University of Calgary

PRISM Repository **https://prism.ucalgary.ca**

The Vault Open Theses and Dissertations

2018-06-27

A Psychological Perspective on Image Interpretation in Acute Ischemic Stroke: Factors Affecting Non-Contrast CT ASPECTS Reliability

Wilson, Alexis Terrin Connett

Wilson, A. T. C. (2018). A Psychological Perspective on Image Interpretation in Acute Ischemic Stroke: Factors Affecting Non-Contrast CT ASPECTS Reliability (Master's thesis, University of Calgary, Calgary, Canada). Retrieved from https://prism.ucalgary.ca. doi:10.11575/PRISM/32229 http://hdl.handle.net/1880/107007

Downloaded from PRISM Repository, University of Calgary

UNIVERSITY OF CALGARY

A Psychological Perspective on Image Interpretation in Acute Ischemic Stroke: Factors Affecting Non-Contrast CT ASPECTS Reliability

by

Alexis Terrin Connett Wilson

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

GRADUATE PROGRAM IN NEUROSCIENCE

CALGARY, ALBERTA

JUNE, 2018

© Alexis Terrin Connett Wilson 2018

ABSTRACT

The Alberta Stroke Program Early CT Score (ASPECTS) is a semiquantitative scale to assess the extent of early ischemic changes on non-contrast CT in acute ischemic stroke patients. This is crucial for prognostication and treatment selection. Recent studies have revealed significant heterogeneity in reported measures of inter-rater reliability in ASPECTS, and this thesis aims to investigate the reasons underlying this phenomenon from the perspective of clinicians' cognitive processes.

First, this work explores relevant topics in the psychology of image interpretation and, on this psychological basis, proposes potential causes of inconsistent ASPECTS reliability. Possible strategies to improve clinicians' inter- and intra-rater reliability are also discussed.

The effect of image reading context variables and rater expertise on ASPECTS inter-rater reliability was then investigated. Raters of different experience levels scored ASPECTS on baseline non-contrast CT scans under three prior-information conditions (NCCT only, NCCT with access to clinical information, NCCT with access to clinical information and multiphase CT angiography) and three reading-context conditions (high/low ambient light, time pressure). The results indicate that these variables have the capacity to affect ASPECTS reliability.

This work highlights the importance of acknowledging that medical image interpretation can be influenced by seemingly irrelevant external and internal factors like reading environment characteristics or physician-level variables. Giving more consideration to these variables in clinical and educational settings could improve the utility of tools like ASPECTS.

PREFACE

Chapter 2 of this thesis has been published as: Wilson AT, Dey S, Evans JW, Najm M, Qiu W, and Menon BK. Minds treating brains: Understanding the interpretation of non-contrast CT ASPECTS in acute ischemic stroke. *Expert Review of Cardiovascular Therapy* 2018;16(2):143-153.

ACKNOWLEDGMENTS

Above all, I must express my wholehearted appreciation to my supervisor, Dr. Bijoy Menon. From agreeing to take me on as a graduate student to going above and beyond to help me pursue my career goals, you have been an invaluable mentor and adviser. Thank you for steadfastly encouraging and supporting my personal and professional growth over the past two years.

I would also like to extend my gratitude to the members of my supervisory committee: Dr. Michael Hill, Dr. Andrew Demchuk, and Dr. Gustavo Saposnik. From your tireless work, I have learned so much about the practice of medicine, the principles of scientific research, and the ways that they intersect. Dr. Hill, your willingness to take me on as a summer research student initiated my academic journey. Dr. Demchuk, your attitude of inquiry has taught me to always seek a profound understanding of the effects I observe. Dr. Saposnik, your generosity in including me in projects and in sharing your expertise has enhanced my learning so much.
I am also very thankful to Dr. Sonny Chan for dedicating the time and effort required to act as my internal examiner.

To my labmates and collaborators, Dr. Wu Qiu, Dr. Hulin Kuang, Dr. Ting-Yim Lee, Dr. Sadanand Dey, Dr. James Evans, Dr. Mohammed Almekhlafi, Jessalyn Holodinsky, Dr. Noreen Kamal, Kevin Chung: I am sincerely grateful for your willingness to share your knowledge with me in the form of academic contributions, feedback, and teaching. Thank you, also, for your camaraderie and encouragement.

To my fellow graduate student, Moiz Hafeez: thank you so much for your excellent advice, and for the many car rides. I look forward to being your classmate again next year.

To my colleague and good friend, Mohamed Najm: your unwavering readiness to provide a helping hand or to lend an ear has meant so much to me. Thank you for teaching me, encouraging me, and supporting me.

Finally, I would like to acknowledge the role that my family has played throughout my graduate work; their endless moral and emotional support was instrumental in the completion of this thesis. Mom and Dad – thank you for always being there. Supriya and Mayank – you have been so helpful at every step along the way. Dhruv – you have stood behind me always. I could not have done this without all of you.

TABLE OF CONTENTS

Abstract ii
Preface iii
Acknowledgments iv
Table of Contents vi
List of Tables viii
List of Figures ix
List of Abbreviations & Symbols x
CHAPTER ONE: INTRODUCTION & BACKGROUND 2
1.1. Background 2
1.1.1. Ischemic Stroke Pathology 2
1.1.2. Treatment of Ischemic Stroke 3
1.1.2.1. Thrombolysis 3
1.1.2.2. Endovascular Thrombectomy 4
1.1.3. Imaging in Hyperacute Stroke Care 4
1.1.3.1. Non-Contrast CT 5
1.1.3.2. CT Angiography 8
1.1.3.3. CT Perfusion 8
1.2. Research Objectives & General Themes 9
1.3. Thesis Structure 9
1.4. Contribution of Authors 10
CHAPTER TWO: AN OVERVIEW OF MEDICAL IMAGE INTERPRETATION AND ASPECTS 11
2.1. Overview of ASPECTS 13
2.1.1. Rationale & Purpose 13
2.1.2. Reliability of ASPECTS 14
2.1.2.1. Technical Factors 16
2.1.2.2. Patient Factors 17
2.1.2.3. Reader Factors 18
2.2. Overview of Visual Processing 19
2.2.1. Perception is Selective 19
2.2.2. Perception can be Biased 20
2.3. ASPECTS Reading and Visual Processing 22
2.3.1. Human Visual Search Strategies Affecting ASPECTS Reading 22
2.3.2. Varying Reading Context Affects ASPECTS Reading 23
2.4. Interventions to Optimize ASPECTS Reliability 24
2.4.1. Top-Down Effects 25

2.4.1.1. Task 25
2.4.1.2. Motivation 26
2.4.1.3. Background Knowledge and Clinical Information 26
2.4.2. Bottom-Up Effects 27
2.4.2.1. Improving Display Quality and Learning Windowing Techniques 27
2.4.2.2. Optimizing Post Processing of NCCT Scans 28
2.5. Training 29
2.5.1. Expertise 29
2.5.2. Training Techniques 30
2.6. Conclusion 33
2.7. Expert Commentary 33
2.8. Five-Year View 34
CHAPTER THREE: THE EFFECT OF IMAGE READING CONTEXT FACTORS ON NON-CONTRAST CT ASPECTS RELIABILITY 36
3.1. Introduction 36
3.2. Methods 38
3.2.1. Statistical Analysis 39
3.3. Results 39
3.4. Discussion 50
3.4.1. Summary of Results 50
3.4.2. Exploration of Cognitive Explanations for Observed Effects 51
3.4.3. Limitations 53
3.4.4. Conclusions 53
CHAPTER FOUR: FUTURE DIRECTIONS 55
4.1. Summary 55
4.1.1. Limitations 56
4.2. Future Directions 57
4.3. Conclusion 58
References 59
Appendix A: Reporting Inter-Rater Reliability 66
Appendix B: Copyright Permissions 67

LIST OF TABLES

Table 2.1. Factors that may contribute to variability in ASPECTS scoring 16

Table 2.2. Summary of interventions suggested to improve ASPECTS reliability across individual reading contexts 25

Table 3.1. Baseline demographic characteristics of the patients selected from the PRove-IT database 41

Table 3.2. Inter-rater reliability estimates for total ASPECTS between all three raters 43

Table 3.3. Median image interpretation times (seconds per NCCT scan) for the non-*Time Pressure* subgroups 44

Table 3.4. Inter-rater reliability estimates for trichotomized ASPECTS (0-4, 5-7, 8-10) between all three raters 46

Table 3.5. Intraclass correlation coefficient estimates for all three raters, stratified by baseline patient and imaging characteristics 48

Table 3.6. Intraclass correlation coefficient estimates for ASPECTS regionwise agreement between all three raters 49

Table 3.7. Intraclass correlation coefficient estimates for each rater's agreement with CT perfusion-ASPECTS 50

LIST OF FIGURES

Figure 2.1. The 10 ASPECTS regions of the middle cerebral artery territory at the ganglionic and supraganglionic levels 14

Figure 2.2. Leukoaraiosis (white matter disease), brain atrophy, and motion artifact are patient-derived sources of variability in ASPECTS reading 18

Figure 2.3. Altering the window settings can affect the appearance of early ischemic changes and thus contribute to variability in ASPECTS reading 24

Figure 2.4. Post processing techniques and enhancement algorithms of CT scans contribute to variability in ASPECTS scoring 28

Figure 2.5. Qualitative trichotomization of ASPECTS (good, fair, poor) reflects the clinical application of ASPECTS 31

Figure 3.1. Flowchart illustrating the proposed cognitive framework underlying potential causes of variability between readers in medical image interpretation 36

Figure 3.2. Bland-Altman plots depicting the agreement between each pair of raters for each of the three prior information conditions 42

LIST OF ABBREVIATIONS & SYMBOLS

ASPECTS	Alberta Stroke Program Early CT Score
ATP	Adenosine Triphosphate
CBF/CBV	Cerebral Blood Flow/Cerebral Blood Volume
CI	Confidence Interval
CT	Computed Tomography
CTA	Computed Tomography Angiography
CTP	Computed Tomography Perfusion
DWI	Diffusion-Weighted Imaging
ECASS	European Cooperative Acute Stroke Study
ECG	Electrocardiogram
EIC	Early Ischemic Changes
ESCAPE	Endovascular treatment for Small Core and Anterior circulation Proximal occlusion with Emphasis on minimizing CT to recanalization times
EVT	Endovascular Thrombectomy
FDA	United States Food and Drug Administration
FLAIR	Fluid-Attenuated Inversion Recovery
HERMES	Highly Effective Reperfusion evaluated in Multiple Endovascular Stroke
ICA	Internal Carotid Artery
ICC	Intraclass Correlation Coefficient
IQR	Interquartile Range
IRR	Inter-Rater Reliability
κ	Kappa Statistic
κ_W	Weighted Kappa Statistic
MCA	Middle Cerebral Artery
MCA-M1/M2	Middle Cerebral Artery - M1 or M2 Segment
mCTA	Multiphase Computed Tomography Angiography
MIP	Maximum Intensity Projection
MR	Magnetic Resonance (Imaging)
mRS	Modified Rankin Scale
MTT	Mean Transit Time [s]
NCCT	Non-Contrast Computed Tomography
NIHSS	National Institutes of Health Stroke Scale
NINDS	National Institute of Neurological Disorders and Stroke
PRove-IT	Precise and Rapid assessment of collaterals using multi-phase CTA in the triage of patients with acute ischemic stroke for IA Therapy
rtPA	Recombinant Tissue Plasminogen Activator
TMax	Time to Maximum [s]
tPA	Tissue Plasminogen Activator
TTP	Time to Peak [s]
WW/WL	Window Width/Window Level

Variability is the law of life, and as no two faces are the same, so no two bodies are alike, and no two individuals react alike and behave alike […].

Sir William Osler, On the Educational Value of the Medical Society

CHAPTER ONE: BACKGROUND AND INTRODUCTION

1.1. Background

1.1.1. Ischemic Stroke Pathophysiology

Stroke is a prevalent and devastating condition; it is a leading cause of death and long-term disability worldwide. Ischemic stroke, which refers to hypoperfusion of a brain region due to cerebral artery occlusion, accounts for approximately 80% of stroke cases.[1] The region affected by ischemia is comprised of two zones: the ischemic core is tissue with very low perfusion that is unsalvageable, and the penumbra is tissue with moderately low perfusion that still maintains its structural integrity and which could be salvaged if perfusion were restored in a timely manner.[2]

The damage incurred by brain tissue in ischemic stroke is caused by a localized reduction in cerebral blood flow, leading to cellular hypoxia and the resultant ischemic cascade, where anaerobic metabolism and ATP depletion cause lactic acidosis and the failure of ATP-dependent ion pumps.[3] This disruption of ionic homeostasis leads to increased concentration of sodium and chloride ions within neurons and, subsequently, cytotoxic edema, where water accumulates intracellularly. Another component of this cascade involves the extracellular accumulation of water, or ionic edema, due to the osmotic gradient generated by sodium efflux into the extracellular space.[4] In vasogenic edema, tight junctions between endothelial cells of the blood-brain barrier lose integrity, causing intracellular components to leak from newly fenestrated capillaries.[5] Excitotoxicity due to excess neuronal glutamate release and uptake promotes cellular calcium influx; this contributes to intracellular degradation of proteins and membranes. Moreover, hypoxic cells produce reactive oxygen species, which further damage neurons.[1]

From a clinical perspective, these cellular processes manifest in the acute stage as early ischemic changes (EICs). Imaging markers for EICs include parenchymal hypoattenuation, reduced grey-white matter differentiation, focal swelling (sulcal effacement), and mass effect. The latter two signs are often excluded from EIC assessment, as they may be associated with penumbra rather than core.[6] In the clinical setting, it is crucial to assess EICs in stroke patients

because the extent of these changes is associated with benefit from therapy and may predict functional outcomes and hemorrhage risk.[7–9] Thus, EIC assessment is key for treatment selection and prognosis in acute ischemic stroke.

1.1.2. Treatment of Ischemic Stroke

"Time is brain" is a ubiquitous aphorism in the acute stroke literature. This statement expresses the importance of restoring perfusion as quickly as possible, because every additional minute of hypoxia results in irreversible loss of brain tissue. Specifically, 1.9 million neurons may be lost for every minute that a typical large vessel ischemic stroke goes untreated, and the brain's aging could be accelerated by 3.6 years for every such hour.[10]

1.1.2.1. Thrombolysis

Thrombolytic drugs constitute an established standard of care in acute ischemic stroke; they function by breaking down thrombi contributing to cerebral ischemia. Tissue plasminogen activator (tPA) is an endogenous fibrinolytic protein that activates fibrin-bound plasminogen on the surface of thrombi. When activated, plasminogen is converted to plasmin, a protease that lyses fibrin in the thrombus, thereby dissolving it.[11] Recombinant tPA (rtPA, or alteplase) is presently the only thrombolytic drug approved by the FDA for treatment of acute ischemic stroke.[12] Nearly twenty-five years ago, the National Institute of Neurological Disorders and Stroke (NINDS) rt-PA Stroke Study demonstrated the safety and efficacy of intravenous rtPA within three hours of ischemic stroke onset: relative to placebo, patients treated with rtPA were 30% more likely to have no or minor disability after three months and 55% more likely to achieve a final NIH Stroke Scale (NIHSS) score of 0 or 1.[13] As a result of subsequent trials,[14] the current American Heart Association/American Stroke Association guidelines recommend 4.5 hours from time last seen normal as the upper limit of the rtPA time window.[15]

This stringent time window excludes a large number of patients from receiving intravenous thrombolysis. Furthermore, patients with more severe strokes[16], large vessel occlusions[17], or longer thrombus length[18] experience less benefit from rtPA.[19]

As an alternative to intravenous administration, thrombolytic therapy can be administered locally into the cerebral circulation (intra-arterial thrombolysis).[20] However, the only thrombolytic drug that has been empirically demonstrated to provide benefit when administered intra-arterially (urokinase/prourokinase) is not approved by the FDA, and alteplase has not been subject to a randomized controlled trial for use in intra-arterial thrombolysis.[21] The current American Heart Association/American Stroke Association guidelines recommend thrombectomy using stent-retrievers (discussed below) over intra-arterial thrombolysis as first-line therapy.[15]

1.1.2.2. Endovascular Thrombectomy

Five landmark randomized controlled trials published in 2015 established the role of endovascular thrombectomy (EVT) in acute ischemic stroke patients with occlusion of the proximal anterior artery circulation.[22–26] In this procedure, a catheter is guided into the cerebral vasculature from a puncture at the groin or neck. Then, one of a number of thrombectomy devices (stent-retrievers presently being the foremost) is deployed in the artery to capture and retrieve the thrombus. In a patient-level pooled meta-analysis of these five studies from the Highly Effective Reperfusion evaluated in Multiple Endovascular Stroke Trials (HERMES) collaboration group, it was determined that EVT in addition to best medical therapy was beneficial across many patient subgroups to a significant extent. The adjusted odds ratio for modified Rankin Scale (mRS) score reduction at 90 days relative to best medical management was 2.49.[27] Rates of 90-day mortality, parenchymal hematoma, and symptomatic intracranial hemorrhage did not differ significantly between the control and treatment arms.

1.1.3. Imaging in Hyperacute Stroke Care

Ischemic stroke is a dynamic pathology, and the condition of ischemic brain tissue is constantly evolving prior to reperfusion. Thus, effective brain imaging protocols must 1) be rapid and readily available, to provide up-to-the-minute information, and 2) provide information that meaningfully contributes to the decision-making processes of treatment selection and prognostication. This information includes the presence or absence of intracranial hemorrhage,

the extent of the infarct core and penumbra, vessel status, and identification of any intracranial thrombi.[28,29]

1.1.3.1. Non-Contrast CT

A non-contrast computed tomography (NCCT) scan consists of two-dimensional images resulting from numerous x-ray measurements. Images can be acquired using either a sequential ("stop-and-shoot") or a helical ("spiral") technique, which may have implications with regards to image quality, brain structure visualization, and grey-white matter differentiation.[30] Denser objects, such as bone or calcification, appear brighter than less dense objects like cerebral parenchyma, cerebrospinal fluid, or water. Due to edema associated with infarction, infarcted tissue progressively becomes more hypodense; conversely, blood is denser than brain parenchyma, so hemorrhage appears hyperdense.[31]

NCCT is the fastest and most widely accessible acute brain imaging modality, and it is generally the first imaging obtained for stroke patients.[32] It can reliably distinguish normal and ischemic tissue from hemorrhage, which is a key step in ischemic stroke care.[33] The presence of a hyperdense vessel sign on NCCT has been associated with more severe strokes and poorer three-month outcomes.[17] Conjugate eye deviation, a shift in horizontal gaze that is a reliable indicator of the affected hemisphere, is another sign that can be appreciated using this imaging modality.[34]

EICs in the middle cerebral artery (MCA) territory can also be assessed on NCCT. From a physiological perspective, cerebral ischemia causes increased water content in brain tissue, which is visualized by hypoattenuation on NCCT. An animal study using a rat model of MCA occlusion found an inversely linear relationship between tissue water content and x-ray attenuation, with a decrease of 1.8 Hounsfield units corresponding to a 1% increase in water content.[35] Severe hypoattenuation is likely associated with irreversible tissue damage; the fate of tissue demonstrating subtle attenuation changes is still an open question.[36] As time from ischemia onset increases, salvageable penumbral tissue will be converted into unsalvageable infarct core.

Thus, early recanalization is favourable for increasing the likelihood of good patient outcomes, and the extent of EIC on NCCT can be a key piece of information in prognostication.

Other early ischemic signs on NCCT include cortical swelling and sulcal effacement. However, if these changes are not associated with hypoattenuation, they may reflect reversible tissue changes related to collateral vessel vasodilation.[4,37–39]

Prior to the development of the Alberta Stroke Program Early CT Score (ASPECTS) in the year 2000, EICs were assessed qualitatively by estimating the percentage of MCA territory where CT signs of ischemia are present. The European Cooperative Acute Stroke Study (ECASS) and ECASS-II, which assessed the safety and efficacy of intravenous alteplase, excluded patients with CT hypodensity in more than 33% of the MCA territory; this became known as the 1/3 MCA Rule.[8,40] The ECASS investigators recognized the importance of a systematic method to evaluate EICs in acute stroke treatment, as interventions are much less likely to produce good outcomes in patients with large infarct cores.[41,42] However, subsequent studies have found that achieving a high degree of agreement with the 1/3 MCA Rule can be problematic, even among experienced clinicians.[43,44] ASPECTS was conceived to address this obstacle by encouraging systematic, stepwise assessment of baseline NCCT scans.[9] It is a ten-point score typically assessed on axial NCCT images; a lower score indicates greater extent of EIC. There are ten prespecified ASPECTS regions in the MCA territory of the affected side: six cortical regions (M1-M6), plus the insula, caudate nucleus, lentiform nucleus, and internal capsule. One point is subtracted from the initial score of ten for each region where signs of EICs are present.

ASPECTS is a widely-used clinical tool. Its prognostic value has been demonstrated in a number of studies: in the original ASPECTS study, for instance, dichotomized ASPECTS (0-7, 8-10) was effective in discriminating patients who achieved independent functional outcomes.[9,37] Subsequent studies, such as an analysis from the Canadian Alteplase for Stroke Effectiveness Study (CASES), have found a graded relationship between baseline ASPECTS and 90-day functional outcome, particularly for ASPECTS > 5.[45] However, NCCT-ASPECTS has not been

shown to have a treatment-modifying effect for intravenous thrombolysis, and patients therefore should not be excluded from this treatment on the basis of ASPECTS alone.[7,46] Following EVT, patients with baseline NCCT-ASPECTS ≤ 7 experienced significantly poorer functional outcomes than those with ASPECTS > 7. Patients who were treated early (<5 hours onset-to-recanalization) and with favourable ASPECTS >7 had the best outcomes, but patients with ASPECTS 5-7 also experienced benefit from early recanalization. If recanalization was achieved in a later time stage, patients with ASPECTS > 7 were more likely to have a good clinical outcome than those with ASPECTS ≤ 7.[47]

In several of the recent EVT trials, ASPECTS was used as a patient exclusion parameter.[23,24,26] For instance, in the Endovascular Treatment for Small Core and Anterior Circulation Proximal Occlusion with Emphasis on Minimizing CT to Recanalization Times (ESCAPE) trial, potential participants with ASPECTS less than 6 were excluded, because this corresponds to a moderate-to-large infarct core. Thus, the evidence for treatment benefit of EVT in patients with low ASPECTS is weak. A number of ongoing trials, including TENSION (ClinicalTrials.gov identifier NCT03094715) and IN EXTREMIS, seek to elucidate the degree to which EVT benefits low-ASPECTS patients.

Although ASPECTS is clinically relevant, pragmatic, and easy to implement, it presents certain challenges. A recent systematic review enumerated thirty studies that have reported measures of inter-rater reliability for ASPECTS; the authors found that results were highly heterogeneous, with kappa (κ) values for total ASPECTS ranging from 0.26 to 0.97, and intraclass correlation (ICC) values ranging from 0.57 to 0.83.[48] The study methods were also heterogeneous, with discrepancies in variables including (but not limited to) rater population, rater training or experience level, specific elements of ASPECTS methodology, environmental or ambient reading conditions, and display settings (window/level).

1.1.3.2. CT Angiography

Collateral vessels are minor vessels in the vicinity of the occluded artery; if a patient has good collateral status, their brain tissue is likely to be sustained for a longer period of time relative to a patient with poor collaterals due to compensatory perfusion. There is some suggestion that patients' differential extents of EIC can be at least partially attributed to differences in collateralization.[49]

CT angiography (CTA) is a CT scan acquired concurrently with intravenous injection of a contrast medium, permitting visualization of vessel lumens in the cerebral arterial tree. This allows for occlusion detection and assessment of collateral circulation, as well as identification of vascular features such as stenosis.[28] Traditionally, single-phase CTA has been performed; however, this technique is limited in temporal resolution, so there is little capacity for collateral grading. Thus, multiphase CTA (mCTA) has been developed, where multiple (typically two) skull base-to-vertex scans are performed in addition to the initial scan following contrast material injection.[50] Features that are crucial to collateral grading, like quality of pial artery filling, can be more easily appreciated by this method.[51]

1.1.3.3. CT Perfusion

Like CTA, CT perfusion (CTP) requires intravenous injection of a contrast agent. This imaging modality involves acquisition at multiple time points, generating a time-attenuation curve for each voxel as the contrast agent is temporally traced through the vasculature. Post-processing techniques produce colour maps based on various parameters, including cerebral blood volume (CBV), cerebral blood flow (CBF), and time to peak/mean transit time (TTP/MTT).[52] These parameters can be correlated with tissue characteristics consistent with penumbra and core, providing information regarding the extent of infarction.[53]

1.2. Research Objectives & General Themes

In clinical practice, physicians make numerous significant decisions each day, taking into consideration ambiguous information and substantial potential risks. For instance, in acute ischemic stroke, the decision to treat a patient with endovascular thrombectomy or thrombolysis is not always straightforward, as innumerable variables must be carefully weighed. Despite this, the investigation of stroke physicians' decision-making processes from a cognitive perspective is a relatively unexplored topic.

One component of physician decision-making is medical image interpretation. Image interpretation is a complex cognitive skill where visual search and sophisticated judgments must be coordinated.

This research aims to address previously described issues in NCCT-ASPECTS inter-rater reliability by, first, exploring plausible relationships between ASPECTS scoring on NCCT and principles of the cognitive psychology of visual perception, and then experimentally assessing the effect of reading-context variables on the inter-rater reliability of NCCT-ASPECTS scoring by raters of different experience levels. Taken together, the findings will provide valuable insight into physicians' cognitive processes underlying medical image interpretation in acute stroke care and potentially identify targets for improving the reliability of ASPECTS scoring.

1.3. Thesis Structure

This thesis consists of one published manuscript and one original study. Chapter Two is a narrative review discussing concepts in cognitive psychology that are relevant to ASPECTS interpretation in acute ischemic stroke. In this paper, potential sources of inter-rater variability in ASPECTS and proposed strategies to mitigate these effects are discussed. It was published in *Expert Review of Cardiovascular Therapy*.[54] Chapter Three describes original research investigating the effects of reading context and background knowledge conditions on ASPECTS inter-rater reliability in raters of different levels of expertise.

1.4. Contribution of Authors

Wilson AT, Dey S, Evans JW, Najm M, Qiu W, Menon BK. Minds treating brains: Understanding the interpretation of non-contrast CT ASPECTS in acute ischemic stroke. *Expert Review of Cardiovascular Therapy* 2018;16(2):143-153. doi: 10.1080/14779072.2018.1421069

ATW, SD, and BKM conceived of this narrative review. ATW collected information and wrote the manuscript. SD, JWE, MN, WQ, and BKM provided feedback and edited the manuscript. ATW assumes responsibility for the integrity of the review.

CHAPTER TWO: AN OVERVIEW OF MEDICAL IMAGE INTERPRETATION AND ASPECTS

Minds treating brains: Understanding the interpretation of non-contrast CT ASPECTS in acute ischemic stroke (published in *Expert Review of Cardiovascular Therapy*)

Wilson AT, Dey S, Evans JW, Najm M, Qiu W, Menon BK

Affiliations:

Wilson AT – Department of Clinical Neurosciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

Dey S – Department of Clinical Neurosciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

Evans JW – Department of Clinical Neurosciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

Najm M – Department of Clinical Neurosciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

Qiu W – Department of Clinical Neurosciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

Menon BK – Departments of Clinical Neurosciences, Radiology, Community Health Sciences; Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

Keywords: Stroke, Computed tomography, Medical imaging, Brain imaging, Image interpretation

Word count: 5266

Tables: Table 2.1: Factors that may contribute to variability in ASPECTS scoring.

Table 2.2: Summary of interventions suggested to improve ASPECTS reliability across individual reading contexts.

Abstract

Introduction: The Alberta Stroke Program Early CT Score on non-contrast CT is a key component in prognostication and treatment selection in acute stroke care. Previous findings show that the reliability of this scale must be improved to maximize its clinical utility.

Areas Covered: This review discusses technical, patient-level, and reader-level sources of variability in ASPECTS reading; relevant concepts in the psychology of medical image perception; and potential interventions likely to improve inter- and intra-rater reliability.

Expert Commentary: Approaching variability in medical decision making from a psychological perspective will afford cognitively informed insights into the development of interventions and training techniques aimed at improving this issue.

2.1. Overview of ASPECTS

2.1.1. Rationale & Purpose

In acute ischemic stroke, the assessment of early ischemic changes (EIC) on non-contrast computed tomography (NCCT) imaging is instrumental in treatment selection, as evidence suggests that it predicts functional outcomes and the risk of intracranial hemorrhage.[7]

EIC were previously quantified using the 1/3 Middle Cerebral Artery (MCA) rule, which was used in the European Cooperative Acute Stroke Study (ECASS) to predict benefit from thrombolysis. By this method, patients were excluded if more than 33% of the MCA territory was affected by EIC.[8] However, subsequent studies have found that achieving a high degree of agreement using the 1/3 MCA rule can be problematic, even among experienced clinicians.[43]

The Alberta Stroke Program Early CT Score (ASPECTS) was developed in 2000 to serve as an alternative to the 1/3 MCA rule in evaluating EIC.[9] It is a semiquantitative scale involving assessment of 10 regions in the MCA territory: M1-M6 (cortex), caudate, lentiform nucleus, insula, and internal capsule. These regions are evaluated at the ganglionic and supraganglionic levels (Figure 2.1). For each region in which parenchymal hypoattenuation, loss of grey-white differentiation, or sulcal effacement is observed, one point is subtracted from 10; thus, the nearer ASPECTS is to 0, the greater the extent of EIC. It is important to note that this methodology is not completely standardized. The ASPECTS regions are imprecisely delineated, and it is not specified to what extent the region must be affected by EIC in order to warrant subtracting a point. The original ASPECTS study used only two NCCT slices to assign scores, but current methods overwhelmingly use the whole scan.[9,37] Another source of variation is the characteristics that are considered evidence of EIC: for example, due to recent pathophysiological research, isolated cortical swelling is not considered a sign of EIC in many studies evaluating ASPECTS.[37] An additional criticism of ASPECTS is that some regions – for instance, the internal capsule – are much smaller than others, yet they are equally weighted in the total score; thus, two patients with the same ASPECTS score may not have the same extent of EIC.[42] In this review, we discuss

reasons for variability in ASPECTS reading, including a detailed exploration of visual perception and resultant cognitive biases that likely affect ASPECTS interpretation. We then discuss strategies with the potential to help physicians improve their ability to read and interpret ASPECTS.

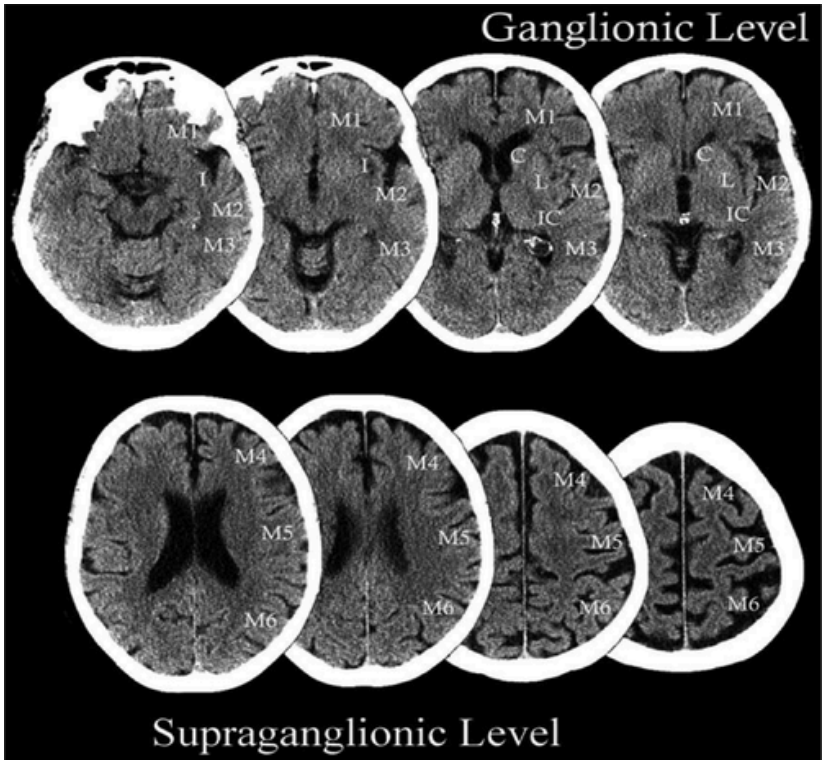

Figure 2.1. The 10 ASPECTS regions of the middle cerebral artery territory at the ganglionic and supraganglionic levels. Note that the cortical regions are not clearly delineated.
M1-M6: Cortical MCA regions; I: insula; L: lentiform nucleus; C: caudate; IC: internal capsule.

2.1.2. Reliability of ASPECTS

Detecting EICs on NCCT is not easy, especially when patients present early after ischemic stroke onset. NCCT has a low signal-to-noise ratio in EIC detection, unlike magnetic resonance diffusion weighted imaging (MR DWI). This, along with a lack of standardization of reporting parameters, has raised concerns about the reliability of this method in assessing extent of EIC in patients with acute ischemic stroke.

There have been a modest number of studies investigating the reliability of ASPECTS scoring on NCCT. In a systematic review, Farzin et al. [48] include 30 such studies, each using between 2 and 5 readers (most readers being expert neurologists or neuro-radiologists). A striking finding from this review is that the study methodologies differ from each other on several characteristics, including if the readers were provided with clinical information when reading the scans or not, the time assigned to read a scan, readers' access to all CT slices, readers' ability to set their own window settings and the inclusion of ASPECTS training as part of the study. (We discuss these variables in greater detail below.) The study populations are also heterogeneous. Perhaps as a result, the findings from this review on the current state of EIC ASPECTS reliability reflect a wide degree of variability in inter-rater reliability (IRR; measured by Kappa statistics and correlation coefficients). For instance, unweighted kappas from the studies included in this review ranged from 0.26 and 0.97 for total ASPECTS, and from 0.16 to 0.93 for dichotomized ASPECTS.[48]

In addition to the ambiguities of the ASPECTS methodology mentioned above, there are a number of sources of variation that could introduce heterogeneity into the process of scoring NCCT scans. An overview of these is provided in Table 2.1. These factors are not necessarily specific to ASPECTS: they have the potential to influence any form of medical image scoring or interpretation.

Lack of Methodological Standardization
Inclusion of cortical swelling
Extent of early ischemic changes in a region
Number of slices to include when scoring
Technical Factors
Scan generation parameters
Scan vendor
Slice thickness
Scan quality
Motion artifact
Display quality
Patient Factors
Age
Presence of old infarcts
Presence of brain atrophy
Presence of leukoaraiosis/white matter disease
Stroke severity (NIH Stroke Scale)
Affected hemisphere
Time from stroke onset to NCCT imaging
Reader Factors
Experience/Expertise
Training
Personality factors (Ambiguity aversion, risk aversion)
Geography; Health jurisdiction
Reading context
Lighting
Time of day
Time pressure
Stress
Fatigue
Task structure
Window/level settings

Table 2.1. Factors that may contribute to variability in ASPECTS scoring.

2.1.2.1. Technical Factors

Not all NCCT scans are created equal; there are several technical variables that could affect readers' scores by introducing perceptual discrepancies. These may include scan generation

parameters, scan quality, scan parameters such as peak X-ray energy (keV/meV), scan vendor, and image processing and display procedures.[55]

2.1.2.2. Patient Factors

The ASPECTS reliability studies discussed above used varied patient populations.[48] Median NIH Stroke Scale (NIHSS, stroke severity at presentation) ranged from 4 to 19. Some studies exclusively used patients eligible for thrombolysis, and others included those eligible for endovascular thrombectomy. Factors such as patient age, presence of old infarcts, brain atrophy, and leukoaraiosis (white matter disease) could influence ASPECTS scoring. Patient motion introduces further limitations in image interpretation (Figure 2.2). Stroke onset-to-CT time also likely contributes to variability; one study found that agreement between readers for ASPECTS EIC assessment was significantly lower in scans acquired 0-90 minutes from stroke symptom onset when compared to scans acquired at subsequent time periods (91-180, 181-360, >360 min).[56]

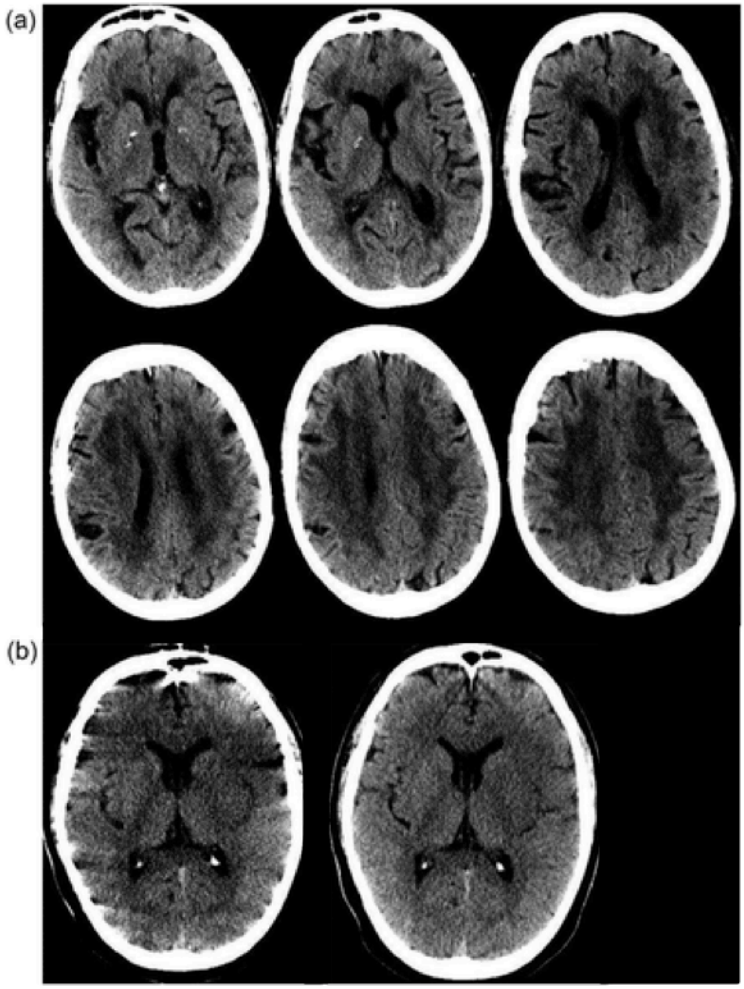

Figure 2.2. Leukoaraiosis (white matter disease), brain atrophy, and motion artifact are patient-derived sources of variability in ASPECTS reading. Leukoaraiosis and brain atrophy can affect image quality and the appearance of early ischemic changes (a). Motion artifact (b) can obscure true early ischemic changes; in this pair of scans, the caudate and M1 ASPECTS regions appear affected in the presence of motion artifact (left), but a better scan (right) reveals that these two regions are spared.

2.1.2.3. Reader Factors

Additional sources of inter-rater variability could be related to individual readers and reader populations. One's level of experience, training, medical specialty, and personality factors such as ambiguity aversion or risk aversion can come into play. The majority of studies that

tested reliability used experienced stroke neurologists and/or neuro-radiologists; only a few used more novice readers, like residents or fellows.[57–60] However, even within expert readers, discrepancies in ASPECTS training may exist across different geographic areas or healthcare jurisdictions.[61] Other important but unaddressed reasons for ASPECTS variability include individual contextual elements like task structure, context/situation of the task, reading environment, time pressure, and time of day.

2.2. **Overview of Visual Processing**

2.2.1. Perception is Selective

When we use our senses to experience the world, it can seem as though our perception represents every detail. However, human cognitive resources are finite, and the world is simply too detail-rich for human cognition to represent each aspect of it simultaneously. Work in psychology reveals that our conscious visual phenomenology (visual perception) is selective; certain facets of reality jump out at us or fade into the background, based on one's current task, problem, or cognitive processes. For example, the amount of information taken in by our retinas far surpasses the information processing limitations inherent to our brains; indeed, perceiving *all* this information would overwhelm cognitive function. Also, other brain areas extensively process data entering the primary visual cortex before our conscious experience of "seeing" is realized. As a result, our visual perception is generated through complex interactions between visual inputs and higher cognitive functions; it is not the case that our visual experience is presenting an "objective reality."[62,63] This is the distinction between sensation, which pertains to direct sensory information, and perception, which is a dynamical process between the brain and the world.

It is theorized that perception can be influenced by two broad categories of factors: bottom-up and top-down. Bottom-up or data-driven processes involve the stimulus properties of incoming external information – in the case of visual processing, retinal inputs. Top-down, conceptually-driven processes are derived from higher brain areas; these types of elements can

include the task one is engaged in and mental attitudes, such as one's motivation or expectations.[64]

Task structure can also have a significant effect on what one perceives. For instance, Clark et al. [65] sought to investigate the effect of two different task conditions on the accuracy of visual search. Participants had to search a screen for one or two targets distributed among distractors; selecting correct targets would accumulate points. The Fixed Duration cohort was instructed to collect the most points possible in a given amount of time, while the Fixed Objective cohort was told to collect a certain number of points as quickly as possible. The results showed a significant difference in error rates between the two cohorts; the Fixed Objective group's accuracy was decreased when there were multiple targets present. In other words, the Fixed Duration group was more effective at finding subsequent targets in dual-target trials. In this experiment, the same optimal strategy (maximizing search efficiency) applies to both conditions, so it is interesting that a discrepancy in cognitive performance was observed. The authors suggest that the different task instructions caused participants to conceptualize the task differently; for instance, the Fixed Objective (time pressure) task may have induced a sense of stress or anxiety relative to the Fixed Duration task. In this sense, one's implicit concept of the task at hand is a top-down factor that can impair perceptual performance.

2.2.2. Perception can be Biased

In the 1970s, researchers in psychology began describing a seemingly worrying trend: human judgment was frequently found to be at odds with what rational choice theory would deem 'objective rationality'. This finding was robust and replicable. A widely-cited example involves a problem now known as the "Linda Problem".[66] Participants are given a profile of a person and are asked to judge which of two alternatives is more likely. For example:

Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.

Which alternative is more probable?

Linda is a bank teller. (Option 1)
Linda is a bank teller and is active in the feminist movement. (Option 2) [66]

Option 2 is a conjunction; it is the probability that Linda is a bank teller AND that she is a feminist. Thus, it is necessary that Option 1 is equally or more likely than Option 2. However, in one experiment, 85% of university student participants selected Option 2, an ostensibly irrational choice. Daniel Kahneman and Amos Tversky were seminal players in proposing cognitive biases and heuristics as an explanation for these apparent failures of rationality. By their theory, heuristics are cognitive rules of thumb or shortcuts that are often sufficient for us to make appropriate judgments. These shortcuts are cognitively economical, given the limited processing power of the human brain. One heuristic employed when estimating the distance of an object is to use the object's visual clarity as a proxy for nearness. While it is generally true that farther objects are less clear, this is not necessarily always the case. If a heuristic is used in a situation where the shortcut rule does not apply, systematic errors in judgment called biases can result.[67] Visual perceptual experiences and judgments based upon these experiences can therefore be biased.[68]

2.3. ASPECTS Reading and Visual Processing

2.3.1. Human Visual Search Strategies Affecting ASPECTS Reading

As Krupinski [55] outlines in her review of medical image perception, there are several challenges specific to radiological image interpretation from a psychological perspective. A fundamental cognitive difficulty with NCCT scans is that readers must generate a three-dimensional mental representation of anatomy and lesions using two-dimensional slices. Although the introduction of helical head CT protocols has mitigated this issue to some extent, this type of cognitive challenge still exists in ASPECTS, which is typically assessed using axial slices. Also, in contrast to some other visual search tasks, reading NCCT scans in acute stroke can involve multiple targets in a single image; moreover, in acute ischemic stroke care, the question is often not, "Is a lesion present or absent?" (detection) but, instead, "Is a lesion present, and what is the extent of the lesion?" (detection and interpretation). These multifactorial objectives add complexity to an already-challenging undertaking.

"Satisfaction of search" is a phenomenon in radiology whereby, in multiple-target scans, readers miss subsequent findings after positively identifying an initial target. This can be precipitated by many contextual variables, including stress and, as demonstrated in the experiment discussed above in which time pressure decreased search performance, task structure.[65,69] The methodology of ASPECTS was designed to avoid this issue by requiring sequential, region-by-region assessment of the MCA territory, reducing the likelihood of search termination after an initial finding of EIC. However, evidence suggests that satisfaction of search does not exclusively arise from premature termination of search; eye-tracking experiments have shown many instances where image readers have fixated on targets but failed to report corresponding findings.[69] For instance, a radiologist may look at a lesion on a lung X-ray, but stress, fatigue, or other circumstances could cause them to not 'register' this lesion. Therefore, 'forcing' ASPECTS raters to consider specific regions of the MCA territory may in itself be insufficient to avoid missed findings due to satisfaction of search.

Synthesis of eye-tracking evidence suggests that the general strategy for medical image interpretation first involves generating a broad 'gist' of the image, followed by more detailed search in relevant areas.[70] This has significant implications for potential training opportunities, discussed below. Matsumoto et al. [71] produced saliency maps of NCCT scans from stroke patients, which involves using a computational program to predict image regions that are more visually salient in a bottom-up manner (e.g. regions of high contrast). Using eye-tracking, they found that stroke neurologists fixated on salient regions for the same duration as non-neurologist controls, but neurologists also fixated on additional clinically relevant areas that controls ignored. Thus, it seems that experienced ASPECTS readers are able to use additional clinical and imaging information to focus better on salient regions when compared to readers with less experience.

2.3.2. Varying Reading Context Affects ASPECTS Reading

Only a small number of studies have investigated the effect of changing reading context on ASPECTS reliability. Despite this, there is clear evidence that contextual variables can affect ASPECTS reading. An optimized window setting (Figure 2.3) improved NCCT-ASPECTS IRR compared to standard window settings, and optimized-window NCCT-ASPECTS more closely reflected DWI or FLAIR MR-ASPECTS.[58] Some studies have compared treating stroke neurologists' real-time ASPECTS scores to expert neuro-radiologists' scores assigned at a later review. Zerna et al. [72] found fair ($\kappa = 0.51$) agreement between these groups, with real-time neurologist scores equally underestimating and overestimating the radiologists' scores. Puetz et al. [73] had similar results ($\kappa_W = 0.62$), and real-time neurologists tended to score higher ASPECTS than the reviewing radiologists. Coutts et al. [74] found substantial agreement between readers in real-time settings ($\kappa_W = 0.69$). In all of these investigations, the expert reviewers were blinded to clinical information except affected side, whereas the real-time stroke neurologists had knowledge of all clinical information.

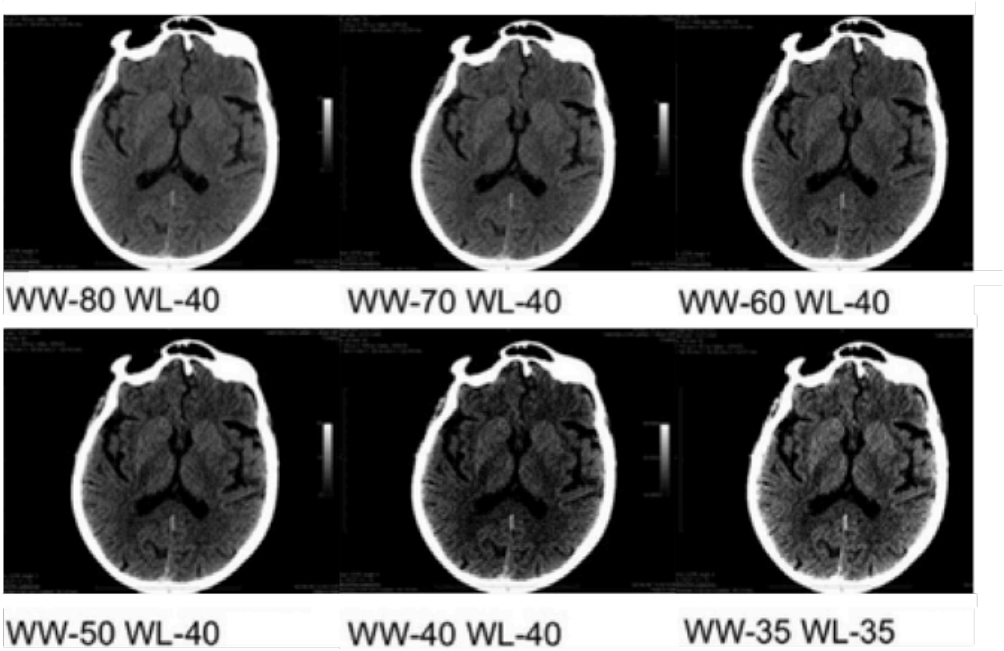

Figure 2.3. Altering the window settings can affect the appearance of early ischemic changes and thus contribute to variability in ASPECTS reading. WW, window width; WL, window level.

2.4. Interventions to Optimize ASPECTS Reliability

By combining findings of clinicians' inter-rater reliability and theories of cognitive psychology, we can propose several paths of action that could improve the reliability of ASPECTS scoring. There is the potential to standardize ASPECTS reading procedures based on both top-down, or conceptually-driven, and bottom-up, or stimulus-driven, factors influencing visual perception. The interventions discussed below are summarized in Table 2.2.

Top-Down Interventions
Considering the task structure
Setting a time limit
Managing one's motivation
Fatigue
Time of day when patient presents
Accessing clinical information
Using additional imaging to 'check' ASPECTS
Bottom-Up Interventions
Higher bit-depth displays
Choosing one's own window settings
Post-processing: Maximal Intensity Projections on NCCT
Color enhancement of grey/white matter
Training & Expertise
Experts have greater speed and accuracy
Generating a gestalt impression first
Perceptual and conceptual training
Error recovery training

Table 2.2. Summary of interventions suggested to improve ASPECTS reliability across individual reading contexts.

2.4.1. Top-Down Effects

2.4.1.1. Task

Although clinicians have not explicitly been assigned a 'task' like participants in psychological experiments, their performance could vary based on how they implicitly or explicitly frame the process. For instance, Clark et al's data [65] suggest that giving oneself a predetermined time limit (as opposed to trying to finish as quickly as possible) could reduce the likelihood of missing multiple targets. Time pressure is counterintuitively not always detrimental to performance in medical image interpretation. A low to moderate degree of time pressure has been demonstrated to not affect accuracy relative to no time pressure, and it could be that time pressure encourages non-analytical processing and discourages 'overthinking', or the consideration of suboptimal cues.[75,76] In other words, time pressure may facilitate a processing

style that increases performance in certain tasks, but further investigation is needed to determine if this applies to ASPECTS reading. In preliminary results, our group has shown that ASPECTS reading is more reliable when readers are provided less than a minute to read the scan, a task similar to Clark's fixed duration task.[65,77]

2.4.1.2. Motivation

One outcome of perception being selective and biasable is that we see what we hope to see. For instance, Balcetis and Dunning [78] performed a series of experiments demonstrating that participants shown ambiguous stimuli (such as a figure that could be the letter B or the number 13) were more likely to interpret the stimuli in the manner that provided a more desirable outcome. The authors posit that the different interpretations of perceptual stimuli are like hypotheses, and top-down processes such as motivation can bias a person to favouring one hypothesis over others. Motivation is a complex concept and can be modified by diverse factors ranging from one's long-term goals to one's present state of hunger. Some factors that could particularly affect the motivation of stroke physicians reading ASPECTS may include fatigue or eye strain [55] and time of day when the patient presents; this latter variable has many treatment implications based on how long it may take to assemble the team, which could unconsciously affect the reader's perception of the severity of EIC.

2.4.1.3. Background Knowledge & Clinical Information

In the context of ASPECTS, clinical information (e.g. affected hemisphere, specific deficits, stroke severity) and additional information from other imaging modalities can provide a great deal of background information while interpreting the NCCT. It is possible that this background information could increase ASPECTS reading accuracy because it prespecifies where to search for EIC; however, this information could also mislead readers and cause them to miss or misinterpret findings. Some results suggest that providing clinical information (age, sex, stroke

severity, affected side) to readers generally improves total ASPECTS inter-rater reliability, but additional CT angiography (CTA) did not confer any additional benefit.[77] Thus, clinical details may be the most beneficial background information to access when scoring ASPECTS – indeed, additional imaging is not likely to be available in the earliest stages of patient assessment. It may be more advisable to use CTA and other subsequent imaging as a post-hoc verification of the ASPECTS score, rather than as a component of initial ASPECTS scoring.

2.4.2. Bottom-Up Effects

Krupinski [55] reviews numerous stimulus-based factors that could facilitate medical image reading. Of particular relevance to ASPECTS interpretation are the bit-depth of display monitors (grey levels), image resolution and signal-to-noise, and colour versus greyscale images.

2.4.2.1. Improving Display Quality and Learning Windowing Techniques

It may seem that increasing display bit-depth would increase reader accuracy, but this does not seem to be the case in practice: it was found that readers interpreting chest images did not perform differently when using an 11-bit display (2048 grey levels) compared to a standard 8-bit display (256 grey levels), although overall visual dwell time was less for 11-bit displays.[79] Thus, higher bit-depth displays may not improve ASPECTS accuracy or reliability, but they may improve the efficiency of ASPECTS reading.

Window setting is another issue that could influence ASPECTS scoring performance (Figure 3). Previous work in ASPECTS reliability has varied on this issue, with some studies prescribing a particular window setting and others encouraging readers to adjust it themselves. Arsava et al. [58] showed that allowing readers to choose their own window settings led to greater concordance between NCCT-ASPECTS and the ground truth of MR-ASPECTS relative to standard settings (width 80, center 20) irrespective of reader experience level.

2.4.2.2. Optimizing Post Processing of NCCT Scans

Image resolution may not have a great effect on lesion detection performance on NCCT, although there is some indication that a decreased signal-to-noise ratio degrades performance after a certain point.[80] Indeed, ASPECTS may be a more nuanced task than lesion detection, as it requires full assessment of potential tissue abnormalities in multiple delineated regions. It is therefore plausible that noise reduction on NCCT could have a marked effect on ASPECTS variability. Of course, one must be careful when using noise-reduction techniques so as to maintain the level of detail necessary to assess ASPECTS. Our group has recently shown that NCCT post-processing techniques may affect reliability of ASPECTS reading (Figure 2.4, A-D). Maximum Intensity Projections (MIPs) of NCCT are more reliable than average or thin slices in EIC detection.[81] However, further investigation in this area is required.

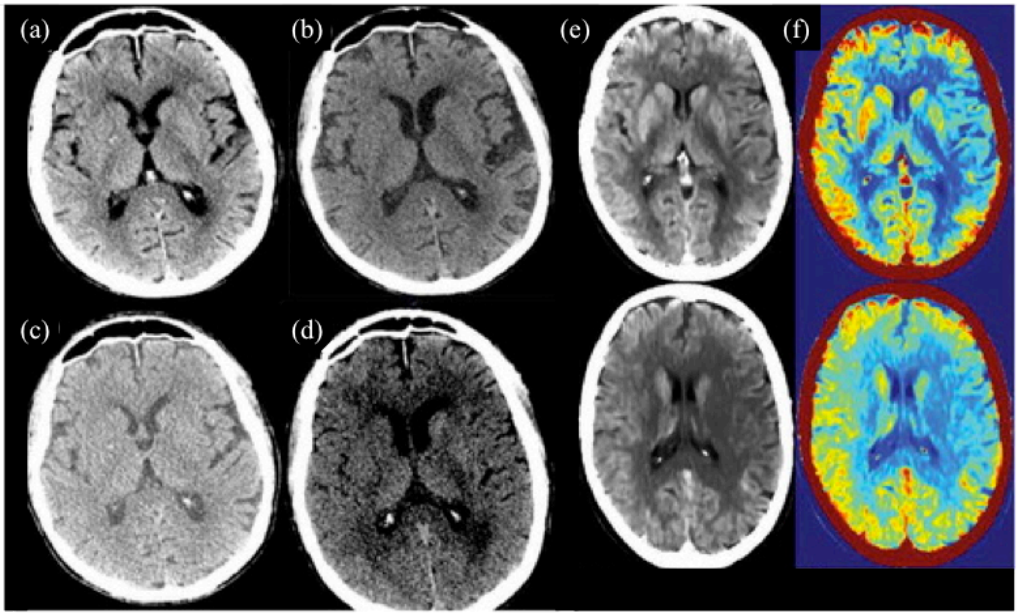

Figure 2.4. Post processing techniques (a–d) and enhancement algorithms (e–f) of CT scans contribute to variability in ASPECTS scoring. (a) Standard thickness 5 mm average CT; (b) Minimum intensity projections (mIPs) reconstructed to 5 mm; (c) Thin slices (0.625 mm); (d) Maximal intensity projections (MIPs) reconstructed to 5 mm; (e) Algorithm-enhanced grey-white matter, greyscale; (f) Algorithm-enhanced grey-white matter, color.

A potential area to explore in ASPECTS is the use of post-processing algorithms to generate colour CT scans, emphasizing grey-white matter differentiation (Figure 2.4, E-F). Our group has recently shown a benefit of this strategy in improving ASPECTS reading.[82]

2.5. Training

Perhaps the most evident way to mitigate the effects of the countless variables that can bias medical image interpretation is effective training. This refers not only to the process of teaching the ASPECTS system itself, but also teaching techniques to optimize environmental and cognitive conditions for ASPECTS reading. The effect of expertise on image interpretation has been well studied; specific training techniques have been explored to a lesser extent.

2.5.1. Expertise

Expertise can be an ambiguous topic. The usual psychological discourse defines expertise as possessing a certain level of domain-specific knowledge or proficiency in a skill domain. Radiologists could be expert image interpreters, for instance, because of a higher sensitivity to discrepant image features and greater clinical knowledge than non-experts. Dror [83] elaborates on this definition, claiming that expertise can be categorized based on the domains of *biasability*, which is one's susceptibility to being influenced by irrelevant external information, and *reliability*, the consistency between expert decisions in the absence of these irrelevant 'biasers'. By this framework, the highest level of expert performance would be maximally reliable and minimally biasable, both within and between individual experts. In addition to the level of expertise, these values can be affected by the strength of the biasing information, the difficulty of the decision being made, and the direction of the bias (and the risk of each bias).

Nakashima et al. [84] found that expert radiologists do not have a greater ability to detect lesions overall compared to novices, but their detection performance was better for clinically relevant lesions (cancer) than for non-significant lesions (bullae). Novices detected both types of

lesions at the same rate. Another study found experts to be faster and more accurate in ECG interpretation than novices; eye-tracking found that experts dwelled on findings for less time than novices.[85] Interestingly, these performance measures were not significantly affected in the experts or novices by the provision of clinical information. Other findings in mammography echo this relationship between speed and accuracy in experts compared to residents.[86]

The importance of experience in ASPECTS interpretation has been pointed out a number of times,[61,87] and it seems intuitive that expertise will increase performance. There have been very few studies on ASPECTS reliability that used raters of different experience levels; one did not report differences between novices and experts,[57] and another found that junior and senior readers did not differ significantly in terms of inter-class correlation,[48] although these readings were not compared to any gold standard to indicate if one group's scores were "more correct".

2.5.2. Training Techniques

There exist a multitude of training techniques for teaching medical image interpretation, especially now that the possibility for online and electronic modules exists. It remains unclear which of these techniques is most beneficial for turning novices into experts, but there are several key points worth discussing.

First, it is not the case that teaching purely analytic strategies is necessarily better than teaching non-analytic strategies. Although image interpretation can go awry when unconscious, automatic processing is unchecked by stepwise, logical analytic reasoning (resulting in cognitive bias), non-analytic processes seem to play a role in radiologic performance. For instance, it was found that students who were instructed to generate a diagnosis first and then list relevant features of the image performed better than those who were told to list relevant features and then diagnose based on the list.[88] Kok et al. [89] failed to demonstrate a performance benefit of systematic (assessing specific regions in a particular order) or full-coverage (assessing specific regions in any order) search strategies over non-systematic search, where readers were told to start

inspecting “whatever caught their attention”. More systematic viewing was associated with greater image coverage, but the full-coverage group showed significantly less sensitivity than non-systematic readers. Thus, there seems to be some benefit to generating a gestalt impression prior to beginning analytic search. In the Calgary Stroke Program, we teach residents and fellows to first look at the NCCT in a gestaltian manner to identify the extent of EIC; in the next step, we suggest that they identify if the EIC may be considered extensive, intermediate or small/minimal in size before interpreting the entire 10-point ASPECTS scale. We find that trichotomizing ASPECTS in this manner without formally scoring helps improve reliability (Figure 2.5).

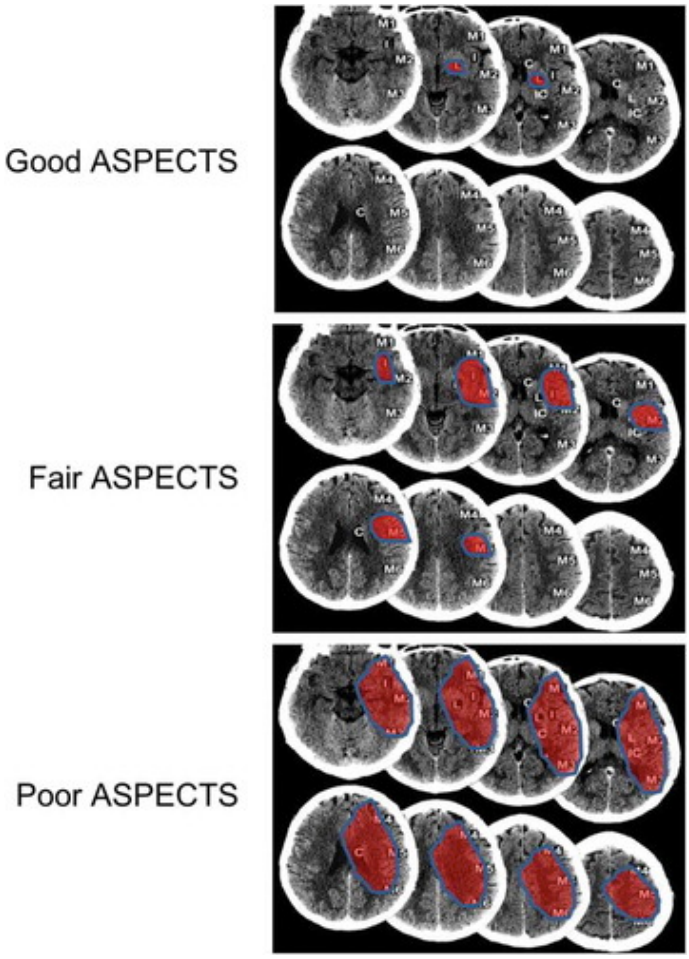

Figure 2.5. Qualitative trichotomization of ASPECTS (good, fair, poor) reflects the clinical application of ASPECTS.

Other strategies used by our program to inform ASPECTS scoring include recognizing that the internal capsule and M1 regions are particularly error-prone and using additional imaging signs such as the location of the dense vessel sign to hone in on affected regions.

Many errors or discrepancies in ASPECTS interpretation could derive from judgment errors, rather than perceptual errors; thus, it seems that teaching cognitive debiasing techniques could be effective against such errors. However, data suggest that simply lecturing to students about biases is insufficient to curtail the prevalence of these biases in practice. For instance, Sherbino et al. [90] taught medical students about the satisfaction of search bias and the availability bias in an interactive seminar with examples from clinical practice; this intervention failed to alter students' diagnostic behaviour or error rate .

Schuster et al. [91] propose a distinction between perceptual and conceptual training techniques for perceptual tasks, aimed at optimizing perception and judgment/interpretation processes, respectively. Perceptual learning is primarily a bottom-up process, driven by exposure to many instances of stimuli; conceptual learning is top-down, and involves developing one's "ability to categorize and differentiate things according to their features and characteristics".[91] The process of learning how to score ASPECTS as a medical professional often occurs on the job, which ought to capture both types of training. Repeated exposure to NCCT scans with varied infarcts on a day-to-day basis would constitute perceptual training, and would improve a student's ability to discriminate visual features on NCCT. Conversely, a lecture from an expert neuro-radiologist explaining how to differentiate between tissue affected and unaffected by EIC would inform the student's concept of EIC on NCCT, which is conceptual training.

From a cognitive neuroscience perspective, Dror [76] recommends that medical training should focus on error recovery techniques, in addition to error reduction. This requires the trainee to first learn how to detect a wide range of errors in others' and their own performance. Then, tools are provided to help induce recovery from such errors. At the Calgary Stroke Program, we apply some of these strategies for training residents and fellows on ASPECTS reading. In real

time and during stroke rounds, trainees have ample opportunity to compare their ASPECTS reads with those of experts. In addition, error recovery is enhanced by the availability of CTA collateral imaging, especially using multi-phase CTA. This serves as a further check on ASPECTS interpretation, as evidence suggests that patients with poor collateral circulation identified on multi-phase CTA are likely to have low ASPECTS and vice versa.[50]

2.6. Conclusion

Evaluating ASPECTS on NCCT is a crucial stage in acute stroke care, but it is a complex, cognitively demanding task. Few studies have directly investigated the many factors contributing to low-moderate inter-rater reliability in ASPECTS, and even fewer have measured intra-rater reliability. Features related to the ASPECTS methodology, image acquisition, patient history, reader variables, and reading conditions could lead to fluctuations in ASPECTS scoring. A number of top-down, bottom-up, and training-related interventions could help mitigate these effects by optimizing reading across individual contexts.

2.7. Expert Commentary

Because medicine requires a great deal of specialized training and expertise, it is sometimes assumed that all physicians behave equivalently. For instance, clinical tools are extensively validated, but we can fail to take into account individual variations in decision-making. Thus, minor or major fluctuations in the deployment of these clinical tools may influence patient care. Research involving clinicians' behavior requires the acknowledgement that medical staff are human, and therefore are subject to the same biases that affect all other professions. We feel that exploring this topic from a psychological perspective allows the application of cognitive theories of problem solving and bias, and offers the opportunity to incorporate cognitively informed solutions into medical training and practice.

We feel that future research in this area would benefit from more heterogeneous groups of image interpreters. Most prior studies investigating ASPECTS inter-rater reliability used only expert neurologists or neuro-radiologists. As we have stressed above, all humans behave differently; this can be attributable to any number of individual factors, from experience level to gender to geographical location of medical training. The inclusion of a wider range of image readers will be a crucial step in shedding light on the effect of these diverse variables on medical practice.

Moving forward in this field of research, it will also be important to report the contextual and environmental conditions during image interpretation sessions, including time of day, lighting, and the nature of the task. Psychological findings have demonstrated that the effect of such variables on individual performance can be substantial; accordingly, these conditions must be controlled between readers and reading sessions as much as possible.

2.8. Five-Year View

The development of automated computational tools to assess ASPECTS on NCCT is well underway. Machine learning techniques are becoming more prevalent and can now perform various image interpretation tasks, such as differentiating grey and white matter. We predict that, as automated tools become increasingly integrated into clinical contexts, the inter-rater reliability issue may become less pertinent than the issue of human versus computer performance. Nevertheless, we feel that variation in behavior between clinicians is a valuable topic of investigation with significant implications for medicine as a whole.

Key Issues

- The Alberta Stroke Program Early CT Score (ASPECTS) was developed as a semiquantitative tool for assessing the extent of early ischemic changes on non-contrast CT (NCCT) following acute ischemic stroke. It requires systematic visual search of 10 cortical and deep brain regions in the middle cerebral artery territory.
- A modest number of studies (approximately 30) have reported ASPECTS inter-rater reliability, and there is significant heterogeneity in these results. This variability could originate from technical, patient-level, reader-level, or reading context-level factors.
- Visual perception is a complex cognitive process involving interaction between bottom-up stimulus properties (retinal signals) and top-down cognitive influences. The use of cognitive shortcuts, or heuristics, can bias perception and affect our interpretation of what we see.
- Top-down interventions to improve ASPECTS reliability could include altering the task structure, working under a time limit, managing one's motivation and expectations, accessing clinical information to guide visual search, and using additional imaging to verify the ASPECTS score.
- Bottom-up interventions to improve ASPECTS reliability could include using higher bit-depth displays, employing post-processing techniques such as Maximal Intensity Projection (MIP), and using algorithms to enhance grey-white matter differentiation or to colorize NCCT images.
- Training techniques likely to develop reader expertise include teaching trainees to generate a gestalt/gist impression prior to initiating systematic search, combining elements of perceptual and conceptual training, and encouraging the development of error recovery strategies.

Additional Information

Funding: This paper was not funded.

CHAPTER THREE: THE EFFECT OF IMAGE READING CONTEXT FACTORS ON NON-CONTRAST CT ASPECTS RELIABILITY

3.1. Introduction

Image interpretation may seem like a simple task, but it is a complex cognitive problem with considerable opportunity for error. In medicine, image interpretation requires careful attention in order to accurately report the salient radiologic features.[92] However, human cognitive processes can be subject to external factors, resulting in systematic biases that colour "accurate" reading of an image.[93] It is therefore in the clinician's best interest to develop strategies and tools to reduce the effect of these external factors and homogenize image interpretation.

Figure 3.1 depicts a proposed simplified framework for the general cognitive functions involved in interpreting medical images, and the influences of certain external factors.

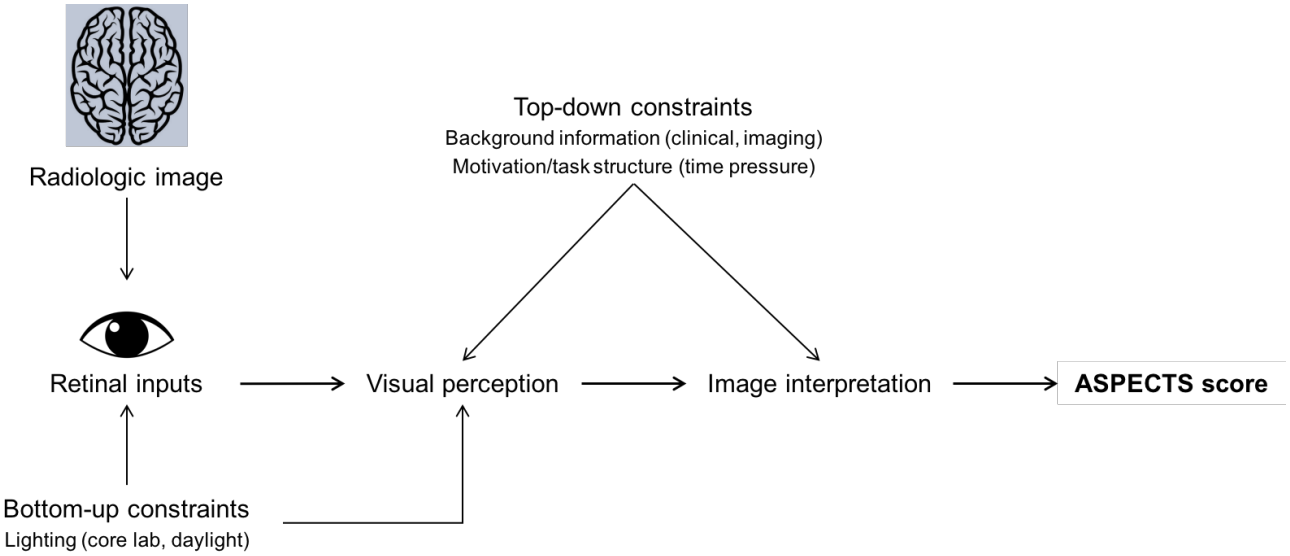

Figure 3.1. Flowchart illustrating the proposed cognitive framework underlying potential causes of variability between readers in medical image interpretation. Bottom-up and top-down variables, which can vary between individuals and contexts, constrain visual perception and subsequent processes of interpretation.

Both intuition and data tell us that expertise is key in medical image interpretation.[61,87] Expert image readers demonstrate a greater specificity for serious lesions when searching radiologic images.[84] Cognitive scientists have hypothesized that expert performance is marked by reduced biasability (being swayed by irrelevant contextual information) and increased reliability (consistent performance between and within individuals).[83] Thus, it is reasonable to posit that a radiologic image reader with more experience would show less susceptibility to the biasing effects of bottom-up and top-down contextual factors between reading sessions.[54]

NCCT-ASPECTS interpretation is subject to these cognitive constraints. Despite the strengths of ASPECTS, recent systematic reviews have found a wide degree of variability in measures of its IRR. Farzin et al. review the 30 prior studies that have investigated the reliability of ASPECTS; the reported agreement measures from these studies vary dramatically.[94] Such findings may reduce clinicians' confidence in the ASPECTS system and limit its use; therefore, it is crucial to endeavor to understand the diverse cognitive factors that could underlie the observed variability.

No prior studies have explicitly investigated the effect of external variables involving ASPECTS reading context, such as room lighting, time pressure, or the presence of clinical information or additional imaging. Thus, this study aims to manipulate specific contextual factors (ambient lighting, time pressure of <60 seconds per scan) and background information factors (presence or absence of clinical information and/or CT angiography scan) and observe how this affects NCCT-ASPECTS IRR. Additionally, NCCT-ASPECTS scores from readers of different experience levels will be compared to CTP-ASPECTS.

We hypothesize that varying the availability of prior information will affect IRR, with baseline clinical information and mCTA images each conferring benefit in this regard. We also predict that reading environment conditions will affect IRR, and greatest reliability will be associated with the *Core Lab* setting relative to the *Real-Life Lighting* or *Time Pressure*

scenarios. Finally, we predict that the expert rater's NCCT-ASPECTS scores will agree more with CTP-ASPECTS than the non-expert raters' scores.

3.2. Methods

Study population: 150 acute ischemic stroke patients who underwent baseline imaging less than 12 hours from stroke symptom onset and who had evidence of a symptomatic intracranial occlusion (intracranial ICA and/or M1- or proximal M2-MCA) were selected from the PRove-IT (Precise and Rapid assessment of collaterals using multi-phase CTA in the triage of patients with acute ischemic stroke for IA Therapy) study.[50] Baseline imaging included NCCT, mCTA, and CTP. Five mm average thickness baseline NCCT scans and TMax (>16 sec and >20 sec) CTP maps were used for the study.

Image readers: For all patients, NCCT-ASPECTS was scored by a trainee stroke neurology fellow with no ASPECTS reading experience, a senior stroke neurology fellow with >1 years of ASPECTS reading experience, and a neuro-radiologist with >5 years of ASPECTS reading experience. Two experienced stroke neurologists scored CTP-ASPECTS by consensus. Background patient or scan characteristics including leukoaraiosis, old infarcts, and motion artifact were evaluated by an independent neurologist.

Reading conditions: For each reader, image interpretation occurred over three reading sessions, each session separated by at least two weeks. In each reading session, all 150 NCCT scans were presented in a random order and scored (total ASPECTS, individual regions). For the first session, readers were only shown NCCTs; in the second session, they viewed NCCTs after being provided with specific clinical information (patient's age, sex, NIH Stroke Scale at baseline, and affected hemisphere), and in the third session, they viewed NCCTs and baseline mCTAs while being provided with the same clinical information as in the previous session. These comprise the three background information conditions.

Within each reading session, 50 scans each were allocated to three contextual conditions: *Real-Life Lighting*, where the reading environment had bright ambient lighting to reflect the environment of the emergency room; *Core Lab*, with low ambient lighting and minimal noise distraction; and *Time Pressure*, where each scan had to be interpreted within 60 seconds. These contextual conditions occurred in a random order for each reader. Readers were free to set their own window/level as desired and had access to all scan slices. The images were displayed on the same computers across all reading sessions and contextual conditions.

3.2.1. Statistical Analysis

Intraclass correlation coefficient (ICC) estimates and their 95% confidence intervals were calculated using a two-way, absolute-agreement, single-rater random effects model for total ASPECTS, trichotomized ASPECTS (0-4, 5-7, 8-10), and individual ASPECTS regions. Trichotomized ASPECTS was included because this may directly influence the clinical use of ASPECTS. Region level involvement on the ASPECTS template and trichotomized ASPECTS IRR were also analysed using Light's kappa, a kappa-like Pi-family statistic for more than two coders. In order to obtain Light's kappa, linearly weighted Cohen's kappa was calculated for each rater pair and the arithmetic mean of these values was taken. A discussion of the appropriateness of ICC and kappa statistics in the present study can be found in Appendix A.

Bland-Altman plots were employed to compare total ASPECTS scores between raters of different expertise levels.

These measures were calculated using R statistical software (R Foundation for Statistical Computing, Vienna, Austria) and MedCalc for Windows, version 14 (MedCalc Software, Ostend, Belgium).

3.3. Results

Of the 150 patients from the PRove-IT database, 91 (61%) were males, and the median age was 71 (IQR 63-78). Median NIHSS at baseline was 8 (IQR 5-16). 19 patients (12.7%) had

internal carotid artery-MCA occlusions, 72 (48%) had exclusively M1-MCA or proximal M2-MCA occlusions, and 36 (24%) had no visible occlusion (Table 3.1).

The median onset-to-CT time overall was 132 minutes (IQR 81-244 min). The 50 cases assessed under *Time Pressure* had a median onset-to-CT time of 123 minutes (IQR 85-186 min); the 50 *Real-Life Lighting* cases had a median time of 151 minutes (IQR 91-311 min), and the 50 *Core Lab* cases had a median time of 107 minutes (IQR 69-257 min) (Table 3.1).

	All (n=150)	Real-Life Lighting (n=50)	Core Lab (n=50)	Time Pressure (n=50)
Age (years) – Median (IQR)	71 (63-78)	72 (64.3-81.8)	70.5 (65-75.8)	68.5 (59.3-78.8)
Sex (male) – %	60.6	64	64	54
Right hemisphere affected – N (%)	76 (50.7)	23 (46)	25 (50)	28 (56)
NIHSS – Median (IQR)	8 (5-16)	8.5 (6-15)	9 (5-17)	8 (3-17)
Occlusion location – N (%)				
MCA	72 (48)	23 (46)	25 (50)	24 (48)
ICA/M1-MCA	19 (12.7)	5 (10)	8 (16)	6 (12)
Other (ACA, PCA, Basilar, Vertebral)	23 (15.3)	13 (26)	6 (12)	4 (8)
None visible	36 (24)	9 (18)	11 (22)	16 (32)
Onset-to-CT time (min) – N (%)				
0-90	47 (31.3)	20 (40)	12 (24)	15 (30)
91-180	46 (30.7)	10 (20)	15 (30)	21 (42)
181-270	21 (14)	6 (12)	7 (14)	8 (16)
>270	36 (24)	14 (28)	16 (32)	6 (12)

Table 3.1. Baseline demographic characteristics of the patients selected from the PRove-IT database. IQR: Interquartile range, MCA: Middle cerebral artery, ICA: Internal carotid artery, M1: M1 segment of the MCA, ACA: Anterior cerebral artery, PCA: Posterior cerebral artery.

Agreement for total ASPECTS between all raters is presented in Table 3.2. There was a general trend that more available information (clinical information, mCTA) improved reliability. ICC values for all 150 cases were 0.187, 0.385, and 0.473 for *NCCT Only*, *NCCT + Clinical Information*, and *NCCT + Clinical Information + CTA* conditions, respectively. Bland-Altman plots for each rater pair are shown in Figure 3.1.

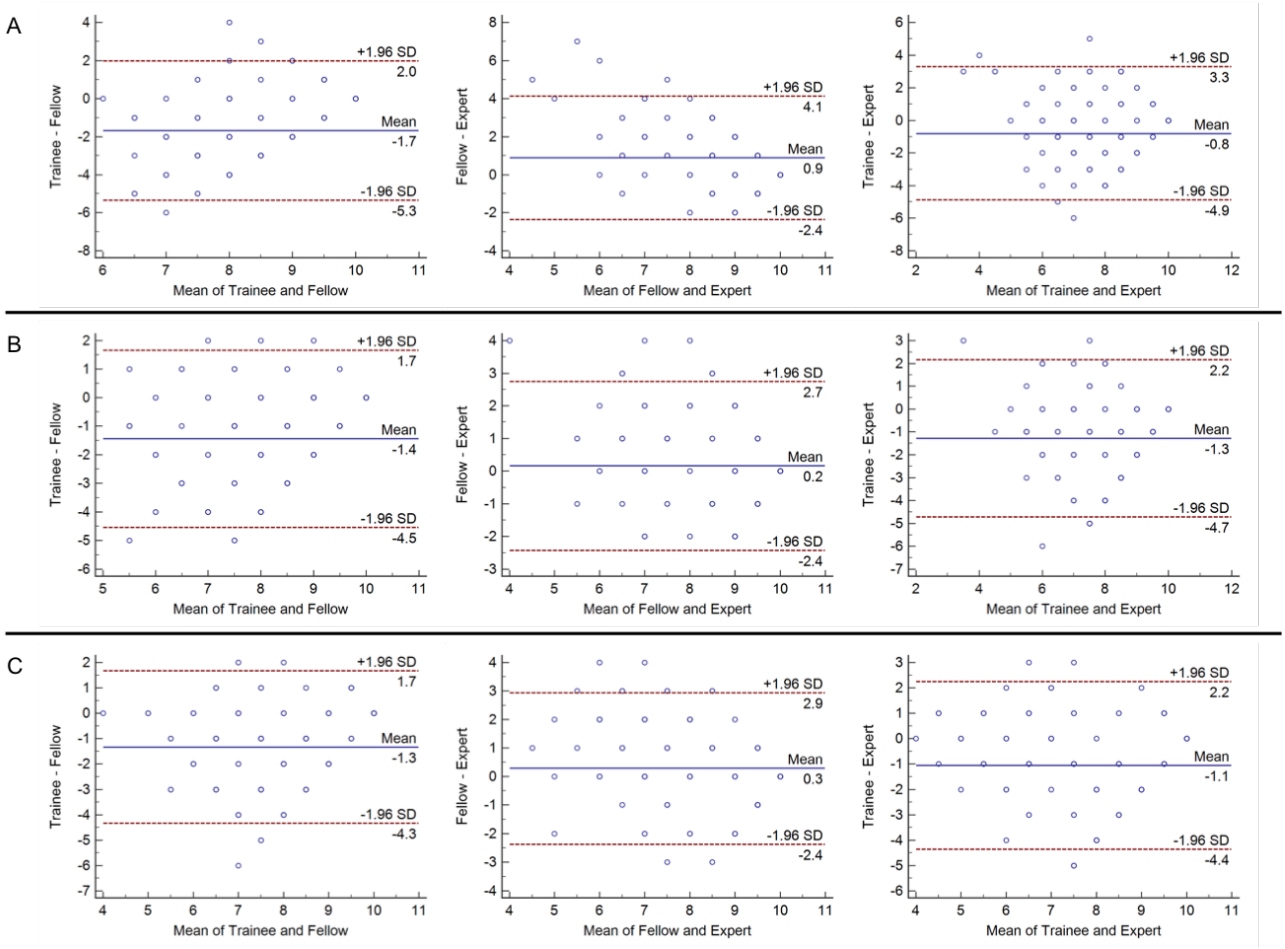

Figure 3.2. Bland-Altman plots depicting the agreement between each pair of raters for each of the three prior information conditions. A) NCCT only; B) NCCT + Clinical information; C) NCCT + Clinical information + CTA. *Trainee* is the junior stroke fellow, *Fellow* is the senior stroke fellow, and *Expert* is the neuro-radiologist. SD: Standard deviation, NCCT: Non-contrast computed tomography, CTA: Computed tomography angiography.

	Overall (n=150)		Real-Life Lighting (n=50)		Core Lab (n=50)		Time Pressure <60 sec (n=50)	
	ICC (95% CI)	Light's κ	ICC (95% CI)	Light's κ	ICC (95% CI)	Light's κ	ICC (95% CI)	Light's κ
NCCT	0.187 (0.070-0.307)	0.110	0.180 (0.028-0.358)	0.119	0.205 (0.046-0.386)	0.101	0.186 (0.012-0.381)	0.123
NCCT + Clin	0.385 (0.191-0.544)	0.277	0.320 (0.023-0.579)	0.230	0.387 (0.178-0.576)	0.253	0.558 (0.393-0.702)	0.398
NCCT + Clin + CTA	0.473 (0.282-0.618)	0.324	0.510 (0.281-0.687)	0.342	0.231 (0.035-0.436)	0.187	0.672 (0.508-0.793)	0.450

Table 3.2. Inter-rater reliability estimates for total ASPECTS between all three raters for the three prior-information conditions (rows) and in the three reading environment subgroups (columns). Clinical information included age, sex, baseline NIHSS, and affected side. ICC: Intraclass correlation coefficient, CI: Confidence interval, NCCT: Non-contrast computed tomography, Clin: Clinical information, CTA: Computed tomography angiography.

Providing clinical information resulted in a statistically significant increase in agreement in the *Time Pressure* condition, and providing clinical information and mCTA improved agreement in the *Core Lab* condition relative to NCCT alone (Table 3.2). There was a non-significant trend that agreement was better in *Time Pressure* conditions relative to the other environmental conditions. Times taken to score individual scans in the non-*Time Pressure* conditions are listed in Table 3.3; generally, raters scored scans in less than sixty seconds even without any time constraint.

	Core Lab (n=50) Median (IQR)	**Real-Life Lighting (n=50)** Median (IQR)
NCCT		
Trainee	34 (24-45.5)	33 (28-40)
Fellow	66 (53-76.8)	71 (59.3-85.5)
Expert	47.5 (40-57)	64 (54.5-79.5)
NCCT + Clin		
Trainee	25.5 (22-36.3)	36 (28.3-45)
Fellow	48 (41.3-58.5)	45 (35-52.5)
Expert	65 (52.3-72.5)	44.5 (37.3-50)
NCCT + Clin + CTA		
Trainee	36.5 (32.3-42.8)	36 (30-45)
Fellow	52 (45-59)	43 (36.3-58.5)
Expert	32 (28-38.8)	38 (33.3-53.8)

Table 3.3. Median image interpretation times (seconds per NCCT scan) for the non-*Time Pressure* subgroups. In the *Time* Pressure condition, scoring time was prescribed and not measured. IQR: Interquartile range, NCCT: Non-contrast computed tomography, Clin: Clinical information, CTA: Computed tomography angiography.

For trichotomized ASPECTS (0-4, 5-7, 8-10), overall IRR was comparable to that of total ASPECTS (Table 3.4). For the *Real-Life Lighting* and *Time Pressure* subgroups, IRR was

significantly greater in the *NCCT + Clinical + CTA* condition than in the *NCCT-only* condition. No contextual condition consistently improved performance relative to the others.

	Overall (n=150)		Real-Life Lighting (n=50)		Core Lab (n=50)		Time Pressure <60 sec (n=50)	
	ICC (95% CI)	Light's κ	ICC (95% CI)	Light's κ	ICC (95% CI)	Light's κ	ICC (95% CI)	Light's κ
NCCT	0.118 (0.027-0.219)	0.184	0.104 (-0.025-0.267)	0.115	0.208 (0.047-0.391)	0.215	0.074 (-0.032-0.216)	0.046
NCCT + Clin	0.278 (0.155-0.399)	0.291	0.193 (0.020-0.385)	0.243	0.342 (0.165-0.521)	0.281	0.370 (0.199-0.544)	0.384
NCCT + Clin + CTA	0.396 (0.270-0.511)	0.361	0.516 (0.345-0.669)	0.484	0.171 (0.021-0.348)	0.187	0.508 (0.338-0.661)	0.408

Table 3.4. Inter-rater reliability estimates for trichotomized ASPECTS (0-4, 5-7, 8-10) between all three raters for the three prior-information conditions (rows) and in the three reading environment subgroups (columns). Clinical information included age, sex, baseline NIHSS, and affected side. ICC: Intraclass correlation coefficient, CI: Confidence interval, NCCT: Non-contrast computed tomography, Clin: Clinical information, CTA: Computed tomography angiography.

Regionwise reliability is shown in Table 3.5. In general, most regions had greatest reliability in the *NCCT + Clinical + CTA* condition, although reliability was poor overall (ICC < 0.3 for all regions, for all conditions).

ICC values for the background information conditions stratified by patient-level variables (presence or absence of motion artefact, old infarct, and leukoaraiosis; onset-to-CT time; NIHSS at baseline; and site of occlusion) are presented in Table 3.6. Once again, IRR tended to increase with the availability of more background information. No statistically significant differences in reliability were noted by these various patient-level variables. However, reliability may have improved somewhat in the *NCCT* condition in patients without old infarcts and leukoaraiosis and in patients with more clinically severe strokes.

	N	**NCCT** ICC (95% CI)	**NCCT + Clin** ICC (95% CI)	**NCCT + Clin + CTA** ICC (95% CI)
Motion artifact				
Present	18	0.257 (0.013-0.553)	0.440 (0.163-0.704)	0.347 (0.076-0.635)
Absent	132	0.176 (0.062-0.297)	0.371 (0.167-0.538)	0.495 (0.300-0.641)
Old infarct				
Present	18	0.052 (-0.100-0.311)	0.216 (-0.091-0.509)	0.422 (0.143-0.692)
Absent	132	0.201 (0.078-0.328)	0.404 (0.211-0.562)	0.476 (0.279-0.626)
Leukoaraiosis				
Present	27	-0.017 (-0.143-0.184)	0.313 (0.086-0.555)	0.480 (0.234-0.695)
Absent	123	0.243 (0.101-0.383)	0.399 (0.190-0.567)	0.465 (0.264-0.619)
Onset-to-CT time (min)				
0-90	47	0.214 (0.047-0.403)	0.330 (0.101-0.542)	0.572 (0.327-0.742)
90-180	46	0.135 (-0.010-0.313)	0.336 (0.121-0.541)	0.343 (0.140-0.539)
180-270	21	0.350 (0.065-0.629)	0.498 (0.208-0.736)	0.431 (0.165-0.682)
>270	36	0.063 (-0.074-0.252)	0.426 (0.202-0.629)	0.429 (0.190-0.841)
Baseline NIHSS				
0-5	45	0.029 (-0.073-0.173)	0.316 (0.099-0.525)	0.268 (0.094-0.459)
6-15	64	0.197 (0.052-0.358)	0.267 (0.082-0.451)	0.281 (0.060-0.489)
>15	41	0.194 (0.026-0.392)	0.202 (0.031-0.400)	0.544 (0.336-0.713)
Site of occlusion				
None visible	36	0.063 (-0.054-0.230)	0.104 (-0.033-0.288)	0.015 (-0.084-0.166)
MCA	72	0.193 (0.055-0.345)	0.368 (0.138-0.520)	0.408 (0.237-0.564)
PCA	11	-0.041 (-0.165-0.256)	0.324 (0.006-0.872)	-0.030 (-0.121-0.208)
MCA and ICA	19	0.144 (-0.050-0.424)	0.073 (-0.124-0.365)	0.346 (0.052-0.638)
Other	12	0.113 (-0.119-0.488)	0.030 (-0.093-0.305)	0.031 (-0.090-0.305)

Table 3.5. Intraclass correlation coefficient point estimates and 95% confidence intervals for total ASPECTS across all three raters, stratified by baseline patient and imaging characteristics. *MCA* refers to M1-MCA or proximal M2-MCA. ICC: Intraclass correlation coefficient, CI: Confidence interval, NCCT: Non-contrast computed tomography, Clin: Clinical information, CTA: Computed tomography angiography, NIHSS: National Institutes of Health Stroke Scale, MCA: Middle cerebral artery, PCA: Posterior cerebral artery, ICA: Internal carotid artery.

	NCCT			NCCT + Clin			NCCT + Clin + CTA		
	ICC	95% CI	Light's κ	ICC	95% CI	Light's κ	ICC	95% CI	Light's κ
Caudate	0.071	-0.016-0.169	0.072	0.116	0.021-0.220	0.101	0.228	0.128-0.334	0.203
Lentiform	0.145	0.052-0.247	0.154	0.085	-0.004-0.186	0.173	0.212	0.114-0.317	0.222
Insula	0.097	0.012-0.193	0.098	0.189	0.090-0.296	0.211	0.249	0.145-0.357	0.257
Internal Capsule	0.018	-0.042-0.092	0.070	0.070	-0.007-0.158	0.058	0.024	-0.046-0.106	0.019
M1	0.010	-0.054-0.087	0.049	-0.002	-0.066-0.075	0.052	0.035	-0.023-0.105	0.025
M2	0.171	0.073-0.277	0.163	0.170	0.075-0.275	0.162	0.165	0.070-0.269	0.150
M3	0.022	-0.063-0.120	0.078	0.087	-0.007-0.191	0.195	0.054	-0.033-0.154	0.075
M4	0.171	0.072-0.278	0.170	0.034	-0.056-0.136	0.158	0.098	0.009-0.198	0.097
M5	0.070	-0.022-0.173	0.067	0.205	0.104-0.311	0.203	0.210	0.112-0.315	0.199
M6	0.144	0.048-0.248	0.150	0.193	0.095-0.298	0.186	0.238	0.138-0.343	0.229

Table 3.6. Intraclass correlation coefficient point estimates with their 95% confidence intervals and Light's κ values for ASPECTS regionwise agreement between all three raters. ICC: Intraclass correlation coefficient, CI: Confidence interval, NCCT: Non-contrast computed tomography, Clin: Clinical information, CTA: Computed tomography angiography, M1-M6: M1-M6 regions of the middle cerebral artery cortical territory.

The expert rater agreed more with CTP-ASPECTS (TMax >16 and >20) than the senior fellow or trainee fellow (Table 3.7). In this test, raters used NCCT, clinical information, and CTA to score NCCT-ASPECTS.

	Tmax >16 s ICC (95% CI)	**Tmax >20 s** ICC (95% CI)
Trainee	0.47 (0.33-0.59)	0.49 (0.35-0.60)
Fellow	0.51 (0.38-0.63)	0.51 (0.38-0.62)
Expert	0.69 (0.59-0.77)	0.68 (0.58-0.76)

Table 3.7. Intraclass correlation coefficient point estimates and 95% confidence intervals for each rater's total ASPECTS agreement with CT perfusion-ASPECTS, scored using two different Tmax thresholds. ICC: Intraclass correlation coefficient; CI: Confidence interval.

3.4. Discussion

3.4.1. Summary of Results

The results of the present study demonstrate that factors involving the reading context and environment can significantly affect the inter-rater reliability of NCCT-ASPECTS. The addition of background clinical information regarding the patient generally resulted in greater reliability than NCCT alone. mCTA images on top of the clinical information conferred additional benefit in most cases with certain exceptions, such as in the *Core Lab* setting.

ASPECTS IRR in real life conditions (high ambient levels of light) was comparable to that in core lab conditions (low ambient levels of light). Time pressure (each patient scored in <60 seconds) did not significantly affect IRR; in fact, most cases were scored in less than one minute in the non-time pressure conditions. It appears that the novice rater generally took less

time to score ASPECTS, suggesting that a potential training technique may consist of encouraging junior raters to spend more time on scoring.

Trichotomized ASPECTS did not demonstrate greater IRR values than total ASPECTS. Reliability remained relatively poor between all three raters, with a maximal ICC of 0.516 (95% CI: 0.345-0.669) in the *Real-Life Lighting* subgroup of *NCCT + Clinical + CTA*.

As hypothesized, the expert neuro-radiologist rater agreed most with CTP-ASPECTS, which is an approximation of ground truth. The junior and senior fellows did not differ significantly in this regard, suggesting that ASPECTS proficiency requires more than one year of training in the methodology.

3.4.2. Exploration of Cognitive Explanations for Observed Effects

According to Krupinski, an expert in the psychology of medical image perception, "Image perception is likely the most prominent, yet least appreciated, source of error in diagnostic imaging."[55] The vast body of cognitive psychology research has not been applied to medical image perception to any great extent. Thus, drawing any firm cognitive conclusions from these results is challenging. However, some broad inferences can be drawn to explain the results.

a. Prior information (clinical information)

Top-down inputs, such as knowledge of a patient's stroke lateralization based on clinical symptoms, can alter performance on perceptual tasks like ASPECTS scoring by differentially tuning neurons, altering the 'salience landscape' that guides image interpretation. An altered salience landscape causes different features to 'jump out', even if the stimulus itself does not change.[95] It seems that background clinical information and, in most cases, mCTA information guide image interpretation in a manner that facilitates ASPECTS scoring.

b. Time pressure

Time pressure, where one feels that they have less time available than would usually be required to complete a task, is another top-down factor in image interpretation. In response to

time pressure, people seem to rely on different cognitive strategies than those they would have otherwise employed. One example of such coping strategies is using simplified heuristics in probability judgments; for instance, one simplified heuristic is the anchoring heuristic, where the initial information that one receives serves as an 'anchor' or reference point for the interpretation of subsequent information.[75] In many tasks, this can reduce accuracy and increase the risk of error; however, the present results do not indicate that time pressure is detrimental to IRR in the context of ASPECTS. It may be that a modest amount of time pressure fosters a cognitive style that is beneficial to ASPECTS scoring.

c. Ambient lighting

In addition to top-down factors, bottom-up signals from the retina can affect the cognitive processes of image interpretation. Seemingly irrelevant conditions that can vary between reading sessions, such as ambient lighting, can change the incoming visual signals and thereby could conceivably affect image interpretation.[64] In the present study, lower ambient lighting (core lab) did not facilitate ASPECTS reliability relative to higher ambient lighting (real life); this seems counterintuitive, as the core lab is viewed as an ideal radiologic reading environment. However, a plausible explanation of this result is that novice raters were less comfortable in the core lab environment than the real life environment they are used to, thus affecting their scoring. Further research in this area is needed to shed more light on these findings.

d. Expertise

Experts perform differently than novices in medical image interpretation tasks. For instance, in a lung lesion detection task, radiologists demonstrated better specificity to pathologically significant lesions even though they showed comparable lesion sensitivity as novice (non-radiologist) readers.[84] In the context of stroke (visual inspection of head CT scans), an eye-tracking experiment revealed that experienced neurologists dwelled on visually salient image features to the same extent as novices, but they also focused on additional clinically relevant regions, such as an anterior cerebral artery infarction area.[71]

As discussed above, it is plausible to posit that expert neuro-radiologists are less susceptible to irrelevant contextual factors than image interpreters with less expertise. The distinction between high versus low ambient light, for instance, is less likely to affect experts' performance in ASPECTS interpretation. The inclusion of raters with different experience levels in this study could constitute a major reason for the low observed IRR estimates.

It is important to note that expertise, in medical image interpretation or any other area, is domain-specific: an expert at detecting lung lesions, for instance, is not necessarily an expert at detecting early ischemic neurological changes.

3.4.3. Limitations

The results of this study could be generalized more readily with more raters for each level of experience. Although we have associated some observed effects to raters' expertise, it must be acknowledged that these differences could be due to variation between individuals.

Another limitation is that the 60-second time limit in the *Time Pressure* condition may not have truly imposed time pressure on the raters; this is evidenced by the observation that the majority of scans in the non-time pressure conditions were scored in <60 seconds. Psychological findings indicate that time pressure induces a distinct cognitive processing style, as discussed above;[75] therefore, it would be informative to repeat the time pressure experiment with a more stringent time limit. Nevertheless, the 60-second limit may have been sufficient to induce a facilitative cognitive style, resulting in improved IRR in some cases.

3.4.4. Conclusions

Altering the reading context (ambient lighting, time pressure) or background information (clinical information, mCTA) affected the inter-rater reliability of ASPECTS scored on baseline non-contrast CT by three raters of different experience levels. The NCCT-ASPECTS scores of the expert rater (neuro-radiologist) showed the greatest concordance with CTP-ASPECTS, and

the scores of the trainee showed the least. These results support the hypothesis that NCCT-ASPECTS interpretation is susceptible to factors that can vary between individual reading sessions, particularly in novice raters. It is important to maintain as many of these variables as possible constant between image reading sessions, and to recognize that individual rater factors can influence medical image interpretation in unpredictable ways.

CHAPTER FOUR: FUTURE DIRECTIONS

4.1. Summary

In order to limit the extent of ischemia in acute ischemic stroke, interventions (thrombolysis, EVT) have been developed to rapidly restore perfusion to cerebral tissue. In order for clinicians to select between therapeutic alternatives, the extent of EICs must be evaluated using NCCT; in current practice, ASPECTS is a straightforward semiquantitative score commonly used for this purpose. However, the inter-rater reliability of ASPECTS has been under-investigated, and the existing results are somewhat inconsistent: there is significant heterogeneity in reported measures of reliability, which brings into question the clinical applicability of this score.

Plausible links can be drawn between the process of NCCT interpretation for ASPECTS scoring and the psychological literature pertaining to visual perception and radiologic performance.

The cognitive process of visual perception does not represent an objective photocopy of the world; it is a complex process where visual input interacts with top-down cognitive inputs to produce a dynamic and subjective saliency map. Variables deriving both from retinal inputs (for example, room lighting) and from cognitive inputs (for example, one's understanding of the task they are engaged in) can affect perception.

Medical image interpretation requires a unique set of cognitive skills. ASPECTS may be cognitively challenging because it requires reconstruction of a three-dimensional space using two-dimensional axial brain slices, and it requires both lesion detection and lesion interpretation, increasing the complexity of the task.

There are several variables that could introduce heterogeneity into ASPECTS scoring, in three broad categories: technical factors relating to the medical images themselves; patient factors such as age, presence of old infarcts, or white matter disease; and reader factors relating to

characteristics of the reader (e.g. experience, training, risk or ambiguity aversion) or the reading conditions (e.g. reading environment, task structure, fatigue).

Prior knowledge, including clinical information and additional vascular imaging, is a contextual factor that can affect the IRR of NCCT-ASPECTS by raters of different experience levels. The availability of this background information was shown to generally improve IRR. Reading environment factors, including level of ambient lighting and time pressure, also affected ASPECTS IRR, although not to a significant extent. The novice rater agreed the least with CTP-ASPECTS, and the expert rater agreed the most. These results suggest that reading-context variables are a contributing factor to the extent of IRR, and that experts may be more reliable in their ASPECTS scoring across different contexts.

4.1.1. Limitations

This work is not without its challenging aspects. Firstly, the present work investigated only a small subset of candidate reading session variables. There are many more potential factors that could not be explored due to practical constraints (Table 2.1). However, the factors that were considered here (knowledge of patient's clinical presentation, additional vascular imaging, ambient lighting, and time pressure) are relevant to all radiologic image interpretation in stroke (especially in the emergent setting) and the results provide promising evidence that other factors could also influence ASPECTS reading.

As discussed above, the experiments carried out in Chapter 3 included only one rater from each experience level, and we cannot extend the findings to all raters with similar expertise; indeed, the present results do not shed light on the degree of heterogeneity within rater groups. However, the finding that novice trainees, senior fellows, and expert neuro-radiologists perform differently when scoring ASPECTS has significant educational and clinical implications.

The lack of a true gold standard or ground truth to corroborate early ischemic changes on hyperacute NCCT is a further limitation, as we cannot determine which raters are more "correct" than others. Other imaging modalities such as MR (DWI) can be used for more precise lesion

visualization, but NCCT may not include equivalent informational content to MR images. CTP-ASPECTS assessed on cerebral blood volume (CBV) maps may be predictive of infarct core size and clinical outcomes if truncation artifacts are avoided, so CBV-ASPECTS may be a potential target of future investigation.[96]

A final limitation involves the concept that inter-rater reliability is merely a proxy for 'correctness'. We assume that the expert rater is able to detect EICs more accurately than other raters, but they may be subject to their own systematic biases. As a result, improved reliability may not reflect more accurate scoring in an absolute sense. As discussed above, there is no gold standard for ASPECTS scoring, so it may not be possible to circumvent this limitation. Ultimately, though, it is important to investigate and optimize ASPECTS reliability between raters of different experience levels, as this will improve the rigour and consistency of clinical applications of ASPECTS.

4.2. Future Directions

First and foremost, future studies in this area of ASPECTS reliability would be most effective and meaningful if investigators held constant as many environmental and contextual factors as possible (except the variable being investigated, of course) between raters and between image reading sessions. Additionally, studies investigating other physician-level factors, including age, gender, or medical specialty, as well as personality variables such as risk aversion or ambiguity aversion would shed more light on which physician demographics may be more susceptible to greater fluctuations in intra-rater reliability. In the context of medical practice, cognitive factors pertaining to the consistency of image reading, such as decision fatigue, visual fatigue, and date or time of assessment may be relevant targets for investigation.

As the present work suggests that rater expertise influences medical image interpretation, subsequent studies to assess training programs for ASPECTS novices and to test machine-learning algorithms to help improve ASPECTS interpretation, could be informative and relevant.

4.3. Conclusion

This work imparts an important message: clinicians cannot neglect the fact that decision-making processes are never objective; they can be influenced by innumerable internal and external factors that may seem irrelevant to the problem at hand. For practitioners engaging in ASPECTS scoring in acute ischemic stroke, it would be beneficial to consider the environmental conditions and the availability of prior information. It would also be beneficial to acknowledge when relevant that novice ASPECTS scorers may perform differently than expert scorers.

Even with the increasing application of machine learning and other computational techniques to medical image interpretation tasks, it is necessary to recognize the role that human psychology plays in radiologic assessment. In ischemic stroke, NCCT-ASPECTS is one of the most important clinical tools for prognostication and treatment selection in the hyperacute stage. Insights into the systematic biases and heuristics underlying the cognitive processes of image interpretation have the potential to increase the clinical utility of this tool and to inform the design of educational programs for future trainees.

REFERENCES

1. Allen CL, Bayraktutan U. Oxidative Stress and Its Role in the Pathogenesis of Ischaemic Stroke. Int J Stroke. 2009;4:461–70.

2. Phan TG, Wright PM, Markus R, Howells DW, Davis SM, Donnan GA. Salvaging the ischaemic penumbra: more than just reperfusion? Clin Exp Pharmacol Physiol. 2002;29:1–10.

3. Xing C, Arai K, Lo EH, Hommel M. Pathophysiologic Cascades in Ischemic Stroke. Int J Stroke. 2012;7:378–85.

4. Bahr Hosseini M, Liebeskind DS. The role of neuroimaging in elucidating the pathophysiology of cerebral ischemia. Neuropharmacology [Internet]. 2017; Available from: http://www.sciencedirect.com/science/article/pii/S0028390817304495

5. Simard JM, Kent TA, Chen M, Tarasov KV, Gerzanich V. Brain oedema in focal ischaemia: molecular pathophysiology and theoretical implications. Lancet Neurol. 2007;6:258–68.

6. Menon BK, Puetz V, Kochar P, Demchuk AM. ASPECTS and Other Neuroimaging Scores in the Triage and Prediction of Outcome in Acute Stroke Patients. Neuroimaging Clin N Am. 2011;21:407–23.

7. Demchuk AM, Hill MD, Barber PA, Silver B, Patel SC, Levine SR. Importance of Early Ischemic Computed Tomography Changes Using ASPECTS in NINDS rtPA Stroke Study. Stroke. 2005;36:2110–5.

8. Hacke W, Kaste M, Fieschi C, Toni D, Lesaffre E, Kummer R von, et al. Intravenous Thrombolysis With Recombinant Tissue Plasminogen Activator for Acute Hemispheric Stroke: The European Cooperative Acute Stroke Study (ECASS). JAMA. 1995;274:1017–25.

9. Barber PA, Demchuk AM, Zhang J, Buchan AM. Validity and reliability of a quantitative computed tomography score in predicting outcome of hyperacute stroke before thrombolytic therapy. ASPECTS Study Group. Alberta Stroke Programme Early CT Score. Lancet Lond Engl. 2000;355:1670–4.

10. Saver JL. Time Is Brain—Quantified. Stroke. 2006;37:263–6.

11. Huang X, Moreton FC, Kalladka D, Cheripelli BK, MacIsaac R, Tait RC, et al. Coagulation and Fibrinolytic Activity of Tenecteplase and Alteplase in Acute Ischemic Stroke. Stroke. 2015;46:3543–6.

12. Acheampong P, Ford GA. Pharmacokinetics of alteplase in the treatment of ischaemic stroke. Expert Opin Drug Metab Toxicol. 2012;8:271–81.

13. The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. Tissue Plasminogen Activator for Acute Ischemic Stroke. N Engl J Med. 1995;333:1581–8.

14. Hacke W, Kaste M, Bluhmki E, Brozman M, Dávalos A, Guidetti D, et al. Thrombolysis with alteplase 3 to 4.5 hours after acute ischemic stroke. N Engl J Med. 2008;359:1317–29.

15. Powers WJ, Rabinstein AA, Ackerson T, Adeoye OM, Bambakidis NC, Becker K, et al. 2018 Guidelines for the Early Management of Patients With Acute Ischemic Stroke: A Guideline for Healthcare Professionals From the American Heart Association/American Stroke Association. Stroke. 2018;49:e46–99.

16. Saver JL, Yafeh B. Confirmation of tPA Treatment Effect by Baseline Severity-Adjusted End Point Reanalysis of the NINDS-tPA Stroke Trials. Stroke. 2007;38:414–6.

17. Kharitonova T, Ahmed N, Thorén M, Wardlaw JM, von Kummer R, Glahn J, et al. Hyperdense middle cerebral artery sign on admission CT scan--prognostic significance for ischaemic stroke patients treated with intravenous thrombolysis in the safe implementation of thrombolysis in Stroke International Stroke Thrombolysis Register. Cerebrovasc Dis. 2009;27:51–9.

18. Riedel CH, Zimmermann P, Jensen-Kondering U, Stingele R, Deuschl G, Jansen O. The Importance of Size: Successful Recanalization by Intravenous Thrombolysis in Acute Anterior Stroke Depends on Thrombus Length. Stroke. 2011;42:1775–7.

19. Jacquin G. J., Adel B. A. Treatment of acute ischemic stroke: from fibrinolysis to neurointervention. J Thromb Haemost. 2015;13:S290–6.

20. Wardlaw JM, Murray V, Berge E, del Zoppo GJ. Thrombolysis for acute ischaemic stroke. Cochrane Database Syst Rev. 2014;7:CD000213.

21. Abou-Chebl A. Intra-arterial Therapy for Acute Ischemic Stroke. Neurotherapeutics. 2011;8:400–13.

22. Berkhemer OA, Fransen PSS, Beumer D, van den Berg LA, Lingsma HF, Yoo AJ, et al. A Randomized Trial of Intraarterial Treatment for Acute Ischemic Stroke. N Engl J Med. 2015;372:11–20.

23. Goyal M, Demchuk AM, Menon BK, Eesa M, Rempel JL, Thornton J, et al. Randomized Assessment of Rapid Endovascular Treatment of Ischemic Stroke. N Engl J Med. 2015;372:1019–30.

24. Saver JL, Goyal M, Bonafe A, Diener H-C, Levy EI, Pereira VM, et al. Stent-Retriever Thrombectomy after Intravenous t-PA vs. t-PA Alone in Stroke. N Engl J Med. 2015;372:2285–95.

25. Campbell BCV, Mitchell PJ, Kleinig TJ, Dewey HM, Churilov L, Yassi N, et al. Endovascular Therapy for Ischemic Stroke with Perfusion-Imaging Selection. N Engl J Med. 2015;372:1009–18.

26. Jovin TG, Chamorro A, Cobo E, de Miquel MA, Molina CA, Rovira A, et al. Thrombectomy within 8 Hours after Symptom Onset in Ischemic Stroke. N Engl J Med. 2015;372:2296–306.

27. Goyal M, Menon BK, van Zwam WH, Dippel DWJ, Mitchell PJ, Demchuk AM, et al. Endovascular thrombectomy after large-vessel ischaemic stroke: a meta-analysis of individual patient data from five randomised trials. The Lancet. 2016;387:1723–31.

28. El-Koussy M, Schroth G, Brekenfeld C, Arnold M. Imaging of Acute Ischemic Stroke. Eur Neurol. 2014;72:309–16.

29. Menon BK, Campbell BCV, Levi C, Goyal M. Role of Imaging in Current Acute Ischemic Stroke Workflow for Endovascular Therapy. Stroke. 2015;46:1453–61.

30. Pace I, Zarb F. A comparison of sequential and spiral scanning techniques in brain CT. Radiol Technol. 2015;86:373–8.

31. Vu D, Lev MH. Noncontrast CT in Acute Stroke. Semin Ultrasound CT MRI. 2005;26:380–6.

32. Menon BK, Goyal M. Imaging Paradigms in Acute Ischemic Stroke: A Pragmatic Evidence-based Approach. Radiology. 2015;277:7–12.

33. von Kummer R, Meyding-Lamadé U, Forsting M, Rosin L, Rieke K, Hacke W, et al. Sensitivity and prognostic value of early CT in occlusion of the middle cerebral artery trunk. AJNR Am J Neuroradiol. 1994;15:9–15; discussion 16-18.

34. Simon JE, Kennedy J, Pexman JHW, Buchan AM. The eyes have it: conjugate eye deviation on CT scan aids in early detection of ischemic stroke. CMAJ Can Med Assoc J. 2003;168:1446–7.

35. Dzialowski I, Weber J, Doerfler A, Forsting M, Kummer RV. Brain Tissue Water Uptake after Middle Cerebral Artery Occlusion Assessed with CT. J Neuroimaging. 14:42–8.

36. Demchuk AM, Coutts SB. Alberta Stroke Program Early CT Score in Acute Stroke Triage. Neuroimaging Clin N Am. 2005;15:409–19.

37. Puetz V, Dzialowski I, Hill MD, Demchuk AM. The Alberta Stroke Program Early CT Score in Clinical Practice: What have We Learned? Int J Stroke. 2009;4:354–64.

38. Muir KW, Baird-Gunning J, Walker L, Baird T, McCormick M, Coutts SB. Can the ischemic penumbra be identified on noncontrast CT of acute stroke? Stroke. 2007;38:2485–90.

39. Parsons MW, Pepper EM, Bateman GA, Wang Y, Levi CR. Identification of the penumbra and infarct core on hyperacute noncontrast and perfusion CT. Neurology. 2007;68:730–6.

40. Hacke W, Kaste M, Fieschi C, von Kummer R, Davalos A, Meier D, et al. Randomised double-blind placebo-controlled trial of thrombolytic therapy with intravenous alteplase in acute ischaemic stroke (ECASS II). The Lancet. 1998;352:1245–51.

41. Hirano T, Yonehara T, Inatomi Y, Hashimoto Y, Uchino M. Presence of Early Ischemic Changes on Computed Tomography Depends on Severity and the Duration of Hypoperfusion. Stroke. 2005;36:2601–8.

42. Schröder J, Thomalla G. A Critical Review of Alberta Stroke Program Early CT Score for Evaluation of Acute Stroke Imaging. Front Neurol. 2017;7:245.

43. Grotta JC, Chiu D, Lu M, Patel S, Levine SR, Tilley BC, et al. Agreement and Variability in the Interpretation of Early CT Changes in Stroke Patients Qualifying for Intravenous rtPA Therapy. Stroke. 1999;30:1528–33.

44. Kalafut MA, Schriger DL, Saver JL, Starkman S. Detection of Early CT Signs of >1/3 Middle Cerebral Artery Infarctions. Stroke. 2000;31:1667–71.

45. Weir NU, Pexman JHW, Hill MD, Buchan AM, CASES investigators. How well does ASPECTS predict the outcome of acute stroke treated with IV tPA? Neurology. 2006;67:516–8.

46. Menon BK, Puetz V, Kochar P, Demchuk AM. ASPECTS and Other Neuroimaging Scores in the Triage and Prediction of Outcome in Acute Stroke Patients. Neuroimaging Clin N Am. 2011;21:407–23.

47. Goyal M, Menon BK, Coutts SB, Hill MD, Demchuk AM. Effect of Baseline CT Scan Appearance and Time to Recanalization on Clinical Outcomes in Endovascular Thrombectomy of Acute Ischemic Strokes. Stroke. 2011;42:93–7.

48. Farzin B, Fahed R, Guilbert F, Poppe AY, Daneault N, Durocher AP, et al. Early CT changes in patients admitted for thrombectomy: Intrarater and interrater agreement. Neurology. 2016;87:249–56.

49. Liebeskind DS. Collateral lessons from recent acute ischemic stroke trials. Neurol Res. 2014;36:397–402.

50. Menon BK, d'Esterre CD, Qazi EM, Almekhlafi M, Hahn L, Demchuk AM, et al. Multiphase CT Angiography: A New Tool for the Imaging Triage of Patients with Acute Ischemic Stroke. Radiology. 2015;275:510–20.

51. Winship Ian R. Cerebral Collaterals and Collateral Therapeutics for Acute Ischemic Stroke. Microcirculation. 2015;22:228–36.

52. Khandelwal N. CT perfusion in acute stroke. Indian J Radiol Imaging. 2008;18:281–6.

53. d'Esterre CD, Boesen ME, Ahn SH, Pordeli P, Najm M, Minhas P, et al. Time-Dependent Computed Tomographic Perfusion Thresholds for Patients With Acute Ischemic Stroke. Stroke. 2015;46:3390–7.

54. Wilson AT, Dey S, Evans JW, Najm M, Qiu W, Menon BK. Minds treating brains: understanding the interpretation of non-contrast CT ASPECTS in acute ischemic stroke. Expert Rev Cardiovasc Ther. 2018;143–53.

55. Krupinski EA. Current perspectives in medical image perception. Atten Percept Psychophys. 2010;72:1205–17.

56. Bal S, Bhatia R, Menon BK, Shobha N, Puetz V, Dzialowski I, et al. Time Dependence of Reliability of Noncontrast Computed Tomography in Comparison to Computed

Tomography Angiography Source Image in Acute Ischemic Stroke. Int J Stroke. 2015;10:55–60.

57. van Seeters T, Biessels GJ, Niesten JM, Schaaf IC van der, Dankbaar JW, Horsch AD, et al. Reliability of Visual Assessment of Non-Contrast CT, CT Angiography Source Images and CT Perfusion in Patients with Suspected Ischemic Stroke. PLOS ONE. 2013;8:e75615.

58. Arsava EM, Saarinen JT, Unal A, Akpinar E, Oguz KK, Topcuoglu MA. Impact of window setting optimization on accuracy of computed tomography and computed tomography angiography source image-based Alberta Stroke Program Early Computed Tomography Score. J Stroke Cerebrovasc Dis Off J Natl Stroke Assoc. 2014;23:12–6.

59. Aviv RI, Mandelcorn J, Chakraborty S, Gladstone D, Malham S, Tomlinson G, et al. Alberta Stroke Program Early CT Scoring of CT Perfusion in Early Stroke Visualization and Assessment. Am J Neuroradiol. 2007;28:1975–80.

60. Phuttharak W, Sawanyawisuth K, Sangpetngam B, Tiamkao S. CT interpretation by ASPECTS in hyperacute ischemic stroke predicting functional outcomes. Jpn J Radiol. 2013;31:701–5.

61. Pexman JHW, Hill MD, Buchan AM, Demchuk AM, Barber PA, Simon JE, et al. Hyperacute Stroke: Experience Essential When Reading Unenhanced CT Scans. Am J Neuroradiol. 2004;25:516–8.

62. Dror I. Perception is far from perfection: The role of the brain and mind in constructing realities. Behav Brain Sci. 2005;28:763–763.

63. Styles E. The Psychology of Attention. 2nd ed. Taylor & Francis; 2014. 351 p.

64. van Zoest W, Donk M. Bottom-up and Top-down Control in Visual Search. Perception. 2004;33:927–37.

65. Clark K, Cain MS, Adcock RA, Mitroff SR. Context matters: The structure of task goals affects accuracy in multiple-target visual search. Appl Ergon. 2014;45:528–33.

66. Tversky A, Kahneman D. Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. Psychol Rev. 1983;90:293–315.

67. Tversky A, Kahneman D. Judgment under Uncertainty: Heuristics and Biases. Science. 1974;185:1124–31.

68. Siegel S. The Rationality of Perception. Oxford, UK: Oxford University Press; 2016. 248 p.

69. Cain MS, Adamo SH, Mitroff SR. A taxonomy of errors in multiple-target visual search. Vis Cogn. 2013;21:899–921.

70. Nodine CF, Kundel HL. Using eye movements to study visual search and to improve tumor detection. RadioGraphics. 1987;7:1241–50.

71. Matsumoto H, Terao Y, Yugeta A, Fukuda H, Emoto M, Furubayashi T, et al. Where Do Neurologists Look When Viewing Brain CT Images? An Eye-Tracking Study Involving Stroke Cases. PLoS ONE. 2011;6:e28928.

72. Zerna C, von Kummer R, Gerber J, Engellandt K, Abramyuk A, Wojciechowski C, et al. Telemedical Brain Computed Tomography Misinterpretation by Stroke Neurologists Is Not Associated with Thrombolysis-Related Intracranial Hemorrhage. J Stroke Cerebrovasc Dis. 2015;24:1520–6.

73. Puetz V, Bodechtel U, Gerber JC, Dzialowski I, Kunz A, Wolz M, et al. Reliability of brain CT evaluation by stroke neurologists in telemedicine. Neurology. 2013;80:332–8.

74. Coutts SB, Demchuk AM, Barber PA, Hu WY, Simon JE, Buchan AM, et al. Interobserver Variation of ASPECTS in Real Time. Stroke. 2004;35:e103–5.

75. Fraser-Mackenzie PAF, Dror IE. Dynamic reasoning and time pressure: Transition from analytical operations to experiential responses. Theory Decis. 2011;71:211–25.

76. Dror I. A novel approach to minimize error in the medical domain: Cognitive neuroscientific insights into training. Med Teach. 2011;33:34–8.

77. Dey S, Evans J, Tham C, Assis Z, Teleg E, Pordeli P, et al. Abstract TP51: When Can Aspects be Read Reliably? Stroke. 2017;48:ATP51–ATP51.

78. Balcetis E, Dunning D. See What You Want to See: Motivational Influences on Visual Perception. J Pers Soc Psychol. 2006;91:612–25.

79. Krupinski EA, Siddiqui K, Siegel E, Shrestha R, Grant E, Roehrig H, et al. Influence of 8-bit vs. 11-bit digital displays on observer performance and visual search: A multi-center evaluation. J Soc Inf Disp. 2007;15:385.

80. Saunders RS, Baker JA, Delong DM, Johnson JP, Samei E. Does image quality matter? Impact of resolution and noise on mammographic task performance. Med Phys. 2007;34:3971–81.

81. Evans JW, Dey S, Eesa M, Eswaradass P, Lun R, Horn M, et al. Abstract WP55: Through Thick and Thin: Improved Aspects Grading and Dense Vessel Detection Using Simple Ncct Post-processing. Stroke. 2017;48:AWP55–AWP55.

82. Hafeez M, Qiu W, Quang H, Najm M, Wilson AT, Bobyn A, et al. Algorithm Enhanced Gray-White Matter Non-Contrast CT improves reliability of ASPECTS scoring. International Stroke Conference; 2018; Los Angeles, USA.

83. Dror I. A Hierarchy of Expert Performance. J Appl Res Mem Cogn. 2016;5:121–7.

84. Nakashima R, Watanabe C, Maeda E, Yoshikawa T, Matsuda I, Miki S, et al. The effect of expert knowledge on medical search: medical experts have specialized abilities for detecting serious lesions. Psychol Res. 2015;79:729–38.

85. Wood G, Batt J, Appelboam A, Harris A, Wilson MR. Exploring the Impact of Expertise, Clinical History, and Visual Search on Electrocardiogram Interpretation. Med Decis Making. 2014;34:75–83.

86. Nodine CF, Kundel HL, Mello-Thoms C, Weinstein SP, Orel SG, Sullivan DC, et al. How experience and training influence mammography expertise. Acad Radiol. 1999;6:575–85.

87. Coutts SB, Hill MD, Demchuk AM, Barber PA, Pexman JHW, Buchan AM. ASPECTS Reading Requires Training and Experience. Stroke. 2003;34:e179–e179.

88. Kok EM, van Geel K, van Merriënboer JJG, Robben SGF. What We Do and Do Not Know about Teaching Medical Image Interpretation. Front Psychol. 2017;8:309.

89. Kok EM, Jarodzka H, de Bruin ABH, BinAmir HAN, Robben SGF, van Merriënboer JJG. Systematic viewing in radiology: seeing more, missing less? Adv Health Sci Educ Theory Pract. 2016;21:189–205.

90. Sherbino J, Kulasegaram K, Howey E, Norman G. Ineffectiveness of cognitive forcing strategies to reduce biases in diagnostic reasoning: a controlled trial. CJEM J Can Assoc Emerg Physicians. 2014;16:34–40.

91. Schuster D, Rivera J, Sellers BC, Fiore SM, Jentsch F. Perceptual training for visual search. Ergonomics. 2013;56:1101–15.

92. Bruno MA, Walker EA, Abujudeh HH. Understanding and Confronting Our Mistakes: The Epidemiology of Error in Radiology and Strategies for Error Reduction. RadioGraphics. 2015;35:1668–76.

93. Todd PM, Gigerenzer G. Bounding rationality to the world. J Econ Psychol. 2003;24:143–65.

94. Farzin B, Fahed R, Guilbert F, Poppe AY, Daneault N, Durocher AP, et al. Early CT changes in patients admitted for thrombectomy: Intrarater and interrater agreement. Neurology. 2016;87:249–56.

95. Gilbert CD, Li W. Top-down influences on visual processing. Nat Rev Neurosci. 2013;14:350–63.

96. Padroni M, Bernardoni A, Tamborino C, Roversi G, Borrelli M, Saletti A, et al. Cerebral Blood Volume ASPECTS Is the Best Predictor of Clinical Outcome in Acute Ischemic Stroke: A Retrospective, Combined Semi-Quantitative and Quantitative Assessment. PLOS ONE. 2016;11:e0147910.

97. Hallgren KA. Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. Tutor Quant Methods Psychol. 2012;8:23–34.

98. Light RJ. Measures of response agreement for qualitative data: Some generalizations and alternatives. Psychol Bull. 1971;76:365–77.

APPENDIX A: REPORTING INTER-RATER RELIABILITY

Inter-rater reliability is an estimated measure quantifying the covariance between multiple coders' independent scores of the same subjects. This provides an estimate of the extent to which the coders perform similarly in scoring. When assessing reliability, statistics such as intraclass correlation coefficient (ICC) or the various permutations of κ are preferable to percentages of agreement, as the former measures account for chance agreement.[97]

In the study described in Chapter 3, all subjects (acute ischemic stroke patients from the PRove-IT database) were scored for ASPECTS by all three raters in a fully crossed design. Because there are more than two coders, Cohen's original conceptualization of κ is not appropriate. Therefore, Light's κ, consisting of the arithmetic mean of the linearly weighted κ for each rater pair, has been employed.[98] Linear weights, where disagreements of a greater magnitude are penalized more, were appropriate for total ASPECTS and trichotomized ASPECTS.

ICC is also an appropriate statistic to apply here. Like κ, ICC mathematically distinguishes 'true score' from measurement error, or systematic deviation from the mean. Additionally, like weighted κ, ICC encompasses the magnitude of disagreement, which is relevant for ASPECTS: two raters who respectively score a patient 9 and 8 should reflect a greater degree of reliability than if they had scored 9 and 5.

COPYRIGHT PERMISSIONS

Taylor & Francis
Taylor & Francis Group

Our Ref: KA/IERK/P18/0948

22 May 2018

Dear Alexis Wilson,

Material requested: Alexis T. Wilson, Sadanand Dey, James W. Evans, Mohamed Najm, Wu Qiu & Bijoy K. Menon (2018) Minds treating brains: understanding the interpretation of non-contrast CT ASPECTS in acute ischemic stroke, Expert Review of Cardiovascular Therapy, 16:2, 143-153

Thank you for your correspondence requesting permission to reproduce the above mentioned material from our Journal in your printed thesis entitled "A Psychological Perspective on Image Interpretation in Acute Ischemic Stroke: Factors Affecting Non-Contrast CT ASPECTS Reliability" and to be posted in the university's repository – University of Calgary

We will be pleased to grant permission on the sole condition that you acknowledge the original source of publication and insert a reference to the article on the Journals website: http://www.tandfonline.com

This is the authors original manuscript of an article published as the version of record in Expert Review of Cardiovascular Therapy © 01 Jan 2018 - https://www.tandfonline.com/10.1080/14779072.2018.1421069

This permission does not cover any third party copyrighted work which may appear in the material requested.

Please note that this license does not allow you to post our content on any third party websites or repositories.

Thank you for your interest in our Journal.

Yours sincerely

Kendyl

[redacted]

Taylor & Francis Group
an informa business

Taylor & Francis is a trading name of Informa UK Limited, registered in England under no. 1072954

2&4 Park Square, Milton Park, Abingdon, Oxfordshire OX14 4RN
Tel: +44 (0) 20 7017 6000; Fax: +44 (0) 20 7017 6336

www.tandf.co.uk

Registered in England and Wales. Registered Number: 1072954
Registered Office: 5 Howick Place, London, SW10 1WG

an informa business

May 8, 2018

Faculty of Graduate Studies
University of Calgary
2500 University Dr NW
Calgary AB T2N 1N4

Re: Co-Author Permission for Alexis Wilson's MSc Thesis

To Whom It May Concern:

I give permission for Alexis T. C. Wilson to include the entirety of the following paper that I have co-authored with her in her thesis:

Wilson AT, Dey S, Evans JW, Najm M, Qiu W, and Menon BK. Minds treating brains: Understanding the interpretation of non-contrast CT ASPECTS in acute ischemic stroke. *Expert Review of Cardiovascular Therapy* 2018;16(2):143-153.

I acknowledge that this thesis will be added to the institutional repository at the University of Calgary and the Library and Archives Canada.

Sincerely,

Dr. Bijoy K. Menon

May 8, 2018

Faculty of Graduate Studies
University of Calgary
2500 University Dr NW
Calgary AB T2N 1N4

Re: Co-Author Permission for Alexis Wilson's MSc Thesis

To Whom It May Concern:

I give permission for Alexis T. C. Wilson to include the entirety of the following paper that I have co-authored with her in her thesis:

Wilson AT, Dey S, Evans JW, Najm M, Qiu W, and Menon BK. Minds treating brains: Understanding the interpretation of non-contrast CT ASPECTS in acute ischemic stroke. *Expert Review of Cardiovascular Therapy* 2018;16(2):143-153.

I acknowledge that this thesis will be added to the institutional repository at the University of Calgary and the Library and Archives Canada.

Sincerely,

Mohamed Najm

May 8, 2018

Faculty of Graduate Studies
University of Calgary
2500 University Dr NW
Calgary AB T2N 1N4

Re: Co-Author Permission for Alexis Wilson's MSc Thesis

To Whom It May Concern:

I give permission for Alexis T. C. Wilson to include the entirety of the following paper that I have co-authored with her in her thesis:

Wilson AT, Dey S, Evans JW, Najm M, Qiu W, and Menon BK. Minds treating brains: Understanding the interpretation of non-contrast CT ASPECTS in acute ischemic stroke. *Expert Review of Cardiovascular Therapy* 2018;16(2):143-153.

I acknowledge that this thesis will be added to the institutional repository at the University of Calgary and the Library and Archives Canada.

Sincerely,

Dr. SADANAND DEY

May 8, 2018

Faculty of Graduate Studies
University of Calgary
2500 University Dr NW
Calgary AB T2N 1N4

Re: Co-Author Permission for Alexis Wilson's MSc Thesis

To Whom It May Concern:

I give permission for Alexis T. C. Wilson to include the entirety of the following paper that I have co-authored with her in her thesis:

Wilson AT, Dey S, Evans JW, Najm M, Qiu W, and Menon BK. Minds treating brains: Understanding the interpretation of non-contrast CT ASPECTS in acute ischemic stroke. *Expert Review of Cardiovascular Therapy* 2018;16(2):143-153.

I acknowledge that this thesis will be added to the institutional repository at the University of Calgary and the Library and Archives Canada.

Sincerely,

Dr J.W. EVANS.

May 8, 2018

Faculty of Graduate Studies
University of Calgary
2500 University Dr NW
Calgary AB T2N 1N4

Re: Co-Author Permission for Alexis Wilson's MSc Thesis

To Whom It May Concern:

I give permission for Alexis T. C. Wilson to include the entirety of the following paper that I have co-authored with her in her thesis:

Wilson AT, Dey S, Evans JW, Najm M, Qiu W, and Menon BK. Minds treating brains: Understanding the interpretation of non-contrast CT ASPECTS in acute ischemic stroke. *Expert Review of Cardiovascular Therapy* 2018;16(2):143-153.

I acknowledge that this thesis will be added to the institutional repository at the University of Calgary and the Library and Archives Canada.

Sincerely,

2018, May 12
Dr. Wu Qiu