical professional at different levels of studies would have to be evaluated to look at interobserver agreement. Some of the limitations when looking at physical exams for heart failure diagnosis is that this is a complex syndrome that can be caused by a multitude of ongoing issues. Looking for this physical finding may not help one figure out what is causing the heart dysfunction, but it can help clueing in what the current issue at hand is and if the present is a useful tool to rule in a patient that is experiencing heart failure. If an S3 is found on examination, treatment should be initiated immediately, with follow up investigations looking for the etiology of heart failure.

### **5.3 Alvarado score**

The Alvarado score as described is not a physical exam, it is a simple scoring system that incorporates part of the history, physical exam findings along with inexpensive investigations that small nursing is more likely to have access to. It was first found after reviewing the literature that using the Alvarado score with a cut off for the score of less than 5 for ruling out appendicitis across all groups with overall sensitivities reaching 99%, each group including men (96%), women (99%) and children (99%) performing as well or just slightly under.(11) This could be interpreted that this scoring system is a useful tool that can be applied when a patient presents to any health centre with the initial common complaint of abdominal pain. It should be

known that this specific chief complaint comprises of 5-10% of all reasons for ER visits.(25) As such this can be an effective tool to rule out appendicitis in patients, in which the physician can confidently state that it is likely not appendicitis and which further imaging is not needed at this time. For example, there are many screening tools that are used to guide for further imaging and management using the rule out a method that is used in common practice, including the Ottawa ankle rule or knee rule having sensitivities of 97.6% and 98.5% respectively.(11)

Now when using the score to rule appendicitis and progressing to surgery with a cut off of 7 or more, it does not perform as well showing an overall moderate specificity of 81%, thus indicating it is not a useful test for ruling in surgery. It does, however, increase clinical suspicion of acute appendicitis, warranting further imaging and surgical consultation for expert opinion on the subject. Comparatively, the CT scan was shown to be more consistent having overall sensitivities of 94.2% and specificity of 90%(21), especially when looking in female patients. It was also shown that there were higher false positives found in females using the Alvarado score using the greater than 7 to rule in method, than compared to the CT scan.

With regards to female patients, this makes sense as they have more anatomical issues that can occur in the right lower quadrant due to the presence of female reproductive organs with high amounts of associated diseases, most of which can mimic signs of acute appendicitis. Something we do have to consider is that the CT abdominal scans do cause radiation exposure to the body, which is something that medical professionals do try to avoid in females of reproductive age with concerns of future issues. Clinical suspicion assessing risk vs benefit in the case weighing the likelihood of appendicitis rather than scanning the abdomen just to make sure nothing is going wrong.

Something that was surprising is that when comparing the Alvarado score with a cut off of greater than 7 to Ultrasound done by Nishikant et al they found that these performed equal well in sensitivities and specificities for positive acute appendicitis(22). They also had similar rate of being histopathological negative after surgical intervention took place. Some of the limitations of looking at the Alvarado score is that most studies looked at were compared in an emergent setting, as it increases the clinical suspicion of acute appendicitis then when a patient comes into the walk-in clinic with abdominal pain.

#### **5.4 Limitations of this study**

There were a few limitations with this literature review including that it is, in fact, a literature review and not a compressive systematic review or meta-analysis. This literature looked at studies that were relevant to the topics provided and that were included in the inclusion criteria. The span of articles that would need to be covered in order to have to the completeness of a systematic review was not obtainable in the time permitted. It is also to be considered that not all articles reviewed were large scale collections of data. Some of the more specific questions including comparing a physical exam finding to a specific test and interobserver reliability were considered small scale studies. To truthfully examine these absolute values of these findings, there would need to be more reproducible data collected in order to have a reliable result that is closer to the real value.

#### **5.5 Future aspirations**

In a day in of modern medicine, there is a test for almost everything, but we must consider that even though in a “free” health care system to the public, the money from spending thousands of dollars on each patient for testing must come from somewhere. There have been

Canadian government initiatives including choose wisely Canada trying to minimize the number of tests that are not indicated, in efforts to decrease excessive health care spending. Looking into multiple physical exam findings can provide useful information to help reduce health care costs on the system. It can also be used to develop new scoring systems such as the Alvarado score to help guide clinical suspicion and guidance for further testing.

## **6. Conclusion**

To conclude it was shown that the Lachman test is a reliable test to perform when suspecting an ACL disruption and should be considered prior to order further investigations. The S3 heart sound is a specific finding of left ventricular heart dysfunction and should not be ignored due to the increase in hospital-related death. Finally, the Alvarado score is a simple scoring system that can be utilized to rule out suspecting acute appendicitis with scores of less than 5 which can help minimize unnecessary surgical laparotomies and supplementary diagnostic testing. It was shown that physical exam findings alone with clinical suspicion cannot replace advanced imaging testing and specific blood work testing. Utilizing physical exam findings for clinical suspicion should remain in standard practice when examining patients and should be considered before ordering multiple tests.

## **7.Acknowledgments**

I would like to thank my capstone mentor, Dr Aditya Sharma for his support and recommendations with this project. Also, would also like to give recognition to my fellow classmates of MPAS 2019 for their consistent support, good times and surviving this feat for the last two years. To Dr Deni Pirnat for his support and guidance throughout the MPAS program and this Capstone Project. To Becky Mueller, for her consistent guidance who believed and provided support when I needed it the most, to Ian Jones for advocating for the PA profession and of course to the rest of MPAS staff for always being there for us throughout this journey, and being a friendly face when needed most. I will never have enough thanks for my partner, my parents, family and friends for all of there love and support throughout the MPAS program and writing this capstone project.

## Appendix

Table. 1 from The Hypothesis-Driven Physical Exam as shown in Garibaldi et al 2018

| <b>Table 1</b><br>Interobserver reliability (kappa score) for common physical examination findings and common diagnostic standard tests |              |  |              |
|---|--------------|--|--------------|
| <b>Physical Finding</b>   | <b>Kappa</b> | <b>Diagnostic Standard</b>                                     | <b>Kappa</b> |
| Liver span >9 cm by percussion  | 0.11         | Classification of coronary artery lesions (by catheterization) | 0.33         |
| Delayed carotid upstroke  | 0.26         | Pulmonary infiltrate (by chest radiograph)                     | 0.38         |
| Diminished cardiac dullness   | 0.49         | Cardiomegaly (by chest radiograph)                             | 0.48         |
| Facial palsy (present or absent)  | 0.57         | Severity of valvular regurgitation (by echo)                   | 0.32–0.55    |
| Clubbing (Schamroth sign)   | 0.64         | Cirrhosis (by liver biopsy)                                    | 0.59         |
| Systolic hypertension (SBP >160 mm Hg)  | 0.75         | Calf DVT (by ultrasound)                                       | 0.69         |
| Tachycardia (pulse >100 bpm)  | 0.85         | Diagnosis of narrow complex tachycardia (by ECG)               | 0.70         |
| Abdominal jugular test  | 0.92         | Interstitial edema (by chest radiograph)                       | 0.83         |

## Bibliography

1. Peterson MC, Virginia W, Holbrook JH, Hales DEVON, Smith NLEE, Staker L V, et al. Articles Contributions of the History, Physical Examination, and Laboratory Investigation in Making Medical Diagnoses. 1992;163–5.
2. Garibaldi BT. The Hypothesis-Driven Physical Examination. *Med Clin NA* [Internet]. 2018;102(3):433–42. Available from: <https://doi.org/10.1016/j.mcna.2017.12.005>
3. Statcan. Census in Brief Canada ' s rural population since 1851 Population and dwelling counts , 2011 Census. 2011;(98):1.
4. Walker K, Hall D, Hurst W. *Clinical methods*. 3rd ed. Boston: Butterworths; 1990.
5. Artandi MK, Stewart RW. The Out patient Physical Examination. *Med Clin NA* [Internet]. 2018;102(3):465–73. Available from: <https://doi.org/10.1016/j.mcna.2017.12.008>
6. Zaman JAB. The Enduring Value of the Physical Examination. *Med Clin NA* [Internet]. 2018;102(3):417–23. Available from: <https://doi.org/10.1016/j.mcna.2017.12.003>
7. Denegar CR, Fraser M. How Useful Are Physical Examination Likelihood Ratios. *J Athl Train*. 2006;41(1):201–6.
8. Benjaminse A, Gokeler A. Clinical Diagnosis of an Anterior Cruciate Ligament Rupture : A Meta-analysis. *J Orthop Phys Sport Ther*. 2005;
9. Simel DL, Rennie D. Heart Failure. In: *The Rational Clinical Examination: Evidence-Based Clinical Diagnosis* [Internet]. McGraw-Hill Education - New York, NY - 2019/05/15; 2016. Available from: [jamaevidence.mhmedical.com/content.aspx?aid=1004223315](http://jamaevidence.mhmedical.com/content.aspx?aid=1004223315)
10. Ergul E. Importance of Family History and Genetics for the. *Internet J Surg*. 2006;10(1):4–7.
11. Ohle R, Reilly FO, Brien KKO, Fahey T, Dimitrov BD. The Alvarado score for predicting acute appendicitis : a systematic review. 2011;
12. Alvarado A, Hospital PG, Bennett H. A Practical Score for the Early Diagnosis of Acute Appendicitis. 1986;254(May).
13. Jackson JL, Malley PGO, Kroenke K. Evaluation of Acute Knee Pain in Primary Care. 2003;
14. Mulligan EP. Reliability and Diagnostic Accuracy of the Lachman Test Performed in a Prone Position. 2011;41(10):749–57.
15. Scholten R, Opstelten W, van der Plas C, Bijl D, Deville W, Bouter L. Accuracy of physical diagnostic tests for assessing ruptures of the anterior cruciate ligament: A meta-analysis. *Appl Evid*. 2003;52(September).
16. Rose N, Gold S. A Comparison of Accuracy Between Clinical Examination and Magnetic Resonance Imaging in the Diagnosis of Meniscal and Anterior Cruciate Ligament Tears. *Arthrosc J Arthroscopic Relat Sugery*. 1996;12(4):398–405.
17. Harlan WR, Oberman A, Grimm R, Carolina N. Chronic Congestive Heart Failure in Coronary Artery Disease : Clinical Criteria. *Ann Intern Med*. 1977;86(2):133–8.
18. Patel R, Bushnell DL, City I. Implications of an Audible Third Heart Sound in Evaluating Cardiac Function. *West J Med*. 1993;158(6):606–9.
19. Minami Y, Kajimoto K, Sato N, Aokage T, Mizuno M, Asai K, et al. Third heart sound in hospitalised patients with acute heart failure : insights from the ATTEND study. *Int J Clin*

- Pract. 2015;(August):820–8.
20. Lok CE, Morgan CD. The Accuracy and Interobserver Agreement in Detecting the ‘Gallop Sounds’ by Cardiac Auscultation. *Chest*. 1998;114(5):1283–8.
  21. Col L, Al-faouri AF, Khaled C, Ajarma Y, Abo-zaiton MOY. ScienceDirect The Alvarado score versus computed tomography in the diagnosis of acute appendicitis : A prospective study. *Med J Armed Forces India*. 2016;72:332–7.
  22. Gujar N, Mudhol S, Choudhari RK, Sachin DM. Determination of Sensitivity and Specificity of Modified Alvarado Score and Ultrasonography in Patients with Acute Appendicitis. *J Krishna Inst Med Sci Univ*. 2015;4(2):89–100.
  23. Government of Manitoba. MRI Scan Wait Times Information [Internet]. 2019 [cited 2019 May 20]. Available from: <https://www.gov.mb.ca/health/waittime/diagnostic/mri.html>
  24. Kusch L. WRHA moves to cut growing echocardiogram wait list. *Winnipeg Free Press* [Internet]. 2018 Aug 21; Available from: <https://www.winnipegfreepress.com/local/wrha-moves-to-cut-growing-echocardiogram-wait-list-491409871.html>
  25. Kendall J, Moreira M. Evaluation of the adult with abdominal pain in the Emergency Department [Internet]. *UptoDate*. 2019. Available from: <https://www.uptodate.com/contents/evaluation-of-the-adult-with-abdominal-pain-in-the-emergency-department#>