12-2020

A Signal Detection Framework for Evaluating the Effects of Feedback on Stroke Recognition

Jordan D. Bailey
Western Michigan University, bailejor11@gmail.com

A Signal Detection Framework for Evaluating the Effects of Feedback on Stroke Recognition

by

Jordan D. Bailey

A dissertation submitted to the Graduate College
In partial fulfillment of the requirements
for the degree of Doctor of Philosophy
Psychology
Western Michigan University
December 2020

Doctoral Committee:

Jonathan C. Baker, Ph.D., Chair
Anthony DeFulio, Ph.D.
Jessica Frieder, Ph.D.
Allison Ilem, Ph.D.

A Signal Detection Framework for Evaluating the Effects of Feedback on Stroke Recognition

Jordan D. Bailey, Ph.D.

Western Michigan University, 2020

The impact of stroke on the lives of individuals and the healthcare system is considerable. Damage from stroke can be reduced if the treatment is administered at the appropriate time so early recognition is essential. One problem is that strokes present in a variety of ways that sometimes do not fit into the Facial drooping, Arm weakness, Speech difficulties and Time (FAST; American Heart Association, 2019) acronym. Signal detection is one way to measure decision making under conditions of uncertainty (e.g., discriminating stroke symptoms and risk factors from other symptoms, and non-risk factors). The methodology also allows us to consider motivation or bias toward a particular decision. I examined the effects of levels of feedback on performance of a random sample of participants from Amazon Mechanical Turk (MTurk). Feedback consisted of continuous feedback delivered on every trial, or asymmetric feedback that occurred only on a percentage of trials. These two levels were compared with a non-feedback reminder condition and a control condition. I solicited the opinions of medical professionals with experience in neurology or stroke as part of a social validity survey. In general, medical professionals found the procedures to be acceptable, but thought that it was most appropriate for a high-risk population.

ACKNOWLEDGEMENTS

I want to begin by thanking those who have supported me prior to and throughout the onset of this process. First and foremost, to my wife Phoebe. I am confident that I would not be in this position and would have quit long ago if not for your love and support. I could not have made it without your willingness to listen and provide feedback on talks, listen and critique research ideas (especially medical), assist with study, support us financially at times, and give up your free time. You are an incredibly intelligent, graceful, kind, and selfless individual. You have sacrificed much in the form of opportunity cost while allowing me to pursue my interests and I hope to repay you some day for doing so. I could not ask for a better partner.

To my brother, Jared. Thanks for always being willing to listen, critique, and motivate. Understand that I felt compelled to become a doctor partially because you hurtle around through the sky at hundreds of miles per hour with nothing separating you from certain death but a few inches of aluminum. I have to be able to compete with that somehow. To my parents Dan and Ruth, who initially set me down this path and have endured many a dull (to them) conversation about behavior analysis. Finally, I would like to acknowledge the contributions of our companion Verschlimmbesserung (Vera). As your name implies, you may be more effort than utility at times, but there is no question that you provide net-positive value. We love you dearly.

I want to say thank you to my mentor Dr. Jon Baker, for initially taking a chance on me. It was a serious gamble, and I hope that there is some degree of payoff. Thank you for your constant guidance not only in academic matters but in occupational, emotional, and many other life matters. I so strongly appreciate your willingness to support and take a strong interest in whatever your students are interested in, even if we propose weird projects from time to time.

To Dr. Jessica Frieder. Jessica, thank you for shaping my clinical and research skills as both a clinical supervisor and committee member. Your lab sees great national and international interest primarily due to the skill and tenacity with which you support your students. I am very lucky to have received the same level of support from you that your lab enjoys.

To Dr. Anthony DeFulio. Anthony, I have learned so much from classes you have taught and been heavily inspired from your work. My interest in behavioral economics comes primarily from your teachings and collaborations with you and your lab members. WMU is extremely lucky to have a researcher and teacher of your caliber. I am happy to call you a colleague and look forward to future possible collaboration.

To all my lab mates and other collaborators, we relied heavily upon each other and saw great benefits as a result in conferences, publications, and furthered interests. You are all great behavior analysts. To anyone interested in behavior analysis, behavioral economics, and aging. Your support, though less visible I contact and benefit significantly from. Much like those who experience great significant financial benefit from being part of an interconnected economy, I benefit from your experience and work.

Jordan D. Bailey

TABLE OF CONTENTS

Table of Contents – Continued

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

A Signal Detection Framework for Evaluating the Effects of Feedback on Stroke Recognition

Stroke, which is also known as cerebrovascular accident, is a physiological event that can result in serious neurological damage (Javedani & Zukowski, 2019). Most strokes are of the ischemic variety, where a clot blocks blood and ultimately oxygen from reaching the brain (Cassella & Jagoda, 2017), or hemorrhagic, where a ruptured blood vessel causes bleeding into the brain (Smith & Eskey, 2011).

Patients may also experience symptoms of a transient ischemic attack (TIA) that precedes or occurs in the absence of a stroke. A TIA is a "mini-stroke" where blood vessels to the brain are blocked by a clot. The stroke is transient because the clot naturally dissolves or is dislodged and most last less than one hour (Siket & Edlow, 2012). Around 15% of strokes are preceded by TIAs (Lovett et al., 2003). The prevalence of these attacks is high. It is likely that more than five million people in the United States have experienced a TIA (Johnston, 2002). TIAs may be even more difficult to detect as the symptoms are similar to strokes, but last for only a few moments, which may cause them to be ignored almost completely (Lavallée et al., 2007). Although TIAs may be a less serious event than a full stroke, the likelihood of experiencing a stroke after a TIA is high, and TIAs themselves can result in disabilities in close to 20% of cases (Coutts et al., 2012).

There are at least 700,000 individuals who experience a stroke for the first time each year in the United States (Alpert, 2011). When a patient is experiencing stroke symptoms rapid treatment is essential. From the onset of stroke symptoms to treatment, the recommended window is between 3 hours (Kothari et al., 1999) and 4.5 hours (Kamal et al., 2014). This window includes identification of a possible stroke that is occurring in the home or other

environment, medical assessment by emergency medical services (EMS) workers, physician assessment, and neuroimaging of the potential problem (Tobinet al., 2009), therefore rapid detection is also essential. The importance placed on rapidity of the response to stroke is even greater than the importance of rapidity of response to a heart attack (Rawles, 1996). Despite the importance of rapid response, patients often delay or do not seek medical attention when experiencing stroke symptoms. In one study conducted by Palomeras et al. (2008) found that only 32% of patients went to the emergency room or called an ambulance immediately after experiencing stroke symptoms. When all patients in the study were surveyed regarding the type of health issue they thought they were experiencing, only 34% thought they were experiencing a stroke indicating that there are some significant issues with recognition of the symptoms.

Causes of Delay to Treatment

Two of the causes of failure to seek treatment may be patient understanding of urgency of the situation and knowledge of stroke symptoms. Faiz et al. (2018) collected information that detailed patient attitude toward strokes and knowledge of treatments. One in seven patients did not think that time was an important factor in stroke treatment. One in 11 did not say that stroke was a serious medical issue. Similar reports are true of TIAs. As many as 36% of individuals experiencing a TIA do not seek medical attention shortly after it occurs (Johnston et al., 2003).

Delayed contact with emergency services is principally due to poor recognition of symptoms and motivation to call 911 on the part of patients and their caregivers (Fussman et al., 2010). When experiencing a stroke, patients seem to be motivated to seek medical attention quickly only if the symptoms are extremely salient. In an analysis conducted by Palomeras et al. (2008) only speech disturbance resulted in seeking medical attention in less than one hour. The first line for treatment of stroke rests with the patient themselves, or any family, friends, or

caregivers present at the time of symptoms as patients are often incapacitated with hemiparesis, aphasia, or neglect syndrome (Morgenstern et al, 2007). For these individuals it may not be enough to provide education on stroke symptomology. Instead it may be necessary to train accurate symptom discrimination so that patients and caregivers feel it is more urgent to call emergency services (Schroeder et al., 2000).

Recognition of strokes and TIAs can even be challenging even for medical professionals. EMS workers are highly trained and make medical decisions on a daily basis. Nevertheless, EMS workers have difficulty with recognizing strokes. Recognition of strokes by these professionals is extremely variable (20-93%; Lin et al., 2012). In a qualitative study of EMS workers' views on the barriers to stroke recognition, Hodell et al. (2016) found that the primary source of the problems was the wide variety of stroke symptoms that can occur.

Several tools have been designed to assist EMS workers in identifying stroke. Validated instruments include the Cincinnati Prehospital Stroke Screen (Kothari et al., 1999), the Los Angeles Prehospital Stroke Screen (Kidwell, Starkman, Eckstein, Weems, & Saver, 2000), and the Melbourne Ambulance Stroke Screen (Bray, Matrin, Cooper, Barger, Bernard, & Bladin, 2005), but these assessment tools vary in accuracy and may miss 30% of strokes that are occurring (Brandler, Sharma, Sinert, & Levine, 2014). In addition to the fact that EMS workers sometimes have difficulty recognizing stroke, and the tools are to assist with recognition are imperfect, patients must first place a call to EMS before any expert decision making becomes possible. Therefore, it is crucial that interventions impact ability to recognize stroke for patients, as well as motivate them to contact emergency services if they are uncertain about the symptoms. The following section provides a review of the literature for existing stroke recognition

interventions spanning the areas of rule-based interventions and interventions that model stroke symptoms.

Current State of Intervention Research

Effects of Rules

Rules are verbal stimuli that may specify a contingency for the behaver (Schlinger & Blakely, 1987). Interventions that are designed to increase awareness of stroke symptoms often take the form of rules. For example, if a physician tells a patient "if you see facial drooping and you do not go to the hospital within 4 hours you will die" the rule is specifying a consequence that cannot be experienced by the individual. When the consequences for rule-following cannot be experienced or are abstract it is known as augmenting. Augmenting is a form of rule-governed behavior that may be controlled by history of rule-following rather than a correspondence between a rule and the environment (Hayes et al., 1986). Augmenting is dependent upon the learning history of the individual and may cause one person to behave very differently from someone else under the effects of the same rule (Toerneke et al., 2008).

Public service campaigns have been somewhat successful with generating an acronym (rule) that many people at least recognize (facial drooping, arm weakness, speech difficulties and time [FAST]), though research suggests many people do not know what the letters of the acronym stand for (Bietzk et al., 2012). Bray et al. (2010) reported that of 171 patients and bystanders, only 12% were aware of the FAST campaign, and only 19% of these (i.e., 4) were able to recall all of the symptoms. It is highly unlikely that FAST would have the expected effect if patients cannot recall what the letters of the acronym stand for.

FAST may also have discriminative effects that result in missed stroke symptoms. Berglund et al. (2014) found that FAST may miss as many as 18% of strokes even when it was

applied by emergency services. Many symptoms that occur with stroke are not part of the acronym (e.g., altered mental status, difficulty breathing, difficulty with ambulation, falls, and generalized pain or chest pain; Kothari et al., 1997). Researchers have recognized the need for more accurate detection and have considered and created alternative interventions that can be distributed in the same manner as FAST.

One alternative intervention consists of a modified acronym: balance, eyes, face, arms, speech, time (BEFAST). Even though this intervention may be distributed easily, research is limited in supporting its efficacy with identifying stroke symptoms. In a study of 159 patients with stroke the accuracy of BEFAST in terms of area under the receiver operating characteristic curve was 0.70, while FAST managed 0.69, a difference that was not statistically significant (Pickham et al., 2018). The fact that additional discriminative stimuli added to the rule (i.e., FAST vs. BEFAST) did not result in increased accuracy may still be due to the variety of symptoms that stroke can present with. Another group of researchers have stated that the FAST campaign fails to consider strokes that impact posterior cerebral artery and affect vision (Lawlor et al., 2014). The authors proposed that the United States should consider adopting the modified "Act VFAST" campaign instead of FAST, which adds in the additional visual component. Rule-based interventions of this variety are just one type of intervention aimed at stroke recognition and there are a number of alternative interventions.

Large-scale rule-based intervention studies with an emphasis on training caregivers and family members to recognize strokes are numerous. The history of such methods extends back to the 1980s (Glanz et al., 1986). At least 27 interventions aimed at improving effectiveness of stroke recognition have been implemented in the literature from the years 1999 to 2012 (Rasura et al., 2013). Although the magnitude of these studies is often large, the impact is very small. For

example, Becker et al. (2001) conducted a study with a targeted public service announcement with a cost of $100,000. The ad campaign was run on television and in newspapers. Follow-up surveys were collected after the campaign. The authors observed an increase in the knowledge of risk factors for stroke and increase in reported likelihood that participants would report a stroke if one were occurring, but still only 50% of respondents could name a stroke symptom.

Many other studies (Morgenstern et al., 2003; Hodgson et al., 2007; Fogle et al., 2008; Kleindorfer et al., 2008; Marx et al., 2008; Williams & Noble, 2008; Bell et al., 2009; Marx et al., 2009; Muller-Norhorn et al., 2009; Tadros et al., 2009; Fogle et al., 2010; Jurkowski et al., 2010; Marx et al., 2010; Bray et al., 2011; Mikulik et al., 2011; Miyamatsu et al., 2012; Worthmann et al., 2013) have used advertisement or educational campaigns to increase public knowledge and awareness of stroke. In general, the findings are that these types of studies do not produce large effects on the behavior of participants and have limited effects on their ability to report stroke symptoms after the intervention (Rasura et al., 2013).

Modeling Stroke Symptoms and Rules

The fact that training a verbal skill in the form of a rule does not lead to the performance of seeking medical help has led researchers to consider alternative means of intervention that physically model stroke symptoms. As some behavior analysis research shows, verbal behavior (e.g., behavior that is based on rules) often does not align with performance (e.g., Baker & Wylie, 1950). In fact, Morrow et al. (2019) provided information about patient and witness failure to recognize stroke symptoms that matched the main symptoms highlighted by FAST despite being familiar with FAST. The authors interviewed 13 patients and stroke witnesses. Eight of these participants were familiar with FAST, but only two reported that the campaign led to their identification of symptoms as a stroke. Five of the participants were not able to say that

the symptoms they were experiencing were due to a stroke which appears to be an issue related to bias (Hutsell & Jacobs, 2012).

Wall et al. (2008) conducted a review of existing literature for best practice of training stroke recognition. The authors then created a 3-minute animation of stroke symptoms. This musical animation was created using the FAST acronym with the hope that the music and animation would lead to long-term memorization of the acronym. The authors reported that 76.5% of their 34-participant sample were able to recall the three primary symptoms of stroke (face, arm weakness, affected speech) at the conclusion of the intervention. Although this study took a somewhat novel approach to training participants to recognize strokes, it still relied upon an acronym that does not include *all* of the potential symptoms of strokes. In addition, there was no evidence that participants could actually detect the occurrence of one of the symptoms of in order to report to emergency services.

In a study aimed at improving stroke recognition for schoolchildren, Sharkey et al. (2016) used puppets to teach fourth, fifth, and sixth graders to identify and navigate an ongoing stroke. The puppets provided models of stroke symptoms. Their intervention was designed by a stroke specialist neurologist. The puppet show emphasized FAST, as several of the other interventions have. Posttests were administered immediately as well as at three and six months after the intervention. The authors reported that the stroke symptom subtest saw an immediate increase from 60% to 94%, and 95% at three months, but fell to just 67% after six months. It is unlikely that the effects of many of these interventions will be lasting based on the longevity of the effects observed by Becker et al. (2001). In addition, most testing comes in the form of written format rather than in the presence of real stroke symptoms.

Morgenstern et al. (2007) conducted a three-year study that consisted of 12 hours of classroom instruction that spanned three years and was targeted at helping children in identifying strokes and what they should do in the presence of one. The classroom lectures were taught by a neurologist and health teacher. The study involved role-play of signs and symptoms, as well as the specific behavior that each individual should engage in in each scenario. The study compared experimental and control groups on a pre and posttest. The researchers observed an approximately 20% increase in scores in the experimental group across domains of pathophysiology, stroke symptom knowledge, and behavioral intent to call 911 (as measured by multiple choice scenarios) in the presence of a stroke. There was no significant increase for the control group.

Stern et al. (1999) implemented a study that included 12 minutes of educational slides with audio that demonstrated risk factors, signs, and symptoms of strokes. These slides included The authors also varied whether a medical professional led discussions for the group or not. The authors targeted 657 individuals and measured verbal performance with pre and posttests. According to the authors discussion did not significantly impact scores, but the intervention did increase posttest scores of stroke risk factors, warning signs, and necessary action significantly.

Practical Issues with Intervention Research

Only four of the previously mentioned studies, (Sharkey et al., Wall et al., Morgenstern et al., and Stern et al.), exposed participants to stroke-like stimuli. The others appear to have relied upon the idea that teaching a verbal rule reliably produces performance in the presence of stroke symptoms (stimuli). In a closely related study Miller et al. (2014) warned about the use of educational campaigns to promote early detection of medical symptoms when a more intensive intervention, training discrimination of symptoms, was indicated. According Miller et al., part of

the issue with such campaigns is the focus on easily discriminable symptoms, while less salient symptoms are missed entirely. This aligns with the findings of Palomeras et al. (2008) that patients generally only seek immediate medical help when the symptoms are the most alarming (e.g., aphasia). An alternative model of training recognition consists of exposing participants to real stimuli, involves consideration of relatively uncommon events, consequences, and provides researchers information about recognition accuracy, and performance under conditions of uncertainty. Each of these variables can be analyzed in a signal detection experiment (Commons et al., 2015).

Signal Detection

Signal detection methodology allows researchers to examine decision making under a variety of conditions. Signal detection methods apply to situations that involve an operator, who must decide about whether a signal was present (stimulus present) or not given some level of noise (no stimulus present) in trials (Stanislaw & Todorov, 1999). Signal detection methods were originally created as a means of examining and improving the decisions of radar and sonar operators (Swets et al., 1961). The theory is based upon statistical decision making and assumes that an operator may discriminate between signal and noise and that each individual may possess a bias that makes them more or less likely to report that a signal is present in conditions of uncertainty (Pastore & Scheirer, 1974). Signal detection theory assumes that these are independent characteristics of the operator (Swets et al., 1961).

Common measures in a signal detection task are *d*' and β (See et al., 1997). *d*' is a measure of sensitivity as well as specificity, it measures the distance between the means of the signal distribution and noise distribution and is not affected by bias (Stanislaw & Todorov, 1999). The fact that signal detection measures are *independent* of one another (i.e., a change in

conditions may affect ability to discriminate but may not affect bias) is important because it allows researchers to analyze how both perception and cost affect decisions (Lynn & Barrett, 2014). β demonstrates the degree to which responding is biased (i.e., when uncertain, how much does a participant favor one alternative over another). In a yes/no task, a β measure of less than one demonstrates a bias toward responding yes, while a β measure greater than one represents a bias toward responding no (Stanislaw & Todorov, 1999).

One common type of task in a signal detection experiment is a yes/no task (Stanislaw & Todorov, 1999). In such a task, if the signal is present during a trial and the participant indicates that there is a signal present, it is known as a "hit". If a signal is present and the participant indicates that there is not, it is known as a "miss". A participant signaling that there is no signal when there is not is a "correct rejection", and a participant signaling that there is a signal when there is not is a "false alarm" (McNicol, 2005). For example, a radar detector who is observing their screen must discriminate between an enemy plane (signal) or various other patterns that appear on a radar screen such as flocks of birds (noise). A failure to label an enemy plane as such is a miss. Correctly doing so is a hit. Reporting a plane when there is not one is a false alarm and reporting that there is not a plane when there is not is a correct rejection. This task becomes much more difficult if there is a low number of presentations of the signal relative to noise presentation (Wolfe et al., 2005).

Behavioral detection theory came into existence as a result of recognition that a yes/no task was very similar to that of multiple schedule experimental arrangements within the field of behavior analysis. It was also recognized that signal detection performance was the result of environmental variables such as instructions, and reinforcement or punishment for performance (Nevin, 1969). If a signal was not discriminable, the experiment was a mixed schedule, meaning

discrimination did not occur because there was no clear discriminative stimulus for the participant to respond yes or no in the presence of. In the case of a mixed schedule the participant's responding is solely dependent upon the history of reinforcement for responding yes or no (McCarthy & Davison, 1981). The generalized matching equation describes exactly this type of performance (Davison & Tustin, 1978). If contingencies are discriminable, the experiment is a multiple schedule, and participants should respond "yes" in the presence of the signal and "no" in the absence with a higher degree of accuracy than responding that is controlled by the schedule of reinforcement alone (Hutsell & Jacobs, 2012). Behavioral signal detection researchers recognized that a signal detection task may be less affected by the number of signal presentations alone as it was by the schedule of reinforcement for those stimuli (McCarthy & Davison, 1979). This type of analysis of signal detection also allowed some flexibility for researchers in using single-subject designs rather than group designs where necessary (Baron & Surdy, 1990).

There are several equations that are used within behavioral detection experiments to estimate an individual or group's discriminability and sensitivity. An equation used by McCarthy and Davison (1984) describes discriminability as the following:

$$\text{Log } d = .5 \log \left(\frac{Bw * Bz}{Bx * B\,y}\right) \quad (1)$$

where B_w is the number of hits for an individual or group, B_x is the number of misses, B_y is the number of false alarms, and B_z is the number of correct rejections. This equation is meant to produce a bias free estimate of the ability of the individual or group to discriminate between signal and noise (Commons et al., 2015). A second equation of McCarthy and Davison (1984) describes the bias of an individual or group.

$$\text{Log } b = .5 \log \left(\frac{Bw * By}{Bx * Bz}\right) \quad (2)$$

The factors in this equation are the same as equation one but are rearranged. This equation is said to be independent of any stimulus discriminability. The sources of bias in this equation are the bias from obtained reinforcers, and any tendency to respond yes/no that is not explained by reinforcement. As such this equation is equivalent to the generalized matching equation of Baum (1974) (Commons et al., 2015).

Although the matching equation is tautological (i.e., it cannot be subject to confirmation, as it is always true under every condition) it is not without value. As Rachlin (1971) explained, the matching law has value because it allows researchers to explore variables that contribute to a bias in responding that may not have initially been considered. For example, a group of individuals who receive feedback for responding "yes" would be more likely to do so than a group who did not receive such feedback. This would be a bias that is changed due to an environmental variable. Examining the behavior of these individuals with the generalized matching equation, or the bias term of the behavioral signal detection would allow researchers to see how such an environmental variable shifts behavior.

Signal Detection with Medical Applications

Physicians in Signal Detection Tasks

There is a lengthy history of researchers using signal detection theory to examine and alter medical decision making. Many of the researchers in these studies have used physicians as subject matter experts to assist with discriminating signal from noise or decision-makers to determine identification skill. For example, Boutis et al. (2010) demonstrated the importance of expert detection by asking physicians to categorize radiographs of patient ankles as normal or abnormal. The authors found that for a low experience group a gradual increase in discrimination performance occurred. A more experienced group showed stable high performance, and

importantly far fewer misses than the inexperienced group. The authors were able to ensure that the content of their task was valid by using pre-screened radiographs of fractures determined to be normal or abnormal. If a diagnosis was uncertain the authors had a single staff radiologist make the determination before including the radiograph in the study. Sandor and Swensson (1978) evaluated the performance of observers on detection of small blood vessels in simulated images. They found that variation in blood vessel diameter affected accuracy significantly. The results of the study provided guidelines for appropriate levels of magnification detection of small blood vessels.

Trueblood et al. (2018) examined the ability of pathology residents and faculty to discriminate between various types and degrees of cancer cells. This study was similar to the methods of Boutis et al. (2010), but the researchers were interested in how increased productivity in the workplace affects accuracy. In general, the researchers found once again that experts more reliably discriminated cancer cells and there was no difference in bias of either group, both groups were susceptible to the effects of time pressure, which caused responding to be less cautious. These researchers validated their content by surveying a panel of expert hematopathologists. They reported that this panel annotated hundreds of these images by identifying the cell as blast or non-blast and rating how difficult identification of the cell was (1-5). An image was used in the experiment if all three judges in the panel agreed on the cell type, and it fell into categories of "easy" or "hard" for classification. All images that fell between 1.66 and 2 for average difficulty ratings of non-blasts were excluded as a means of creating clear separation between easy and difficult classifications. For blasts the authors excluded images if the difficulty rating fell between 4.66 and 4. This represented 0.25 and 0.75 quantiles for ratings of difficulty. This study provided a method of rating the samples so that trials can be balanced

according to difficulty and conclusions can be drawn about participant performance on a particular trial.

Registered Nurses in Signal Detection Tasks

Researchers have also used signal detection methods to look at the diagnostic decisions of nurses. Thompson et al. (2012) used a signal detection task and patient health records to determine if experienced nurses and nursing students were able to detect critical events. The researchers used a written task as well as a simulation that involved a computerized human patient. For the written task face validity was assessed for each scenario by a critical care nurse with 12 years of experience. The researchers found that signal detection of both groups of nurses suffered when the task involved a simulated patient mannequin rather than a pen-and-paper task. This piece of research highlights the importance of simulation that is as close to a live as possible. It is also interesting, as a life-like simulation gives a decision-maker more access to stimuli than a written task, but it may be that it also adds a great deal more noise. Thompson and Adderley (2015) looked at the ability to discriminate leg ulcers for generalist nurses and tissue specialist nurses. Their findings indicated that tissue specialist nurses were much more likely to be able to detect venous ulcers than their generalist nurse peers, but there was also significant within-group variation for both groups. The specialists did show a higher bias toward beginning with treatment for such ulcers with compression therapy. The authors validated the contents of their task by asking a panel of tissue specialist nurses to discriminate between venous ulcers versus ulcers of other causes prior to the experiment.

Patients in Signal Detection Tasks

The studies mentioned previously looked at physician and nurse ability to discriminate signal from noise. The literature also extends to patients themselves. A study conducted by

Harver et al. (2013) highlights the importance of feedback for training to improve discrimination. The authors used a signal detection task to demonstrate that children with asthma could discriminate a simulated asthma attack more accurately through ongoing feedback than a control group would be able to. The asthma attack was simulated by having participants breathe through a device that provided resistance to breathing at a level that could be varied by the researcher. The researchers found that when compared with training only, a group that had feedback delivered were much more accurate with discriminating the restrictive load to their breathing.

It is clear that signal detection and feedback can be applied to benefit those faced with medical decision. The use of behavioral detection methodology may allow researchers to observe errors that individuals make in identification of possible stroke or stroke-like events as well as determine whether they tend to over-report or under-report such symptoms. Motivation to call 911, one of the essential processes identified by Fussman et al. (2010), can be interpreted as sensitivity or bias to decide in conditions of uncertainty of symptoms. This type of responding is known as bias, which can be altered through the use of feedback or incentives as consequences in the task (Commons et al., 2015).

Signal Detection in Clinical Assessment

Just as signal detection has a high degree of utility for observing the behavior of expert and novice performers within medicine, it also has a significant amount of value as a means of clinical assessment. Clinical assessment involves discrimination of at least two categories (McFall and Treat, 1999). Within the discipline of medicine, components of signal detection are regularly used in medical imaging tasks, such as those performed by radiologists (Burgess, 2011). Typically, this involves showing participants many x-rays or other types of medical images and asking them to diagnose the presence or absence of disease. Within behavior

analysis, Lerman et al. (2010) used a signal detection framework to evaluate the behavior of individuals observing problem behavior. The authors found the framework to be useful for looking at variables that may affect the accuracy of observers when conducting an assessment. Signal detection has also been applied to situations that require constant vigilance such as luggage screening for harmful or dangerous devices (e.g., Sterchi et al., 2019). This is a task that airport agents are asked to perform all day long as part of their job role and is one more example of clinical assessment using signal detection. Each of these applications can be seen as difficult due to the high level of difficulty of discriminability that can occur in the task. For example, a behavior may be very complex and therefore difficult to discriminate. A dangerous item within luggage may be well hidden or be of an odd shape. A medical image of an illness may be exceedingly atypical. Similar issues may be present in the case of stroke, so signal detection may be an equally good candidate for the task.

Signal Detection Application to Stroke Recognition

As I have previously established, a variety of symptoms occur with stroke which makes interventions aimed at improving accuracy of detection alone challenging. Through signal detection we can determine the degree to which rule-based interventions such as FAST and other interventions change behavior. This is possible if stroke symptoms are treated as signals (S+) and other symptoms as noise (S-). We can determine the degree to which interventions improve detection of stroke symptoms (discriminability) and the effect that an intervention has on symptom recognition in conditions of uncertainty (bias). The fact that we have two independent measures allows identification of interventions that retain a high degree of accuracy, as well as shift patient behavior in the direction of assuming a stroke is occurring in conditions of

uncertainty. Before we carry out such an analysis though it is necessary to determine which stimuli qualify as S+, and which qualify as S- using subject matter experts.

Relevant Signal Detection Variables

Stimulus Presentation Ratio

In human signal detection experiments it has been demonstrated that the stimulus presentation ratio affects bias performance without affecting discriminability (Alsop et al., 1995; Johnstone & Alsop, 1996). Johnstone and Alsop (1996) demonstrated that this effect may not be unique to humans and presented two possible explanations for why stimulus presentation ratio alone would have an effect on bias. The first was that any nonreinforced trials serve to punish the behavior of responding toward the alternative that is presented most often. The second explanation was that behavior matches conditional probabilities (i.e., under the condition that the stimulus present is the less common stimulus, the probability of reinforcement for that stimulus is higher than that of the other stimulus) rather than the absolute distribution of reinforcers. Arranging a large number of S- trials for a given task may more accurately map onto the real world (e.g., stroke is an uncommon event so most symptoms will be non-stroke) but the implications are that bias will be pushed in the direction of S- rather than S+. Therefore, it makes sense to arrange trials in a manner that is nearly equal in distribution, or favors the S+ if the goal is to shift bias toward S+ responding.

Immediacy of Reinforcement/Feedback

Several researchers have made use of feedback in order to shift bias in responding as well as make discrimination more accurate. Mason and Redmon (1993) examined the impact of stimulus presentation rate and immediate versus delayed feedback on signal detection performance. This task was a quality control task that consisted of asking participants to identify

whether a computer hard disk picture was defective based on its components. The authors found that in general, accuracy was higher when feedback was immediate rather than delayed, but accuracy was high in both conditions. Harver (1994) compared the ability to detect restricted airflow for asthmatics in a feedback group versus a group that did not receive feedback. Feedback was provided immediately after the trial and consisted of a text message across a computer screen that stated whether their response was correct or not. Arranging feedback so that it is immediate is an important component of feedback, but other components of feedback are also important in signal detection task performance.

Continuous versus Asymmetric Reinforcement/Feedback

Whether feedback is continuous or asymmetric also appears to matter. Krol and El-Deredy (2012) brought up the importance of asymmetry of performance and consequences. The authors made an argument that the choices in a given decision making task may be unequal. For example, in the case of an individual who is believed to be contemplating suicide, it is more important that the decider reports it and gets help for the person while risking a false positive (i.e., the person was not suicidal and as a result your relationship suffers) than it is to decide the individual does not need help and risk a catastrophic false negative. Krol and El-Deredy (2012) made this argument for asymmetric decisions and payoffs and then conducted an experiment that involved perceptual decisions about whether a target stimulus was more visible than a non-target. A liberal payoff, where participants received more points for responding "yes" resulted in a bias toward responding "yes" relative to control and a more conservative feedback condition.

Furukawa et al. (2018) arranged a behavioral signal detection task that provided further evidence that asymmetric reinforcement or feedback impacts responding. The experiment was designed for children with and without attention deficit hyperactivity disorder. The task was

arranged so that the choice with a higher rate of reinforcement switched after 30 rewards. The authors noticed a significant bias associated with the alternative that was weighted with a higher rate of reinforcement.

As a final example of asymmetric versus continuous reinforcement, Smillie et al. (2014) conducted an auditory signal detection task that consisted of detecting whether a tone presented against a background of noise increased in volume or not. The authors examined four different schedules of feedback and found that asymmetric feedback (i.e., feedback that followed some types of stimuli more consistently than others) impacted bias but also resulted in the same level of discriminability as a continuous feedback group. When feedback occurred in every trial it was predicted that bias would be neutral. When more feedback was delivered for hits or misses relative to all other types of performance it was predicted that responding would be biased toward a "signal present" response. When more feedback was delivered for false alarms or correct rejections it was predicted that responding would be biased toward a "signal absent" response. These predictions made aligned with the observed results of the experiment.

The findings of all three studies that examined the effects of asymmetric reinforcement demonstrated what would be expected of bias in behavioral detection theory (Furukawa et al., 2018). It could be said that feedback for hits and correct rejections served as positive reinforcers, as responding "signal present" for hits, and "signal absent" for misses increased when feedback was delivered. When feedback for misses was delivered it could be said that the tendency to respond "signal absent" was punished. When feedback for false alarms occurred it could be said that the tendency to respond "signal present" was punished. The stimulus presentation ratio, immediacy of feedback, and continuity of feedback are all relevant variables that affect bias in a signal detection task. Before bias manipulation can occur researchers must act to determine the

validity of their stimuli. Furthermore, procedures, and goals for shifting the bias of participants should be examined.

Validity of a Signal Detection Task

Content Validity

In order to determine whether a task's content is valid it may be necessary to involve subject matter experts (Sireci & Geisinger, 1995). It is especially important if a task involves medical feedback and the individuals administering the task are not trained medical professionals. Haynes et al. (1995) provided a definition of content validity and stressed its importance as a part of construct validity. To these authors content validity is important because demonstrates the degree to which parts of the assessment measure the construct of interest. In the case of a signal detection task, those most important parts of the task are likely the categorization of the stimuli used in the task. Lawshe (1975) proposed a quantitative approach to content validity for jobs rather than academic tasks. In his description of a job he included single performances that were operationally defined that a worker must perform. He argued for the use of subject matter experts to identify performance metrics on a test of skill that has significant overlap with what is actually required of a performer. Lawshe (1975) proposed that the highest degree of content validity is demonstrated when there is complete agreement of subject matter experts on whether an item is essential to the job performance, or complete agreement that an item is not essential to the job performance. In the case of stroke identification and report though, it may be best to have liberal criteria for decision making. It may be more appropriate to define an S+ (i.e., the stimulus of interest) as an item that any of the subject matter experts endorse.

Social Validity

Social validity is also an extremely relevant component. In the still-early stages of the field of applied behavior analysis, Wolf (1978) provided a definition that the field had been operating without up to that point. Wolf's definition provided some clarity in the otherwise ambiguous concept of social importance. The definition included consideration of behavioral goals and how they fit within society, the acceptability of the procedures, and how much the effects actually mattered. Fawcett (1991) expanded upon Wolf's (1978) recommendations by providing specific procedures for obtaining social validity measures of a treatment acceptability, the appropriateness of the goals, and assessment of the effects. One of the specific recommendations from Fawcett (1991) was construction of social validity surveys that could be used to solicit feedback from consumers in order to determine whether the treatment was acceptable, effective, and goals were appropriate. Social validity of a signal detection task may be especially relevant as there have been none to date that have attempted to examine the decision making of stroke symptom recognition.

Decision Curve Analysis

Since it is possible to shift bias with consequences, the question becomes how much of a shift results in a benefit for potential stroke patients. One method, known as decision curve analysis (Vickers & Elkin, 2006) allows for examination of the tradeoff between hits and false alarms in terms of net benefit. Net benefit is plotted on a range of risk thresholds (Van Calster et al., 2018) which allows patients or clinicians to make decisions about an acceptable level of risk and net benefit at that level. Risk thresholds may be interpreted as a physician or patient's preference for risk (Vickers et al., 2019). Net benefit can be infinite but is expressed as a percentage. A net benefit of 0.02 means 2 out of 100 people would be identified (Van Calster et

al., 2018). Receiver operator characteristic curves are traditionally used to examine model or intervention accuracy in classification tasks, but decision analysis curves provide information about the accuracy of a model as well as the clinical utility (Rousson & Zumbrunn, 2011). Decision curve analysis provides information about clinical value of a model or intervention (Vickers & Cronin, 2010). Although decision curve analysis allows for examination of the tradeoff between clinical benefit and risk it does not provide information about financial and social costs compared with benefits (Vickers et al., 2019).

Summary of Findings

As a reminder, the issues at hand are that patients, family, and caregivers are not easily able to recognize when strokes are occurring. This may primarily be due to the variations of symptoms that stroke can present with. As a result, patients may delay seeking medical help, or avoid doing so completely, which comes with the potential for long-term disability and death. Previous intervention research is plentiful, but there are issues with the two main categories of research. The first category is rule-based interventions. These may result in behavior that does not correspond with performance of seeking medical help and may easily be forgotten. A second area of research that has not received as much research attention is on exposing participants to real stroke stimuli and teaching them to discriminate between symptoms. This approach teaches reliable discrimination but does not impact or examine the bias toward seeking help when uncertain about symptoms. Signal detection approaches may help alleviate some of the issues present in current intervention research, but a signal detection task must be valid. Interventions must also prove to be better than current interventions when findings from an experiment are extrapolated to the at-risk population.

Purpose

We used behavioral detection methodology to observe errors that individuals make in the identification of possible stroke or stroke-like events. This methodology also allowed us to determine whether participants tended to over-report or under-report such symptoms when they were uncertain. The questions that we sought to answer with this study were:

1. How do continuous feedback and asymmetric feedback affect discrimination for stroke risk-factors and symptoms relative to FAST and no feedback/no FAST?
2. How do continuous feedback and asymmetric feedback affect bias for stroke risk-factors and symptoms relative to FAST and no feedback/no FAST?
3. To what degree does experience with stroke and stroke symptoms impact discriminability and bias for stroke risk-factors and symptoms?
4. What are the net benefits of the interventions relative to the control group at the level of the experiment, and extrapolated to part of the population?
5. To what degree do medical professionals think bias should be shifted toward reporting a potential stroke is occurring when people are uncertain?
6. How acceptable is a feedback-based intervention that is administered online?

We determined the discrimination and bias scores for four groups. The first group received no feedback. For the second group, the program delivered the FAST reminder and was available at all times (FAST continuous). For the third group, the program delivered feedback for hits, misses, false alarms, and correct rejections on every trial (continuous feedback). For the final group, the program delivered asymmetric feedback (i.e., feedback that is delivered on some trials, but not all and is meant to shift bias in some direction).

Experiment One Methods

Content Validation

Prior to conducting either experiment, two medical professionals (one physician and one nurse) responded to 60 videos which included risk factors for stroke or noise risk factors. If either of the medical professionals indicated that a stroke could be occurring in the clip, the clip was considered to be a S+ (positive for possible stroke). The medical professionals were asked to provide rationale for each clip that they indicate a possible stroke is occurring in. This rationale was used as feedback for study participants. The medical professionals were also asked to rate the difficulty of identification of combinations of risk factors and symptoms on a scale of one to five. There were 32 total S+ and 28 total S- videos. The mean for discrimination difficulty for the entire set of videos was 2.7. For the signal detection portion of the study we arranged the three blocks of videos so that each block had an approximate 1:1 ratio of S+ and S-. We also arranged videos so that the average difficulty of each block was within $1/10^{th}$ of a standard deviation of the mean discrimination difficulty (2.7 +/- 0.1).

Power Analysis

The smallest effect size of interest (Lakens, Scheel, & Isager, 2018) was calculated based on the effect size from Smillie et al. (2013). This effect size was calculated based on the difference between a continuous feedback condition and an asymmetric feedback condition, which are the most similar conditions. Cohen's *d* for discriminability and bias was 0.6. Cohen's *f* was calculated by dividing Cohen's *d* by two. Cohen's *f* is the necessary effect size for power calculation for ANOVA. We used an effect size of 0.3, power of 0.8, and four groups to calculate the minimum effect size necessary using G*Power. In order to be appropriately powered a minimum of 128 participants was necessary. It was highly likely that we would see a greater

effect size for bias in this study, as we provided feedback at a more liberal ratio toward a "yes" response than Smilie et al..

Participants and Setting

Participants were recruited through Amazon Mechanical Turk (MTurk). Inclusionary criteria were that an individual was at least 18 years old, had access to a computer, and had an Amazon Mechanical Turk (MTurk) account. We excluded any participants who served as a medical professional, including physicians, advanced practice nurses, registered nurses, physician assistants, paramedics, and nurse assistants. Participants were recruited in batches of 9 participants at a time as the MTurk increases the cost substantially when recruiting exceeds 9. One-hundred-fifty participants from MTurk took part in the task.

Of these 150 participants 17 appeared to have anomalous data that included extremely low latencies, pressing the same button for at least 20 trials in a row, or both. The presence of either of these issues indicated that the participant was not following the instructions and attending to the task, suggesting they should be excluded. We decided to identify these participants on the basis of latencies if the average latency of responses was less than the first quartile of latencies minus the interquartile range (IQR) multiplied by 1.5. This is a commonly used approach to identifying and dealing with outliers after ordering the data (Barbato et al., 2011). One participant did not include their demographic information. Their data were included in the analysis, but not the demographic information. Therefore, we conducted analyses with all 150 participants (intention-to-treat), and a subset of the 150 participants (restricted sample; 133 total; 77 male and 55 female for those submitting demographics).

Participants could select one of four categories for experience with stroke: no experience with stroke, family member or friend has suffered stroke, observed a stroke in person, or

observed multiple strokes in person. See table 1 for participant demographic information. A convenience sample was appropriate given the exploratory nature of this research.

Materials

We modified 60 clips of stroke or non-stroke symptoms as part of experiment one. We searched YouTube and other public video sites for videos of stroke symptoms and symptoms that mimic strokes. We then cut them to 1-2 minutes and removed any voiceover or text that revealed the illness or symptoms. We used a list created by Okano et al. (2018) to identify and search fro common stroke mimics. Common stroke mimics were symptoms of epilepsy, neuropathy, psychiatric diagnoses, hypoglycemia, acute aortic dissection, syncope, sepsis, drug intoxication, and brain tumor. Each of these illnesses represented at least five percent of all stroke mimics examined by the authors. Additionally, video clips were included if they demonstrated stroke symptoms according to the video author. In order to include videos that were less likely to include stroke symptoms we carried out searches for videos that are not common stroke symptoms or mimics. Searches included cough, chest pain, non-limb muscle aches or joint pain, rash, shortness of breath, heart palpitations, nausea or vomiting, abdominal pain, non-stroke neurology symptoms, headaches that are not sudden and intense, and swelling. Clips were cut to 1-2 minutes in length so that symptoms were clearly displayed.

For the first experiment participants responded on a program written in Construct 2 that consisted of instructions, videos, and buttons for selecting whether a stroke was occurring or not. Construct 2 was also used to animate a physician character (Dr. Hazel) to deliver feedback. Google text-to-speech (Hazel voice) was used to deliver each of the feedback messages that went with the Dr. Hazel animation. The program was linked to a Google Sheet that automatically collected responses for each trial, latency to each response, time, date, MTurk worker

identification number, and demographic information. We provided participants with $6.50 for their participation.

Procedures

We arranged the task so that 60 approximately1-2-minute long videos were available for participants to respond to. Text describing risk factors for stroke and non-stroke risk factors preceded each video, but participants were not informed of which were and which were not stroke risk factors. There were 20 unique risk factor and non-risk factor lists that were distributed across these 60 videos. The lists of risk factors and non-risk factors were assigned to each of the first 20 videos, followed by second 20 videos, and final 20 videos. All four groups received the same 60 videos and same risk factors in the same order.

The computer program automatically randomly assigned participants to one of the treatment or control groups. Participants were presented with instructions that prompted them to "press the button labeled 'Stroke symptoms' if you believe that a stroke occurred during the video clip, do so now. Press the button labeled 'Not a stroke' if you believe no stroke has occurred, do so now. Press the button labeled 'Return to Risk Factors' if you would like to play the risk factors and video again, do so now. Participants could not advance to the next instruction until they had selected the appropriate button when prompted to do so. Once all buttons had been selected under the appropriate instructional prompt, a button to advance to the first trial appeared.

Trials began with a display of text that detailed the risk factors for the video that followed. Participants could take as long as they wished to read the risk factors and could select a button to return to the risk factors after seeing the symptoms. Buttons labeled "A stroke may be occurring" and "There is no stroke in this video" were available at all times during the video and

at the end of the video. Participants were forced to select one of the two choices at the end of the video as the task would not proceed until they had done so.

The control group received no reminders or feedback at any time. A message appeared on screen between trials that read “Please wait for the next trial”. The FAST group had reminders available consisting of the FAST acronym at all times during the study. A button labeled “remind me about strokes” fullfilled this function. The asymmetric group received feedback for hits and misses on 100% of trials and feedback for correct rejections and false alarms 10% of the time (c.f., Smillie et al., 2014). Prior to the start of the task participants read a message on the computer screen that told them “You will receive feedback for *some* correct answers and *some* incorrect ones”. We did not inform participants of the asymmetric schedule. There were no reminders available for this group. For the control group we arranged feedback for all hits, misses, correct rejections, and false alarms. There were no reminders available for this group.

Feedback

Dr. Hazel delivered feedback by appearing on the screen at the start of the task during training and at the end of trials (see figure 1). The purpose of the Dr. Hazel was to remove the requirement for participants to read through the feedback, as they were already required to do so for the risk factors. We also wanted to make feedback more interesting for participants. The computer program animated Dr. Hazel at a random location on the screen immediately after participants responded. She delivered auditory feedback.

Feedback was similar to the following for hits: “Correct! The individual in this video was experiencing a sudden, intense headache, which is a symptom of a stroke”. Feedback for misses took the form of: “You missed one here. There was weakness in the left arm, which is a symptom of a stroke.” For correct rejections participants saw messages similar to: “That’s right,

there were no stroke symptoms in this video". For false alarms participants saw: "Incorrect. There was no stroke symptom in this video". Feedback was be displayed across the screen for 10s. For any trials where feedback was not delivered the participants will simply see a blank screen for with the message "please wait for next trial" for 10s.

We programmed the application so that it provided a unique code that participants could copy and paste into the MTurk payment system along with their MTurk worker identification number. We provided payment for participation through the MTurk automated payment system.

Data Analysis

To address research questions one (how does feedback affect discriminability relative to FAST and control) and two (how does feedback affect bias relative to FAST and control) we carried out analyses for null hypothesis testing and post-hoc comparison, visual comparison of hit and false alarm rates, and visual comparison of trial-by-trial accuracy. We used a one-way analysis of variance (ANOVA) for the group scores of discriminability, and the group scores for bias. We used Tukey's test for post-hoc pairwise comparisons for each of the groups to see which scores were significantly different where a main effect was detected with ANOVA.

We plotted the false alarm rate against the hit rate for each group (see figure 2). The hit rate is the probability of responding "yes" on S+ trials, and the false alarm rate is the probability of responding "yes" on S- trials. Hit rate is calculated by dividing the number of hits by the number of hits plus misses, and false alarm rate is calculated by dividing the number of false alarms by the number of false alarms plus correct rejections (Stanislaw & Todorov, 1999).

We examined the trial-by-trial accuracy data compared to the textual description of the symptoms in a figure known as a heatmap. We conducted a visual secondary data analysis on latency averages by group. We also examined experience with strokes and discriminability and

bias scores, and the effect that group placement and experience with stroke had on discriminability and bias scores. One final analysis conducted as part of the first study was the frequency of FAST reminder button pushes and how that affected discriminability and bias scores. This analysis was carried out for the last seven participants in the FAST group.

As with all other groups, we analyzed data at the individual and group level for discriminability and bias. Assuming there were no zeros for hits, misses, correct rejections, for false alarms the highest possible bias score was 4.78, which would indicate an extreme bias toward responding "yes, a stroke is occurring" during both S+ and S- trials. Under the same assumption the maximum possible negative bias was -4.78, which would represent an extreme bias toward reporting "no, there is not stroke occurring". The same minimum and maximum are possible for the discriminability scores. In a signal detection task, it becomes an issue when there are no misses, hits, false alarms, or correct rejections in a single participant's data. The problem is one of interpretability of the data collected, as the measures become infinite (Brown & White, 2005). Therefore, we made a mathematical correction by adding 0.25 to all cells that were zero (Brown & White, 2005).

To explore research question three (to what degree does experience with stroke affect discriminability and bias) we examined mean latency of responding and experience with stroke. We examined the degree to which experience with stroke impacts mean discriminability and bias. We also examined the degree to which experience impacts mean discriminability and bias within each individual treatment group.

To address research question four (What are the net benefits of the interventions relative to the control group at the level of the experiment and extrapolated to part of the population) we carried out decision curves. We created one decision curve to examine the net benefit of the

effects of FAST, asymmetric feedback, and continuous feedback versus the control group. We created a second decision curve with corrections for the estimated effects of the interventions at the population level.

We plotted a decision curve with all interventions using the values obtained from the signal detection task and the stimulus presentation ratio. We plotted a second decision curve where we considered the base rate of stroke. Bayesian estimation was used to calculate the posterior probability of a hit and the likelihood that a stroke would occur within ten years for someone 55-59 years old. According to Westbury (2010) we can determine the probability of a that a stroke is occurring if we have a hit using Bayes theorem. In this case:

$$P(\text{Stroke} \mid \text{Hit}) = \frac{P(Hit)P(Base\ Rate\ of\ \text{Stroke})}{P(\text{Hit})\text{P}(\text{Stroke})+P(Base\ Rate\ of\ No\ Stroke)P(Miss+False\ Alarm)} \quad (3)$$

For example, assuming that we have a hit rate of 90%, and the probability of suffering a stroke is 5.9% we can estimate the posterior probability of a stroke if the test is positive:

$$= \frac{(0.90*0.06)}{(0.90*0.06)+(0.10*0.94)} = 0.36$$

We chose the 55-59 age category as it has one of the lowest probabilities of stroke (5.9%) of any group that is considered at risk of stroke based on the Framingham Stroke Risk Profile (FSRP; Wolf et al., 1991). The FSRP is a tool for profiling risk of stroke based on 10 years of data and 472 strokes. The tool provides a 10-year risk prediction for stroke based on individual risk factors that a patient has. We conducted this calculation for each of the intervention groups and the control group. The posterior probability of these values allowed us to extrapolate the results of our experiment to the probability of a single patient having a stroke rather than relying on the near 1:1 stimulus presentation ratio.

Experiment One Results

For the intention-to-treat analysis (i.e., no participant data were removed for high latency responding or failure to follow instructions), results of the one-way ANOVA to compare the effects of FAST, the control group, asymmetric feedback, and continuous feedback on discriminability did not indicate a significant effect at the $p < .05$ level for the four conditions [$F(3, 146) = 0.29$, $p = 0.6851$]. There was a difference between means of the three experimental groups: FAST ($m = 0.58$, 95% CI = 0.38 – 0.79), asymmetric feedback ($m = 0.59$, 95% CI = 0.38 – 0.81), continuous feedback ($m = 0.63$, 95% CI = 0.43 – 0.83), and the control group: ($m = 0.49$, 95% CI = 0.28 – 0.71). This difference was not enough to be statistically significant (Table 2).

The results of the one-way ANOVA for the effects of FAST, the control group, asymmetric feedback, and continuous feedback on bias with all 150 participants did indicate a significant effect at the $p < .05$ level for the four conditions [$F(3, 146) = 5.17$, $p < 0.0020$]. Post-hoc pairwise comparisons revealed that there was a statistically significant difference between the asymmetric feedback group and FAST groups ($p = 0.02$). The difference of means was 0.92 (95% CI = 0.09 – 1.74). There was a statistically significant difference between the continuous feedback group and FAST group ($p = 0.038$). The difference of means was 0.82 (95% CI = 0.03 – 1.62). There was a statistically significant difference between the asymmetric group and the control group ($p = 0.0243$). The difference of means was 0.93 (95% CI = 0.09 – 1.76). Finally, there was a statistically significant difference between the continuous feedback group and the control group ($p = 0.0432$). The difference between means was 0.83 (95% CI = 0.02 – 1.64). There was no significant difference between the control group and FAST group, or the continuous feedback and asymmetric feedback groups.

When analyzing data for the restricted sample (i.e., the high-latency data and data for those who did not adhere to the instructions were removed), results of the one-way ANOVA to compare the effects of FAST, the control group, asymmetric feedback, and continuous feedback on discriminability did not indicate a significant effect at the $p < .05$ level for the four conditions [$F(3, 130) = 0.50$, $p = 0.6851$]. Therefore, this was not different from the results with anomalous data included. There was a difference between means of the three experimental groups: FAST (m = 0.68, 95% CI = 0.48 – 0.87), asymmetric feedback (m = 0.69, 95% CI = 0.49 – 0.89), continuous feedback (m = 0.68, 95% CI = 0.49 – 0.88), and the control group: (m = 0.54, 95% CI = 0.34 – 0.74). This difference was not enough to be statistically significant (Table 3).

The results of the one-way ANOVA of the restricted sample was consistent with the effects of the intention-to-treat analysis. There was a significant effect on bias at the $p < .05$ level for the four conditions [$F(3, 130) = 7.80$, $p < 0.0001$]. Post-hoc pairwise comparisons revealed that there was a statistically significant difference between the asymmetric feedback group and FAST groups ($p = 0.0020$). The difference of means was 0.90 (95% CI = 0.26 – 1.54). There was a statistically significant difference between the continuous feedback group and FAST group ($p = 0.0083$). The difference of means was 0.78 (95% CI = 0.15 – 1.41). There was a statistically significant difference between the asymmetric group and the control group ($p = 0.0027$). The difference of means was 0.89 (95% CI = 0.24 – 1.53). Finally, there was a statistically significant difference between the continuous feedback group and the control group ($p = 0.0106$). The difference between means was 0.77 (95% CI = 0.13 – 1.39). There was no significant difference between the control group and FAST group, or the continuous feedback and asymmetric feedback groups. For the remainder of the results section we present information without the anomalous data included as the results did not differ significantly.

I present the hit rate and false alarm rate for each group in figure 2. The asymmetric feedback group had the highest hit rate (m = 0.76) and highest false alarm rate (m = 0.45). The continuous feedback had the second highest hit rate (m = 0.74) and second highest false alarm rate (m = 0.44). The control and FAST groups had equivalent hit rates (m = 0.55), but the false alarm rate of the FAST group (m = 0.33) was slightly higher than the false alarm rate of the control group (m = 0.32).

I present pooled accuracy rates by trial and group (see figure 3). This figure depicts the accuracy of each group for each trial, and a description of the symptoms within that trial. In this figure dark shading indicates a high degree of correct responses. Light shading indicates a high degree of incorrect responses. There are several trends present in this figure. The first trend is that the feedback groups tend to have darker cells on S+ trials from approximately trial 32 to trial 60 relative to the FAST and control group. There also appear to be more light-colored cells on S- trials for the feedback groups. Trial-by-trial analysis reveals that the FAST group largely produces hits on trials that include FAST symptoms.

Latencies were analyzed by group. The mean latency for each group was generated. Outliers were removed by removing any trial latencies that were greater than the third quartile of latencies plus the IQR multiplied by 1.5. These data were removed due to the probability that these extremely high latencies were not representative of the task and instead were due to pauses during the task where participants were attending to other things. The control group demonstrated the second longest average latency (m = 53s). The FAST group demonstrated the highest average latency (m = 56s). The asymmetric feedback group had the lowest level of latency (m = 42s). The continuous feedback group had the second lowest average latency (m = 49s, see figure 4).

Experience with Stroke

I examined the discriminability and bias for the four levels of participant experience (see figure 5). Participants who had a family member with a stroke had the highest level of discriminability (log *d* m = 0.86), followed by the group with no experience (log *d* m = 0.63), the group that had observed a single stroke (log *d* m = 0.49), and the group that had observed multiple strokes in person (log *d* m = 0.31). The group who had observed multiple strokes in person though had the highest level of bias (log *b* m = 1.09), followed by the group who had observed a single stroke in person (log *b* m = 0.37), the group that had a family member who suffered a stroke (log *b* m = 0.23), and finally the group that had no experience with stroke (log *b* m = -0.35).

I then examined the discriminability and bias for the four levels of participant experience by experimental group (see figure 6). Within the group of participants who had no prior experience with strokes the highest level of discriminability was the continuous feedback group (log *d* m = 0.74), followed by the FAST group (log *d* m = 0.73), the asymmetric feedback group (log *d* m = 0.68), and the control group (log *d* m = 0.42). Within the group of participants who had a family member who suffered a stroke the highest level of discriminability was the asymmetric feedback group (log *d* m = 1.01), followed by the control group (log *d* m = 0.96), the FAST group (log *d* m = 0.86), and the continuous feedback group (log *d* m = 0.68). Within the group of participants who had witnessed a stroke in person the highest level of discriminability was the continuous feedback group (log *d* m = 0.7), followed by the asymmetric feedback group (log *d* m = 0.56), the FAST group (log *d* m = 0.43), and the control group (log *d* m = 0.33). Within the group of participants who had observed multiple strokes the highest level of discriminability was the FAST group (log *d* m = 0.5), followed by the continuous feedback

group (log *d* m = 0.46), the asymmetric feedback group (log *d* m = -0.22). There were no participants assigned to the control group who had observed multiple strokes.

Within the group of participants who had no prior experience with strokes the highest level of bias was the asymmetric feedback group (log *b* m = 0.5), followed by the continuous feedback group (log *b* m = 0.09), the control group (log *b* m = -0.64), and the FAST group (log *b* m = -0.94). Within the group of participants who had a family member who suffered a stroke the highest level of bias was the asymmetric feedback group (log *b* m = 0.57), followed by the continuous feedback group (log *b* m = 0.44), the control group (log *b* m = 0.41), and the FAST group (log *b* m = -0.73). Within the group of participants who had witnessed a stroke in person the highest level of bias was the FAST group (log *b* m = 1.37), followed by the continuous feedback group (log *b* m = 0.88), the asymmetric feedback group (log *b* m = 0.36), and the control group (log *b* m = -0.32). Within the group of participants who had observed multiple strokes the highest level of bias was the continuous feedback group (log *b* m = 1.59), followed by the asymmetric feedback group (log *b* m = 1.51), and the FAST group (log *b* m = 0.63). There were no participants assigned to the control group who had observed multiple strokes.

For the last seven participants in the FAST group I sampled how often participants were pressing the FAST button as a reminder. I provided the FAST reminder for all participants before the trials began, but within this sample of seven participants some chose to view the reminder more often than others. One participant viewed the reminder 13 times. Another participant viewed the reminder seven times. Two participants viewed the reminder once beyond the initial presentation. Three of the participants did not view the FAST reminder beyond the initial presentation.

Decision Curve Analysis

Within the context of the experiment (figure 8) the highest initial level of net benefit came from decision to treat all symptoms as stroke (52% at the lowest level of risk). The net benefit of the feedback groups was nearly identical (39% at the lowest level of risk). The net benefit of the FAST group was slightly higher (28% at the lowest level of risk) than the control group (26% at the lowest level of risk). The net benefit of the feedback groups exceeded net benefit of treating all symptoms as stroke within the context of the experiment at a risk level of 35%. The net benefit of the FAST and control group exceeded treating all symptoms as stroke at a risk level of approximately 45%. The FAST and control group curves did not intersect the feedback groups at any of the acceptable risk levels examined (0 - 75%).

When extrapolating to the population who are at 5.9% 10-year risk of stroke the highest net benefit was nearly the same for the feedback groups (9.9% at the lowest level of risk). The FAST group had a lower level of net benefit at all acceptable levels of risk (8.6% at the lowest level of risk). The FAST and control group curves did not intersect the feedback groups at any of the risk levels examined (0 – 20%). The risk levels examined were lower for the population curve as the net benefit of intervention is excessively negative at high levels of risk.

Experiment Two Methods

Participants and Setting

For experiment two I solicited responses from local area medical professionals. I sent e-mails to a local area hospital group and group of professionals focused on healthy aging. For this experiment any medical professional who had experience with strokes could respond and I was interested in receiving responses from 5-10 medical professionals. Eight medical professionals responded to the social validity survey. The eight respondents had varying degrees of experience

with stroke. Two of the respondents were medical doctors with experience in gerontology and stroke care. Another respondent was a nurse practitioner with experience in stroke care. Two of the respondents were registered nurses with experience in stroke care, telemetry, and rehabilitation following stroke. One respondent was a paramedic with experience treating and transporting stroke patients. Two respondents reported running home care agencies that provided care for individuals who have had strokes or are at risk of stroke.

Materials

I used Qualtrics to distribute our survey to participants. Participants could win one of three $50 Amazon gift cards for participating in the social validity survey.

Social Validation

I distributed a social validity survey so that the validity of the procedures could be examined. Of particular interest was be the social appropriateness of the procedures, and social importance of the effects (Fawcett, 1991). I examined the importance of the effects of the intervention at the levels of immediate effects. In this case it was of interest whether medical professionals believe that an increase in the discrimination skills and bias toward acting in the presence of a perceived stroke using the current experimental arrangement was socially important. I compared this in a question about the efficacy of FAST according to medical professionals.

I also examined how socially appropriate the procedures were. Specifically, I was interested in how much the medical professionals thought it would be appropriate to provide a 1-2-hour long feedback-based task for individual at risk of stroke, how much the medical professionals liked the procedures, and how much the medical professionals liked the feedback

delivery. Finally, I asked the medical professionals how representative of stroke the S+ clips were, and how likely it was that S- clips contained stroke.

Procedures

I e-mailed a Qualtrics survey link to participants. I sought to collect responses in an anonymous fashion and export them as comma separated values files. There was no identifying information associated with data collected. I provided three gift cards valued at $50 each that were randomly distributed for participation in the survey.

Data Analysis

To address research questions five (to what degree should bias be shifted toward reporting a potential stoke is occurring in conditions of uncertainty) and six (how acceptable is a feedback-based intervention that is administered online) I examined the mean, range, and standard deviation of responses to each question on the survey. I also provided a descriptive analysis of responses.

Experiment Two Results

For a summary of responses to our social validity survey see table 6. In response to research question five, respondents generally indicated that stroke should be assumed in conditions of uncertainty if risk of stroke is relatively high (10-year risk of > 10%). Most respondents indicated that the feedback intervention would be most appropriate for a high-risk group, but also thought that FAST was generally effective for getting patients to seek help when they are experiencing stroke symptoms. I asked respondents to provide information about a goal for hits and false alarms that an intervention should be able to produce. The mean for the acceptable percentage of hits was 67.57. The mean for the acceptable percentage of false alarms was 50.

I asked about any conditions under which a bias toward seeking medical attention in conditions of uncertainty would not be indicated. Two of the respondents stated that this should never be the case, one respondent stated "if someone is terminally ill and admitted to a hospice service it may not be necessary to seek medical attention; especially if not interested in aggressive treatment measures or treatments aimed at extension of life; if they have a signed DNR, for example". Another respondent echoed that sentiment: "when it is not in line with the patient's goals of care". A final respondent stated that if the patient was on a palliative care plan and was already receiving antiplatelet or anticoagulants it would not be indicated. No other reasons were provided.

In response to research question six, I asked participants how likely they would be to recommend a feedback-based training to participants of three different risk categories. Participants generally indicated that they would recommend such an intervention, but only to the highest risk category. I also asked respondents to assume that the effects of feedback would not be long-lasting and then asked how often it would be feasible to provide a booster or reminder intervention. One of the respondents stated that it may be difficult to administer due to the advanced age of some of the individuals at-risk and their experience with technology. One respondent stated that patients may start, but not complete the training and have issues with comprehension or short-term memory. A third respondent stated that the training may be most beneficial for families of the patient instead of the patient.

Discussion

Rapid response in the form of seeking medical attention is crucial, but a large proportion of stroke patients do not seek help within the essential 3 to 4.5-hour window. The most common issue with failure to seek medical attention is failure to recognize the symptoms due to the varied

symptoms that stroke can present with. Intervention research aimed at improving recognition (discriminability in signal detection terms) is abundant but falls short of improving tendency to seek medical for several reasons. First, intervention research does not cover all of the symptoms of stroke. Second, intervention research results in learning rules that do not cause patients to seek help or does not consider changing what someone might do if they are uncertain about symptoms (bias). Finally, most intervention research does not expose participants to real stroke and non-stroke stimuli so that their behavior can come under control of appropriate stimulus conditions. In this study I sought to evaluate the effects of two feedback interventions in comparison with one of the most common rule-based interventions (FAST) and a no-intervention control group. Therefore, in the first study I examined the effects using two measures from signal detection research known as discriminability and bias. After conducting study one I sought to determine whether our intervention was acceptable and to what degree future research should set as goals for discriminability and bias. I determined the acceptability and goals by soliciting the feedback of medical professionals in a social validity survey.

Discriminability Findings and Implications

One of the most interesting findings of the first study was that discriminability of the feedback interventions were not significantly better than the FAST or control groups. It is not entirely surprising that the level of discriminability in the FAST group and both feedback groups was not significantly different, as all three of these interventions should have discriminative effects based on their design. What was most surprising is that none of the experimental groups had significantly higher level of discriminability than the control group. This is most likely due to the fact that stroke discrimination is difficult due to the enormous variety of presenting symptoms with stroke, and stroke mimic symptoms that can occur in cases of non-stroke. This

leads us to believe that conditions of uncertainty are prevalent with stroke, so interventions that attempt to shift bias toward seeking help in conditions of uncertainty (i.e., interventions that impact bias) may be the best option as opposed to interventions that impact just discriminability. This could be taken as far as offering incentives for going to the hospital when certain risk factors are present, and there is uncertainty about symptoms. For example, if an older adult has certain risk factors such as history of heart disease and smoking, and they seek medical care in the presence of symptom uncertainty, they receive a monetary voucher. This is an approach that is effective with increasing other medical adherence behaviors such as medication adherence (DeFulio & Silverman, 2012).

It is important to recall that in our study the details of FAST were available at all times as a reminder for individuals in that group. As I have pointed out previously, people often forget what FAST stands for, so in application FAST discriminability may be even lower. I also pointed out that FAST has a strong discriminatory effect for the three symptoms that it details. FAST may actually contribute to a low level of bias as symptoms are specified in a rule statement, which would potentially lead to excluding other stroke symptoms. This is why longer acronyms exist (e.g., BEFAST; Pickham et al., 2018). There is some evidence in the scores of participants who selected the FAST button during the task, that FAST increased discriminability, but decreased bias (see figure 7). Although there was not enough data for statistical analysis the scores for the seven participants that I did collect appeared to be drastically different. Still, it is quite concerning that the FAST group did not have significantly higher levels of discriminability than the control group given the fact that FAST is mainly meant to improve discriminability of symptoms.

Given the fact that FAST is meant to have a discriminative effect it makes sense that correct rejections contributed to the discriminability score of FAST more than hits did. This is evident in figure 2, as the hit rate of the FAST and control groups is much lower than either of the feedback groups, but the false-alarm rate is also lower. This means that FAST will yield a high number of patients who do not seek medical care in the presence of non-stroke symptoms, and a substantial number of patients who have non-FAST stroke symptoms will also forego medical care. This discriminative effect will also result in a lower number of patients seeking medical care who have stroke symptoms, relative to the feedback groups. Practically speaking this is likely to result in a lower level of clinical benefit.

Discriminability and Experience with Stroke

Based on visual analysis of figure 5, it seems clear that those who have greater experience with stroke actually respond with lower levels of discriminability than those who are less experienced. This is likely to be, once again, due to the fact that stroke can be so difficult to discriminate due to the variety of symptoms it can present with. It is also likely that the effects of FAST are greater for those who have less direct experience with stroke, as it may be the only thing that they can rely on in terms of discriminability.

Bias Findings and Implications

It is clear that both of the feedback interventions increase participant bias toward reporting that a stroke is occurring.

Given the findings of Smilie et al. (2013) we could assume that the asymmetric feedback group would benefit from higher levels of bias than the continuous feedback group. Although the bias for the asymmetric group was slightly higher, it was not enough to be statistically significant. This could be due to several factors. The first factor could be the number of trials. It

is possible that more trials than 60 are required for an asymmetric effect to begin to emerge. The second could be that there are prior learning histories in place for all participants with respect to stroke. This could be enough to nullify the effects that asymmetrically arranged feedback has on bias. The task in which Smilie et al. (2013) observed asymmetric effects was a novel auditory detection task. In studies such as Johnstone and Alsop (2000) asymmetric effects were observed using money, reinforcers that were assumedly higher magnitude.

Bias and Experience with Stroke

An important finding is that the participants who are more experienced with stroke are seemingly more sensitive (i.e., they have higher bias) toward reporting that a stroke might be occurring. A group that had no experience with strokes was far less likely to be biased toward reporting strokes occurring. It is likely that observing strokes in person or having a family member who has had a stroke results in a bias toward reporting that a stroke is occurring rather than greater discriminability. This is most likely, as mentioned previously, due to the large variety of symptoms that stroke may present with.

Clinical Benefit of Interventions

There is a strong focus on accuracy in health interventions and diagnostic testing, but the accuracy of a test or intervention is only one facet and it is also important to consider clinical benefit (Bossuyt et al., 2012). One type of clinical benefit to consider is that which is presented in decision curves (see figures 8 and 9). Within the experiment and in the extrapolation to population both of the feedback groups yielded higher net benefit than the control group or FAST group at all calculated levels of acceptable risk. This means that the clinical net benefit of both feedback interventions is greater than FAST for the population regardless of the level of risk that a patient may be at. Even though decision curve analysis assumes that false positive are

a harm (Martin et al., 2017) both feedback groups had higher net benefit scores than the FAST group. The important takeaway is that even when false alarms are penalized the clinical benefit of the feedback intervention was higher than the FAST intervention.

Determining Risk for Stroke

Respondents to the social validity survey made it clear that a feedback-based intervention would be most appropriate for individuals with a high risk of stroke. Despite the decision curve revealing that the feedback interventions are better for all levels of risk, I also believe that this intervention would be most appropriate for a high-risk group. The primary reason is that decision curves do not take economic costs into consideration. Therefore, it is important to gauge who is at risk for stroke. Risk for stroke can be determined by using the FSRP (Wolf et al., 1991). In fact, an updated risk profile based on the FSRP was released in 2017 (Dufouil et al., 2017). Machine learning approaches can also help to identify those who are at risk of stroke. Cheon et al. (2019) used population medical records and machine learning to detect who is most likely to be at risk of stroke in the Korean population. The model achieved a hit rate of 64% and a false-alarm rate of just 14%. The authors pointed out that using such a model would allow for deploying resources most effectively. Our feedback-based video intervention could be available for individuals who are determined to be at risk based on one of these methods.

Social Validity of Methods

The social validity survey revealed that medical practitioner respondents believed that those who are at high to moderate risk of stroke should assume that a stroke is occurring when they cannot recognize their symptoms. Perhaps unsurprisingly, respondents were not very likely to recommend a 1-2-hour video-based training program. Some of the comments revealed low belief in the technological abilities of the older adult population who is most at-risk, as well as

low belief in the likelihood that those at-risk would sit through a 1-2-hour training at will. This is one area where FAST is appreciable. FAST is a simple intervention that does not require a significant amount of resources or time of the individual to consume it. With that said, the ease and acceptability are much less appreciable if it does not have the intended effect. The survey revealed that practitioners may believe that FAST is more effective that it truly is.

One final area of interest from the survey was that respondents reported strong belief that the S+ clip contained a stroke (m = 8.9, sd = 0.93), but did not appear to agree that the S- clip was not a stroke (m = 3.43, sd = 2.2). One respondent ranked this clip as high as 7, indicating somewhat strong belief that a stroke WAS occurring. This is further evidence that it is difficult to discriminate stroke symptoms, and we must rely upon biological diagnostic testing to have any degree of certainty. According to respondents in this survey, a high level of false alarms is actually acceptable following intervention. Respondents thought that 43% false alarms in a task like ours was an acceptable rate.

Limitations

I have no information about how the effects of the intervention persist over time. FAST also is not lasting though. This may be for several reasons. As Bray et al. (2010) identified, participants may not remember what the acronym stands for. It may also be the case that if patients make their decision to seek medical treatment, and those symptoms turn out to be a false positive, they may abandon the rule in the future. This is an effect that is related to rule-governed behavior, and is known as tracking (Hayes et al., 1986).

Another limitation was that I could not effectively control the stimulus properties. The S+ and S- videos were videos that were freely available from the internet. It is quite possible that the reason for a lack of statistical significance with discriminability was due to the fact that I could

not control exercise enough control over stimulus properties. In the future researchers should attempt to create and use videos where stimulus dimensions can be adequately controlled. For example, it may be possible to control the degree of a facial droop to make it more or less easy to detect. Videos could also be presented with a single stroke symptom and those with multiple symptoms.

Future Directions

A cost-effectiveness analysis should be carried out. Cost-effectiveness analyses are used in healthcare or other areas where it is difficult to monetize the impact of the proposed interventions (Boardman et al., 2018). We can estimate the results of such an analysis using the joint probability of events to estimate the impact of the intervention on a population level. The probability of individuals in the U.S. over 65 having a stroke in a given year is 1.2% without considering any risk-factors. The joint probability of hits for the asymmetric feedback intervention and the one-year probability of stroke is 0.9%. The joint probability of false alarms for the asymmetric feedback intervention and the probability of no stroke is 44%. The joint probability of hits for the FAST group and the one-year probability of stroke is 0.7%. The joint probability of false alarms for the FAST group and the one-year probability of no stroke is 33%.

The lifetime cost for an individual following stroke on average in the United States is $140,048 (Johnson et al., 2016). Assuming that a false alarm would lead to an emergency room visit, the asymmetric intervention would lead to approximately 11% more emergency room visits. The cost of a single visit to the ER in 2020 dollars is approximately $563.31 (Adjusted from 1998 dollars; Bamezi et al., 2005). There are approximately 596,250 people over the age of 65 who have strokes in the U.S. over a one-year period. Given the 33% false alarm rate that I estimated for the FAST intervention, FAST would produce approximately 196,762 false alarms

over a one-year period if the intervention reached every person. This would result in a cost of $110,838,283. Given the 44% false-alarm rate that I estimated for the asymmetric intervention, the asymmetric intervention would produce 262,350 false alarms. This would result in a cost of $147,784,378.

It is also necessary to consider how much is saved by correctly identifying strokes. Given the 0.7% hit rate I estimated for the FAST intervention, FAST would produce 4174 hits in a given year. This would be a savings of $617,952,352 when considering the lifetime cost of stroke. Given the 0.9% hit rate I estimated for the asymmetric intervention, the asymmetric intervention would produce 5366 hits in a given year. This would be a savings of $751,497,568. Assuming that the intervention would prevent the $140,048 average lifetime cost of stroke patient care, the total amount saved by the FAST intervention would be $507,114,069. The total amount saved by the asymmetric intervention would be $603,713,190. Although the feedback intervention appears to save more than $100,000,000 more than FAST neither of those figures include the cost of intervention. Although there were no reports of total costs of FAST, one estimate of a three-month portion of the FAST campaign in the United Kingdom was approximately $923,000 (Flynn et al., 2014). It is also not possible to say that preventing stroke in one instance will prevent it for a lifetime for each individual. Therefore, future researchers may wish to examine cost-effectiveness in greater detail.

I also suggest considering alternative interventions that impact bias or discriminability and bias together. The social validity survey pointed to low levels of potential acceptability of a 1-2-hour feedback-based video intervention. Therefore, although the feedback-based intervention did have the intended effect of increasing bias, other candidates such as incentives should be explored.

Conclusion

In summary, both of the feedback interventions were better than FAST at getting participants to recognize the symptoms of stroke or assume that stroke was occurring in conditions of uncertainty. FAST may have some discriminative effects, but these effects appear to result in correct rejections more than hits. FAST may also decrease bias toward reporting a stroke in conditions of uncertainty. Conditions of uncertainty appear to be relatively common for patients experiencing strokes. I also found that medical professionals believed that the feedback intervention was appropriate to apply to a group of high-risk patients. I believe that a feedback-based intervention such as ours, or other intervention aimed at increasing bias could be used to great effect if coupled with a method for detecting individuals who are at high-risk, such as a machine learning approach.

References

Allen, K. D., Loiben, T., Allen, S. J., & Stanley, R. T. (1992). Dentist-implemented contingent escape for management of disruptive child behavior. *Journal of applied behavior analysis*, *25*(3), 629–636. doi:10.1901/jaba.1992.25-629

Alsop, B., Rowley, R., & Fon, C. (1995). Human symbolic matching-to-sample performance: Effects of reinforcer and sample-stimulus probabilities. *Journal of the experimental analysis of behavior*, *63*(1), 53–70. doi:10.1901/jeab.1995.63-53

Baker, K. E., & Wylie, R. C. (1950). Transfer of verbal training to a motor task. *Journal of Experimental Psychology, 40*(5), 632-638. doi:http://dx.doi.org.libproxy.library.wmich.edu/10.1037/h0057374

Balci, K., Utku, U., Asil, T., & Celik, Y. (2011). Ischemic stroke in young adults. *The Neurologist*, *17*(1), 16–20. doi: 10.1097/nrl.0b013e3181f954a7

Bamezai, A., Melnick, G., & Nawathe, A. (2005). The Cost of an Emergency Department Visit and Its Relationship to Emergency Department Volume. *Annals of Emergency Medicine*, *45*(5), 483–490. https://doi.org/10.1016/j.annemergmed.2004.08.029

Barbato, G., Barini, E. M., Genta, G., & Levi, R. (2011). Features and performance of some outlier detection methods. *Journal of Applied Statistics*, *38*(10), 2133–2149. doi: 10.1080/02664763.2010.545119

Baum, W. M. (1974). On two types of deviation from the matching law: bias and undermatching1. *Journal of the Experimental Analysis of Behavior*, *22*(1), 231–242. doi: 10.1901/jeab.1974.22-231

Bell, M., Lommel, T., Fischer, J. G., Lee, J. S., Reddy, S., & Johnson, M. A. (2009). Improved recognition of heart attack and stroke symptoms after a community-based intervention for older adults, Georgia, 2006-2007. *Preventing chronic disease*, *6*(2), A41.

Berglund, A., Svensson, L., Wahlgren, N., & Euler, M. V. (2014). Face Arm Speech Time Test Use in the Prehospital Setting, Better in the Ambulance than in the Emergency Medical Communication Center. *Cerebrovascular Diseases*, *37*(3), 212–216. doi: 10.1159/000358116

Bossuyt, P. M., Reitsma, J. B., Linnet, K., & Moons, K. G. (2012). Beyond Diagnostic Accuracy: The Clinical Utility of Diagnostic Tests. *Clinical Chemistry*, *58*(12), 1636–1643. doi: 10.1373/clinchem.2012.182576

Boutis, K., Pecaric, M., Seeto, B., & Pusic, M. (2010). Using signal detection theory to model changes in serial learning of radiological image interpretation. *Advances in Health Sciences Education*, *15*(5), 647–658. doi: 10.1007/s10459-010-9225-8

Brandler, E. S., Sharma, M., Sinert, R. H., & Levine, S. R. (2014). Prehospital stroke scales in urban environments: A systematic review. *Neurology*, *82*(24), 2241–2249. doi: 10.1212/wnl.0000000000000523

Bray, J. E., Martin, J., Cooper, G., Barger, B., Bernard, S., & Bladin, C. (2005). Paramedic identification of stroke: community validation of the Melbourne Ambulance Stroke Screen. *Cerebrovascular Diseases*, *20*(1), 28–33. doi: 10.1159/000086201

Bray, J. E., Mosley, I., Bailey, M., Barger, B., & Bladin, C. (2011). Stroke public awareness campaigns have increased ambulance dispatches for stroke in Melbourne, Australia. *Stroke*, *42*(8), 2154–2157. doi: 10.1161/strokeaha.110.612036

Briard, J. N., Zewude, R. T., Kate, M. P., Rowe, B. H., Buck, B., Butcher, K., & Gioia, L. C. (2018). Stroke Mimics Transported by Emergency Medical Services to a Comprehensive Stroke Center: The Magnitude of the Problem. *Journal of Stroke and Cerebrovascular Diseases*, *27*(10), 2738–2745. https://doi.org/10.1016/j.jstrokecerebrovasdis.2018.05.046

Burgess, A. E. (2011). Visual Perception Studies and Observer Models in Medical Imaging. *Seminars in Nuclear Medicine*, *41*(6), 419–436. https://doi.org/10.1053/j.semnuclmed.2011.06.005

Cassella, C. R., & Jagoda, A. (2017). Ischemic stroke. *Emergency Medicine Clinics of North America*, *35*(4), 911–930. doi: 10.1016/j.emc.2017.07.007

Commons, M. L., Nevin, J. A., & Davison, M. C. (2015). *Signal detection: mechanisms, models, and applications*. New York: Psychology Press.

Coutts, S. B., Modi, J., Patel, S. K., Aram, H., Demchuk, A. M., Goyal, M., & Hill, M. D. (2012). What Causes Disability After Transient Ischemic Attack and Minor Stroke? *Stroke*, *43*(11), 3018–3022. doi: 10.1161/strokeaha.112.665141

Critchfield, T. S., & Reed, D. D. (2016). Does hearing about cancer influence stimulus control? An exploratory study of verbal modulation of stimulus generalization. *The Analysis of verbal behavior*, *32*(1), 46–59. doi:10.1007/s40616-016-0055-5

Davison, M. C., & Tustin, R. D. (1978). The relation between the generalized matching law and signal-detection theory. *Journal of the Experimental Analysis of Behavior*, *29*(2), 331–336. doi: 10.1901/jeab.1978.29-331

DeFulio, A., & Silverman, K. (2012). The use of incentives to reinforce medication adherence. *Preventive medicine*, *55 Suppl*(Suppl), S86–S94. https://doi.org/10.1016/j.ypmed.2012.04.017

Dufouil, C., Beiser, A., McLure, L. A., Wolf, P. A., Tzourio, C., Howard, V. J., Westwood, A. J., Himali, J. J., Sullivan, L., Aparicio, H. J., Kelly-Hayes, M., Ritchie, K., Kase, C. S., Pikula, A., Romero, J. R., D'Agostino, R. B., Samieri, C., Vasan, R. S., Chêne, G., Howard, G., … Seshadri, S. (2017). Revised Framingham Stroke Risk Profile to Reflect Temporal Trends. *Circulation*, *135*(12), 1145–1159. https://doi.org/10.1161/CIRCULATIONAHA.115.021275

Flynn, D., Ford, G. A., Rodgers, H., Price, C., Steen, N., & Thomson, R. G. (2014). A time series evaluation of the FAST National Stroke Awareness Campaign in England. *PloS one*, *9*(8), e104289. https://doi.org/10.1371/journal.pone.0104289

Fogle, C. C., Oser, C. S., Mcnamara, M. J., Helgerson, S. D., Gohdes, D., & Harwell, T. S. (2010). Impact of media on community awareness of stroke warning signs: a comparison study. *Journal of Stroke and Cerebrovascular Diseases*, *19*(5), 370–375. doi: 10.1016/j.jstrokecerebrovasdis.2009.06.007

Fogle, C. C., Oser, C. S., Troutman, T. P., Mcnamara, M., Williamson, A. P., Keller, M., … Harwell, T. S. (2008). Public education strategies to increase awareness of stroke warning signs and the need to call 911. *Journal of Public Health Management and Practice*, *14*(3). doi: 10.1097/01.phh.0000316496.78282.47

Furukawa, E., Alsop, B., Caparelli-Dáquer, E. M., Casella, E. B., da Costa, R.,Quimas Molina, Queiroz, P. d. M., . . . Tripp, G. (2018). Behavioral adjustment to asymmetric reward availability among children with and without adhd: Effects of past and current reinforcement contingencies. *ADHD Attention Deficit and Hyperactivity Disorders,* doi:http://dx.doi.org.libproxy.library.wmich.edu/10.1007/s12402-018-0265-x

Fussman, C., Rafferty, A. P., Lyon-Callo, S., Morgenstern, L. B., & Reeves, M. J. (2010). Lack of association between stroke symptom knowledge and intent to call 911. *Stroke*, *41*(7), 1501–1507. doi: 10.1161/strokeaha.110.578195

Glanz, K., Marger, S. M., & Meehan, E. F. (1986). Evaluation of a peer educator stroke education program for the elderly. *Health Education Research, 1*(2), 121-130. doi:http://dx.doi.org.libproxy.library.wmich.edu/10.1093/her/1.2.121

Harver, A. (1994). Effects of feedback on the ability of asthmatic subjects to detect increases in the flow-resistive component to breathing. *Health Psychology, 13*(1), 52-62. doi:http://dx.doi.org/10.1037/0278-6133.13.1.52

Hayes, S. C., Brownstein, A. J., Zettle, R. D., Rosenfarb, I., & Korn, Z. (1986). Rule-governed behavior and sensitivity to changing consequences of responding. *Journal of the experimental analysis of behavior*, *45*(3), 237–256. https://doi.org/10.1901/jeab.1986.45-237

Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment, 7*(3), 238-247. doi:http://dx.doi.org.libproxy.library.wmich.edu/10.1037/1040-3590.7.3.238

Hodgson, C., Lindsay, P., & Rubini, F. (2007). Can mass media influence emergency department visits for stroke? *Stroke*, *38*(7), 2115–2122. doi: 10.1161/strokeaha.107.484071

Hutsell, B. A., & Jacobs, E. A. (2012). Rapid acquisition of bias in signal detection: dynamics of effective reinforcement allocation. *Journal of the experimental analysis of behavior*, *97*(1), 29–49. doi:10.1901/jeab.2012.97-29

Javedani, P. P., & Zukowski, M. (2019). Cerebrovascular accident in a pediatric patient presenting with influenza. *The Journal of Emergency Medicine*, *57*(1). doi: 10.1016/j.jemermed.2019.03.015

Johnson, B., Bonafede, M., & Watson, C. (2016). Short- and longer-term health-care resource utilization and costs associated with acute ischemic stroke. *ClinicoEconomics and Outcomes Research*, 53. https://doi.org/10.2147/ceor.s95662

Johnston, S. C. (2002). Transient ischemic attack. *New England Journal of Medicine*, *347*(21), 1687–1692. doi: 10.1056/nejmcp020891

Johnston, S. C., Fayad, P. B., Gorelick, P. B., Hanley, D. F., Shwayder, P., Husen, D. V., & Weiskopf, T. (2003). Prevalence and knowledge of transient ischemic attack among US adults. *Neurology*, *60*(9), 1429–1434. doi: 10.1212/01.wnl.0000063309.41867.0f

Johnstone, V., & Alsop, B. (1996). Human signal-detection performance: effects of signal presentation probabilities and reinforcer distributions. *Journal of the experimental analysis of behavior*, *66*(2), 243–263. doi:10.1901/jeab.1996.66-243

Johnstone, V., & Alsop, B. (2000). Reinforcer control and human signal-detection performance. *Journal of the experimental analysis of behavior*, *73*(3), 275–290. doi:10.1901/jeab.2000.73-275

Jurkowski, J. M., Maniccia, D. M., Spicer, D. A., & Dennison, B. A. (2010). Impact of a multimedia campaign to increase intention to call 9-1-1 for stroke symptoms, upstate New York, 2006-2007. *Preventing chronic disease*, *7*(2), A35.

Kamal, N., Benavente, O., Boyle, K., Buck, B., Butcher, K., Casaubon, L. K., . . . Hill, M. D. (2014). Good is not good enough: the benchmark stroke door-to-needle time should be 30

minutes. *Canadian Journal of Neurological Sciences / Journal Canadien Des Sciences Neurologiques,41*(6), 694-696. doi:10.1017/cjn.2014.41

Kidwell, C. S., Starkman, S., Eckstein, M., Weems, K., & Saver, J. L. (2000). Identifying stroke in the field. *Stroke*, *31*(1), 71–76. doi: 10.1161/01.str.31.1.71

Kleindorfer, D., Miller, R., Sailor-Smith, S., Moomaw, C. J., Khoury, J., & Frankel, M. (2008). The challenges of community-based research: the beauty shop stroke education project. *Stroke*, *39*(8), 2331–2335. doi: 10.1161/strokeaha.107.508812

Kothari, R., Hall, K., Brott, T., & Broderick, J. (1997). Early Stroke Recognition: Developing an Out-of-hospital NIH Stroke Scale. *Academic Emergency Medicine*, *4*(10), 986–990. doi: 10.1111/j.1553-2712.1997.tb03665.x

Kothari, R., Pancioli, A., Liu, T., Brott, T., & Broderick, J. (1999). Cincinnati Prehospital Stroke Scale: Reproducibility and validity. *Annals of Emergency Medicine,33*(4), 373-378. doi:10.1016/s0196-0644(99)70299-4

Krol, M., & El-Deredy, W. (2012). Payoff changes sensitivity by modulating the processing style. *Perception, 41*(5), 623-625. doi:http://dx.doi.org.libproxy.library.wmich.edu/10.1068/p7124

Lavallée, P. C., Meseguer, E., Abboud, H., Cabrejo, L., Olivot, J.-M., Simon, O., … Amarenco, P. (2007). A transient ischaemic attack clinic with round-the-clock access (SOS-TIA): feasibility and effects. *The Lancet Neurology*, *6*(11), 953–960. doi: 10.1016/s1474-4422(07)70248-x

Lawlor, M., Perry, R., & Plant, G. T. (2015). Is the 'Act FAST' stroke campaign lobeist? The implications of including symptoms of occipital lobe and eye stroke in public education

campaigns. *Journal of neurology, neurosurgery, and psychiatry*, *86*(7), 818–820. https://doi.org/10.1136/jnnp-2014-308812

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, *28*(4), 563–575. doi: 10.1111/j.1744-6570.1975.tb01393.x

Lerman, D. C., Tetreault, A., Hovanetz, A., Bellaci, E., Miller, J., Karp, H., . . . Toupard, A. (2010). Applying signal-detection theory to the study of observer accuracy and bias in behavioral assessment. *Journal of Applied Behavior Analysis,43*(2), 195-213. doi:10.1901/jaba.2010.43-195

Lin, C. B., Peterson, E. D., Smith, E. E., Saver, J. L., Liang, L., Xian, Y., . . . Fonarow, G. C. (2012). Patterns, predictors, variations, and temporal trends in emergency medical service hospital prenotification for acute ischemic stroke. *Journal of the American Heart Association,1*(4). doi:10.1161/jaha.112.002345

Lovett, J. K., Dennis, M. S., Sandercock, P. A., Bamford, J., Warlow, C. P., & Rothwell, P. M. (2003). Very early risk of stroke after a first transient ischemic attack. *Stroke*, *34*(8). doi: 10.1161/01.str.0000080935.01264.91

Martin, M. S., Wells, G. A., Crocker, A. G., Potter, B. K., & Colman, I. (2017). Decision curve analysis as a framework to estimate the potential value of screening or other decision-making aids. *International Journal of Methods in Psychiatric Research*, *27*(1). doi: 10.1002/mpr.1601

Marx, J. J., Gube, C., Faldum, A., Kuntze, H., Nedelmann, M., Haertle, B., … Eicke, B. M. (2009). An educational multimedia campaign improves stroke knowledge and risk perception in different stroke risk groups. *European Journal of Neurology*, *16*(5), 612–618. doi: 10.1111/j.1468-1331.2009.02555.x

Marx, J. J., Klawitter, B., Faldum, A., Eicke, B. M., Haertle, B., Dieterich, M., & Nedelmann, M. (2010). Gender-specific differences in stroke knowledge, stroke risk perception and the effects of an educational multimedia campaign. *Journal of Neurology*, *257*(3), 367–374. doi: 10.1007/s00415-009-5326-9

Marx, J. J., Nedelmann, M., Haertle, B., Dieterich, M., & Eicke, B. M. (2008). An educational multimedia campaign has differential effects on public stroke knowledge and care-seeking behavior. *Journal of Neurology*, *255*(3), 378–384. doi: 10.1007/s00415-008-0673-5

Mason, M. A., & Redmon, W. K. (1992). Effects of immediate versus delayed feedback on error detection accuracy in a quality control simulation. *Journal of Organizational Behavior Management, 13*(1), 49-83. doi:http://dx.doi.org/10.1300/J075v13n01_05

McCarthy, D., & Davison, M. (1979). Signal probability, reinforcement and signal detection. *Journal of the Experimental Analysis of Behavior*, *32*(3), 373–386. doi: 10.1901/jeab.1979.32-373

McCarthy, D., & Davison, M. (1980). Independence of sensitivity to relative reinforcement rate and discriminability in signal detection. *Journal of the Experimental Analysis of Behavior*, *34*(3), 273–284. doi: 10.1901/jeab.1980.34-273

McCarthy, D., & Davison, M. (1984). Isobias and alloiobias functions in animal psychophysics. *Journal of Experimental Psychology: Animal Behavior Processes*, *10*(3), 390–409. doi: 10.1037//0097-7403.10.3.390

Mcfall, R. M., & Treat, T. A. (1999). Quantifying The Information Value Of Clinical Assessments With Signal Detection Theory. *Annual Review of Psychology*, *50*(1), 215–241. https://doi.org/10.1146/annurev.psych.50.1.215

McNicol, D. (2005). A Primer of Signal Detection Theory. doi: 10.4324/9781410611949

Mikulík, R., Goldemund, D., Reif, M., Brichta, J., Neumann, J., Jarkovský, J., & Krýza, J. (2011). Calling 911 in response to stroke: no change following a four-year educational campaign. *Cerebrovascular Diseases*, *32*(4), 342–348. doi: 10.1159/000330339

Miller, J. R., Reed, D. D., & Critchfield, T. S. (2014). Modeling the effects of melanoma education on visual detection: a gradient shift analysis. *The Psychological Record*, *65*(2), 323–335. doi: 10.1007/s40732-014-0108-x

Miyamatsu, N., Kimura, K., Okamura, T., Iguchi, Y., Nakayama, H., Toyota, A., … Yamaguchi, T. (2012). Effects of public education by television on knowledge of early stroke symptoms among a Japanese population aged 40 to 74 years. *Stroke*, *43*(2), 545–549. doi: 10.1161/strokeaha.111.634196

Morgenstern, L. B., Bartholomew, L. K., Grotta, J. C., Staub, L., King, M., & Chan, W. (2003). Sustained benefit of a community and professional intervention to increase acute stroke therapy. *Archives of Internal Medicine*, *163*(18), 2198. doi: 10.1001/archinte.163.18.2198

Morgenstern, L. B., Gonzales, N. R., Maddox, K. E., Brown, D. L., Karim, A. P., Espinosa, N., … Conley, K. M. (2007). A randomized, controlled trial to teach middle school children to recognize stroke and call 911. *Stroke*, *38*(11), 2972–2978. doi: 10.1161/strokeaha.107.490078

Müller-Nordhorn, J., Wegscheider, K., Nolte, C. H., Jungehülsing, G. J., Rossnagel, K., Reich, A., … Willich, S. N. (2009). Population-Based Intervention to Reduce Prehospital Delays in Patients With Cerebrovascular Events. *Archives of Internal Medicine*, *169*(16), 1484. doi: 10.1001/archinternmed.2009.232

Palomeras, E., Fossas, P., Quintana, M., Monteis, R., Sebastián, M., Fábregas, C., … Álvarez-Sabín, J. (2008). Emergency perception and other variables associated with extra-hospital delay in stroke patients in the Maresme region (Spain). *European Journal of Neurology*, *15*(4), 329–335. doi: 10.1111/j.1468-1331.2008.02082.x

Pickham, D., Valdez, A., Demeestere, J., Lemmens, R., Diaz, L., Hopper, S., … Lansberg, M. G. (2018). Prognostic Value of BEFAST vs. FAST to Identify Stroke in a Prehospital Setting. *Prehospital Emergency Care*, *23*(2), 195–200. doi: 10.1080/10903127.2018.1490837

Rasura, M., Baldereschi, M., Carlo, A. D., Lisi, F. D., Patella, R., Piccardi, B., … Inzitari, D. (2013). Effectiveness of public stroke educational interventions: a review. *European Journal of Neurology*, *21*(1), 11–20. doi: 10.1111/ene.12266

Rawles, J. (1996). Magnitude of benefit from earlier thrombolytic treatment in acute myocardial infarction: New evidence from Grampian region early anistreplase trial (GREAT). *Bmj,312*(7025), 212-215. doi:10.1136/bmj.312.7025.212

Sandor, T., & Swensson, R. G. (1978). Evaluation of observer performance in detecting blood vessels on simulated angiographic images. *Medical Physics*, *5*(5), 380–386. doi: 10.1118/1.594510

Schlinger, H., & Blakely, E. (1987). Function-altering effects of contingency-specifying stimuli. *The Behavior analyst*, *10*(1), 41–45. https://doi.org/10.1007/BF03392405

Schroeder, E. B., Rosamond, W. D., Morris, D. L., Evenson, K. R., & Hinn, A. R. (2000). Determinants of use of emergency medical services in a population with stroke symptoms. *Stroke*, *31*(11), 2591–2596. doi: 10.1161/01.str.31.11.2591

Sharkey, S., Denke, L., & Herbert, M. A. (2016). using puppets to teach schoolchildren to detect stroke and call 911. *The Journal of School Nursing*, *32*(4), 228–233. doi: 10.1177/1059840516636197

Siket, M. S., & Edlow, J. A. (2012). Transient ischemic attack. *Emergency Medicine Clinics of North America*, *30*(3), 745–770. doi: 10.1016/j.emc.2012.05.001

Sireci, S. G., & Geisinger, K. F. (1995). Using subject-matter experts to assess content representation: An MDS analysis. *Applied Psychological Measurement*, *19*(3), 241–255. https://doi.org/10.1177/014662169501900303

Smith, S. D., & Eskey, C. J. (2011). Hemorrhagic stroke. *Radiologic Clinics of North America*, *49*(1), 27–45. doi: 10.1016/j.rcl.2010.07.011

Sterchi, Y., Hättenschwiler, N., & Schwaninger, A. (2019). Detection measures for visual inspection of X-ray images of passenger baggage. *Attention, Perception, & Psychophysics*, *81*(5), 1297–1311. https://doi.org/10.3758/s13414-018-01654-8

Stern, E. B., Berman, M., Thomas, J. J., & Klassen, A. C. (1999). Community education for stroke awareness. *Stroke*, *30*(4), 720–723. doi: 10.1161/01.str.30.4.720

Tadros, A., Crocco, T., Davis, S. M., Newman, J., Mullen, J., Best, R., … Teter, S. (2009). Emergency medical services-based community stroke education. *Stroke*, *40*(6), 2134–2142. doi: 10.1161/strokeaha.108.532762

Thompson, C., & Adderley, U. (2015). Diagnostic and treatment decision making in community nurses faced with a patient with possible venous leg ulceration: A signal detection analysis. *International Journal of Nursing Studies*, *52*(1), 325–333. doi: 10.1016/j.ijnurstu.2014.10.015

Thompson, C., Yang, H., & Crouch, S. (2012). Clinical simulation fidelity and nurses' identification of critical event risk: a signal detection analysis. *Journal of Advanced Nursing*, *68*(11), 2477–2485. doi: 10.1111/j.1365-2648.2012.05945.x

Tobin, W. O., Hentz, J. G., Bobrow, B. J., & Demaerschalk, B. M. (2009). Identification of stroke mimics in the emergency department setting. *Journal of Brain Disease*, *1*. doi: 10.4137/jcnsd.s2280

Trueblood, J. S., Holmes, W. R., Seegmiller, A. C., Douds, J., Compton, M., Szentirmai, E., … Eichbaum, Q. (2018). The impact of speed and bias on the cognitive processes of experts and novices in medical image decision-making. *Cognitive Research: Principles and Implications*, *3*(1). doi: 10.1186/s41235-018-0119-2

Westbury C. F. (2010). Bayes' rule for clinicians: an introduction. *Frontiers in psychology*, *1*, 192. https://doi.org/10.3389/fpsyg.2010.00192

Williams, O., & Noble, J. M. (2008). 'Hip-hop' stroke. *Stroke*, *39*(10), 2809–2816. doi: 10.1161/strokeaha.107.513143

Wolf, P. A., D'Agostino, R. B., Belanger, A. J., & Kannel, W. B. (1991). Probability of stroke: a risk profile from the Framingham Study. *Stroke*, *22*(3), 312–318. https://doi.org/10.1161/01.str.22.3.312

Wolf, M. M. (1978). Social validity: the case for subjective measurement or how applied behavior analysis is finding its heart1. *Journal of Applied Behavior Analysis*, *11*(2), 203–214. doi: 10.1901/jaba.1978.11-203

Wolfe, J. M., Horowitz, T. S., & Kenner, N. M. (2005). Rare items often missed in visual searches. *Nature*, *435*(7041), 439–440. doi: 10.1038/435439a

Worthmann, H., Schwartz, A., Heidenreich, F., Sindern, E., Lorenz, R., Adams, H.-A., … Dengler, R. (2012). Educational campaign on stroke in an urban population in northern germany: influence on public stroke awareness and knowledge. *International Journal of Stroke*, *8*(5), 286–292. doi: 10.1111/j.1747-4949.2012.00809.x

Table 1

Participant Demographics

		Experience with stroke				
Gender	All	No experience	Family member	Observed in person	Multiple in person	Average age
	n	*n*	*n*	*n*	*n*	*m*
Male	77	29	23	15	10	35
Female	55	25	21	7	2	38
Total	132	54	44	22	12	36

Note. One participant did not provide a gender or age to use in the study, so the total *n* reported here is 132 instead of 133.

Table 2

One-way ANOVA Summary Table for Discriminability

Group	*n*	*M*	*SD*	95% CI	*p*
Control	33	0.54	0.65	0.34 - 0.74	0.69
FAST	33	0.68	0.57	0.49 - 0.88	
Asymmetric	32	0.69	0.62	0.49 - 0.89	
Continuous	35	0.68	0.45	0.49 - 0.88	

Note. Participants excluded for anomalous data in this table. Therefore, n = 133.

Table 3

One-way ANOVA Summary Table for Discriminability: No Participants Excluded

Group	*n*	*M*	*SD*	95% CI	*p*
Control	36	0.50	0.65	0.29 – 0.71	0.83
FAST	39	0.59	0.69	0.38 – 0.79	
Asymmetric	35	0.59	0.75	0.37 – 0.80	
Continuous	40	0.63	0.46	0.43 – 0.83	

Note. No participants excluded for anomalous data in this table. Therefore, n = 150.

Table 4

Post Hoc Comparisons using Tukey's Test for Bias

Group Comparison	*Difference of Means*	95% CI	*T-Value*	*p*
Control - Continuous	-0.77	0.34 - 0.74	-3.16	0.010
Control - Asymmetric	-0.89	0.49 - 0.88	-3.59	0.003
Control - FAST	0.01	0.49 - 0.89	0.05	0.999
Continuous - Asymmetric	-0.12	0.49 - 0.88	-0.51	0.958
Continuous - FAST	0.78	0.15 – 1.41	3.23	0.008
Asymmetric - FAST	0.90	0.26 – 1.54	3.66	0.002

Note. T-value is based on studentized distribution. It is a t-test with a correction for family-wise error rate.

Table 5

Post Hoc Comparisons using Tukey's Test for Bias: No Participants Excluded

Group Comparison	*Difference of Means*	95% CI	*T-Value*	*p*
Control - Continuous	-0.83	-1.64 – 0.02	-2.66	0.043
Control - Asymmetric	-0.92	-1.76 – 0.09	-2.87	0.024
Control - FAST	-0.005	-0.82 – 0.81	-0.01	1.000
Continuous - Asymmetric	-0.10	-0.91 – 0.72	-0.30	0.990
Continuous - FAST	0.82	0.03 – 1.61	2.70	0.038
Asymmetric - FAST	0.92	0.09 – 1.74	2.91	0.021

Note. T-value is based on studentized distribution. It is a t-test with a correction for family-wise error rate. *No participants excluded for anomalous data in this table. Therefore, n = 150.*

Table 6

Social Validity Survey Responses

Characteristic	Mean	Range	SD
In Condition of Uncertainty Assume Stroke			
Risk of Stroke > 10%	9.4	8.0 – 10.0	0.7
Risk of Stroke 5 – 8%	6.5	0.0 – 10.0	3.4
Risk of Stroke 3 – 4%	6.3	1.0 – 9.0	3.3
Likelihood of Recommending Feedback-based Training			
Risk of Stroke > 10%	5.5	1.0 – 10.0	3.2
Risk of Stroke 5 – 8%	4.8	1.0 – 10.0	3.0
Risk of Stroke 3 – 4%	3.9	0.0 – 10.0	2.9
FAST			
How Effective is FAST	7.8	6.0 – 9.0	1.1
Social Appropriateness of Sample Procedures			
Stroke Occurrence in S+ Clip	8.9	7.0 – 10.0	0.9
Stroke Occurrence in S- Clip	3.4	1.00 – 7.0	2.2
Social Appropriateness of Procedures			
Rating of Feedback Intervention Procedures	8.7	8.0 – 10.0	0.8
Rating of Feedback Delivery (Animation & Description)	7.0	3.0 – 10.0	2.9
Rating of Signal Detection Format (2-button choice)	8.0	3.0 – 10.0	2.5

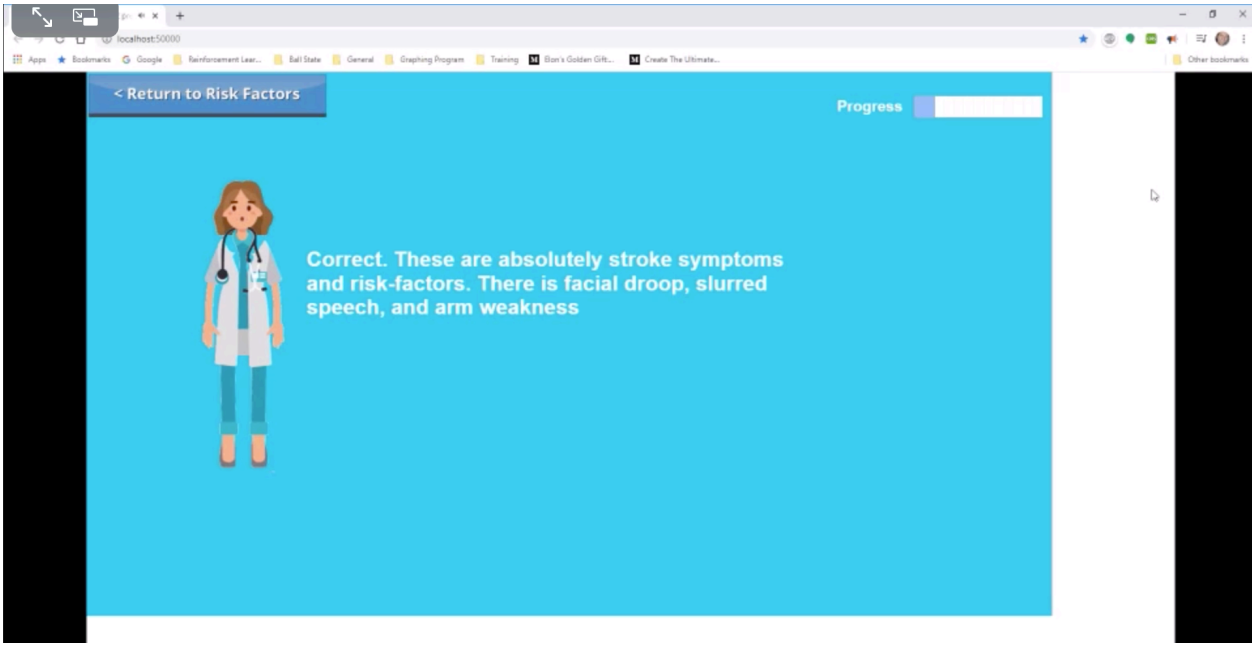
Note. This table displays the responses from medical professionals on the social validity survey.

For the purposes of this table, 0: Strongly Disagree and 10: Strongly Agree or 0: Dislike a Great

Deal, 10: Like a Great Deal or 0: Do not believe a stroke is occurring, 10: Strongly believe a stroke is occurring.

Figure 1

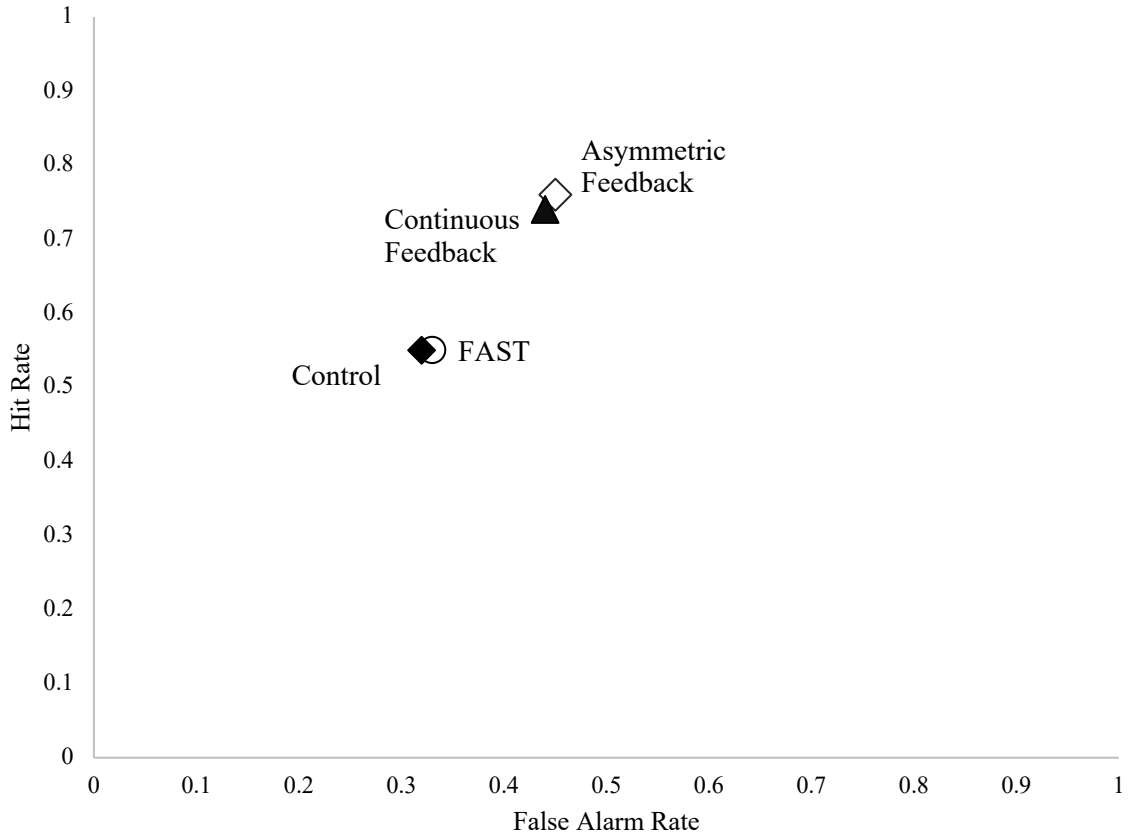
Demonstration of Animated and Auditory Feedback with Dr. Hazel



Note. Dr. Hazel was only present for trials in the continuous feedback condition and some trials of the asymmetric feedback condition. She did not appear for the FAST or control groups other than to introduce participants to the task.

Figure 2

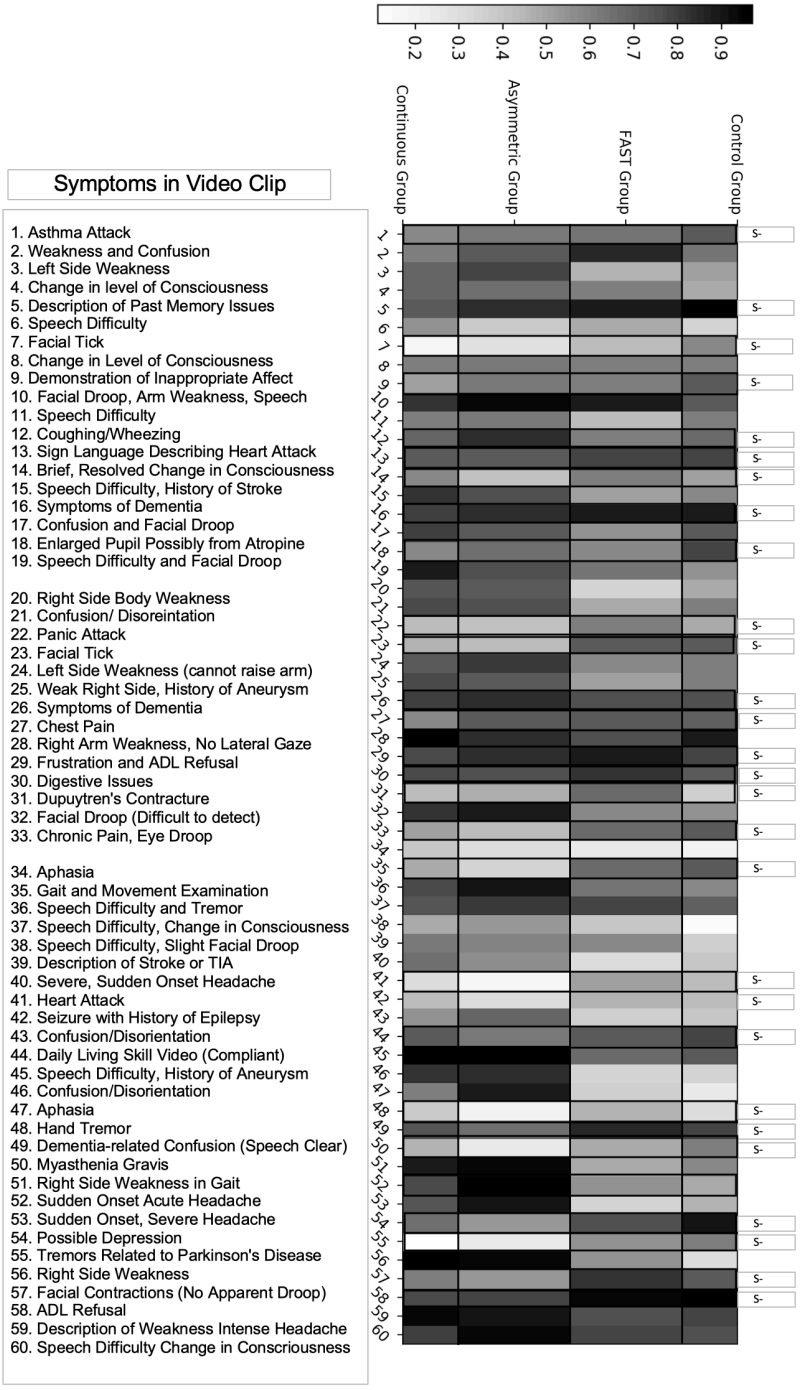
Hit Rate and False Alarm Rate for Groups



Note. This figure shows the hit rate and false alarm rates of all four groups. Higher hit rates indicate a high ratio of hits to misses. Higher false alarm rates indicate a higher ratio of false alarms to correct rejections.

Figure 3

Trial-by-trial Error Accuracy by Group with Symptoms

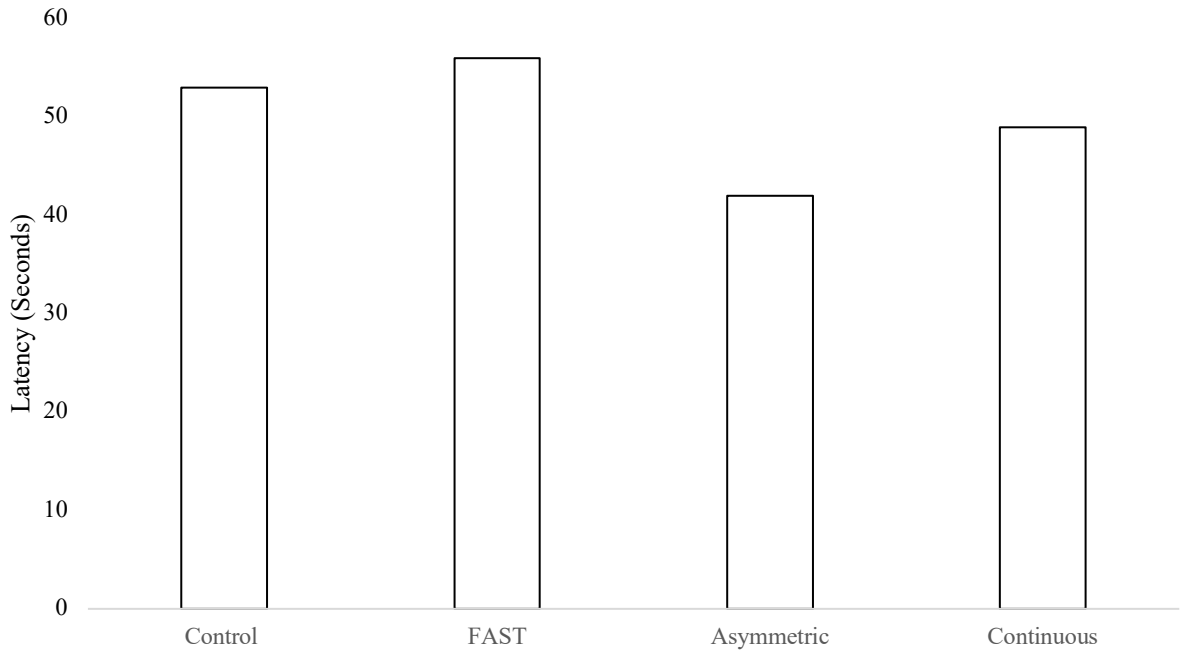


Note. This figure captures trial-by-trial accuracy and the symptoms depicted in each trial.

*The bar at the top left shows higher levels of accuracy as darker colors than lower levels. Trials where no stroke symptoms occurred are denoted with a “S-” above the top panel.

Figure 4

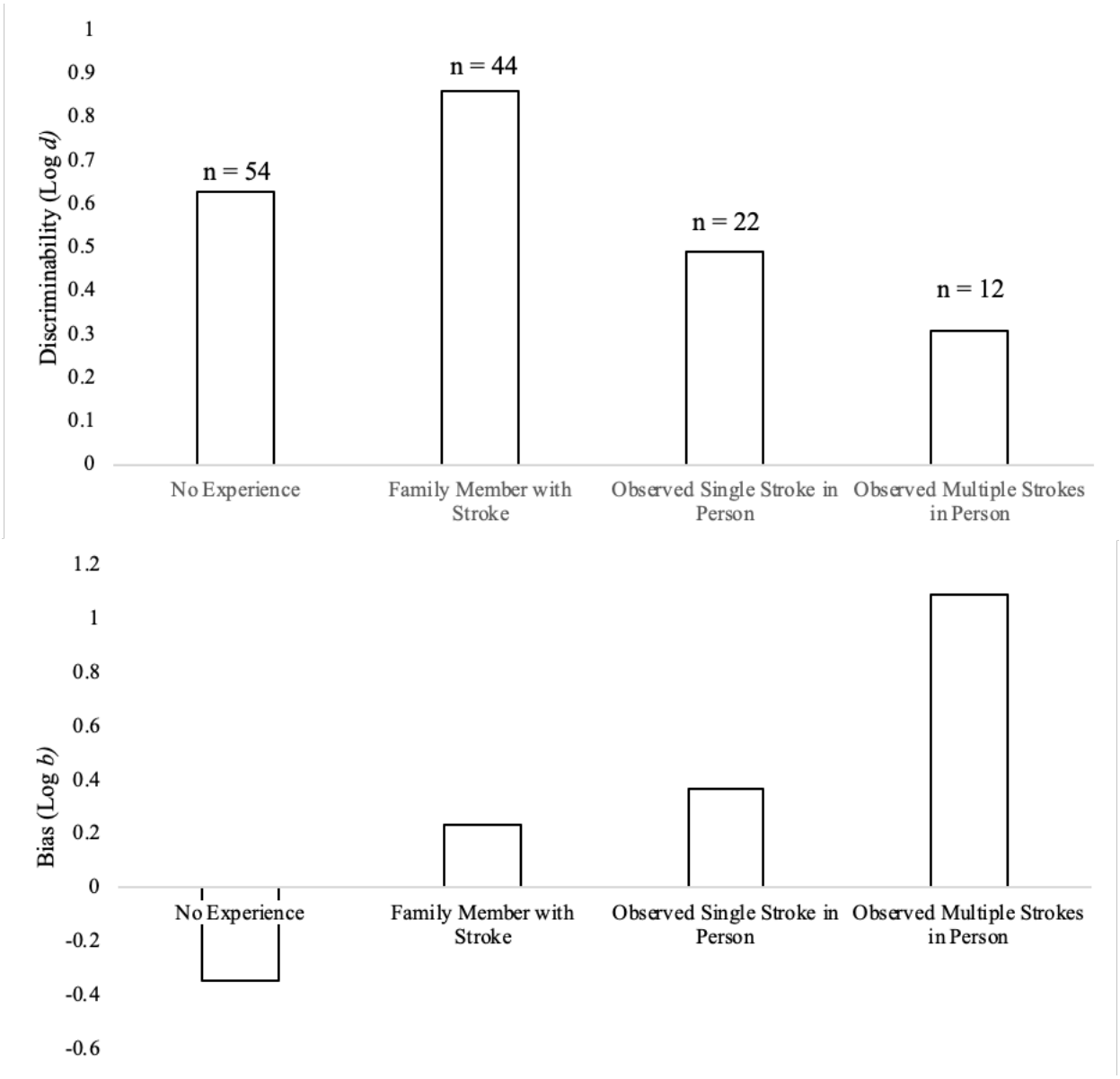
Time to Decision by Group



Note. This figure displays pooled mean decision time by group

Figure 5

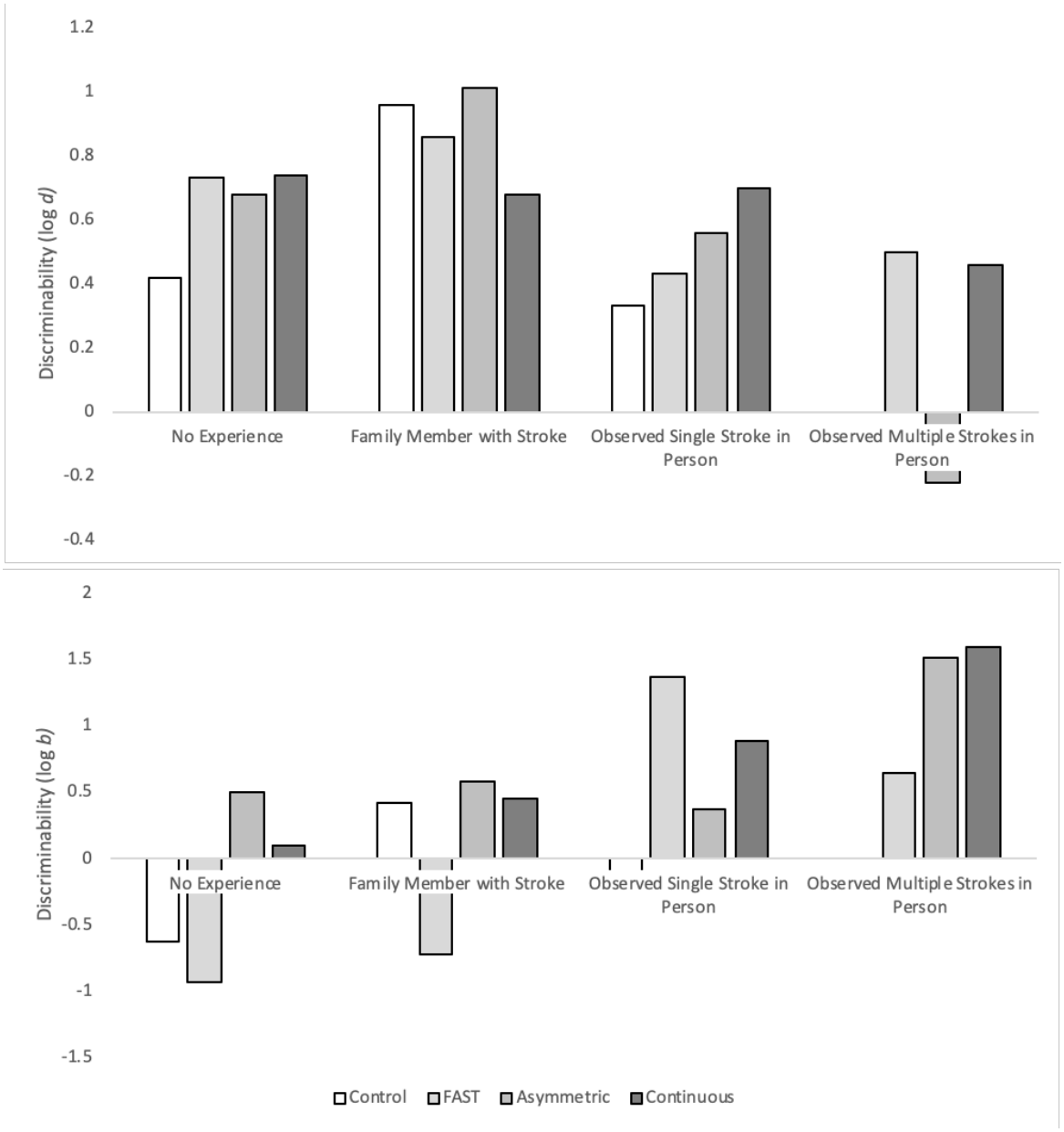
Discriminability and Bias by Experience



Note. This figure demonstrates pooled discriminability by experience (top panel) and pooled bias by experience (bottom panel).

Figure 6

Discriminability and Bias by Experience Level and Group

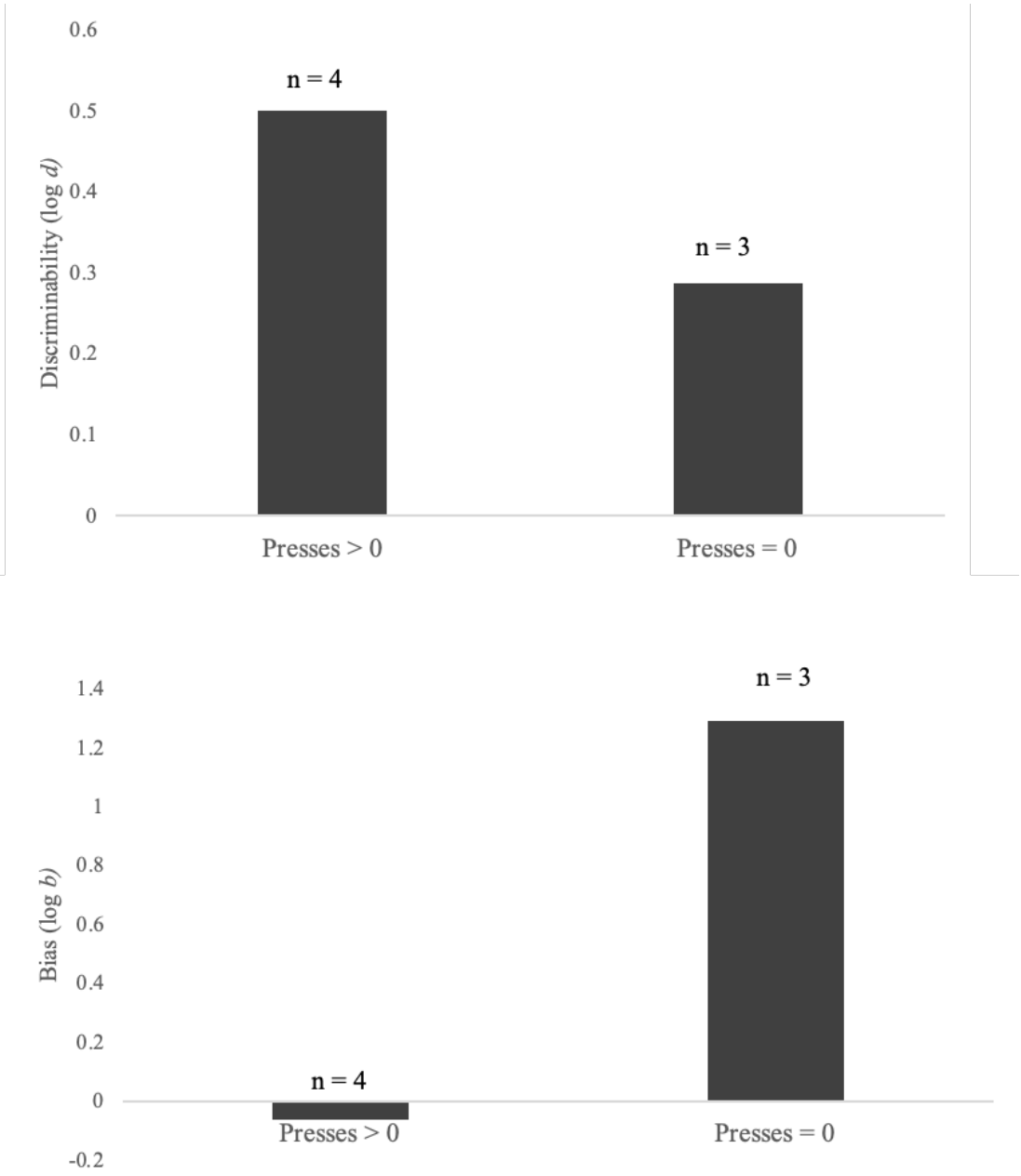


Note. This figure shows the pooled discriminability by experience and group (top panel) and the pooled bias by experience and group (bottom panel).

*There were only 12 total participants who had observed multiple strokes in person and none of them were randomly assigned to the control group.

Figure 7

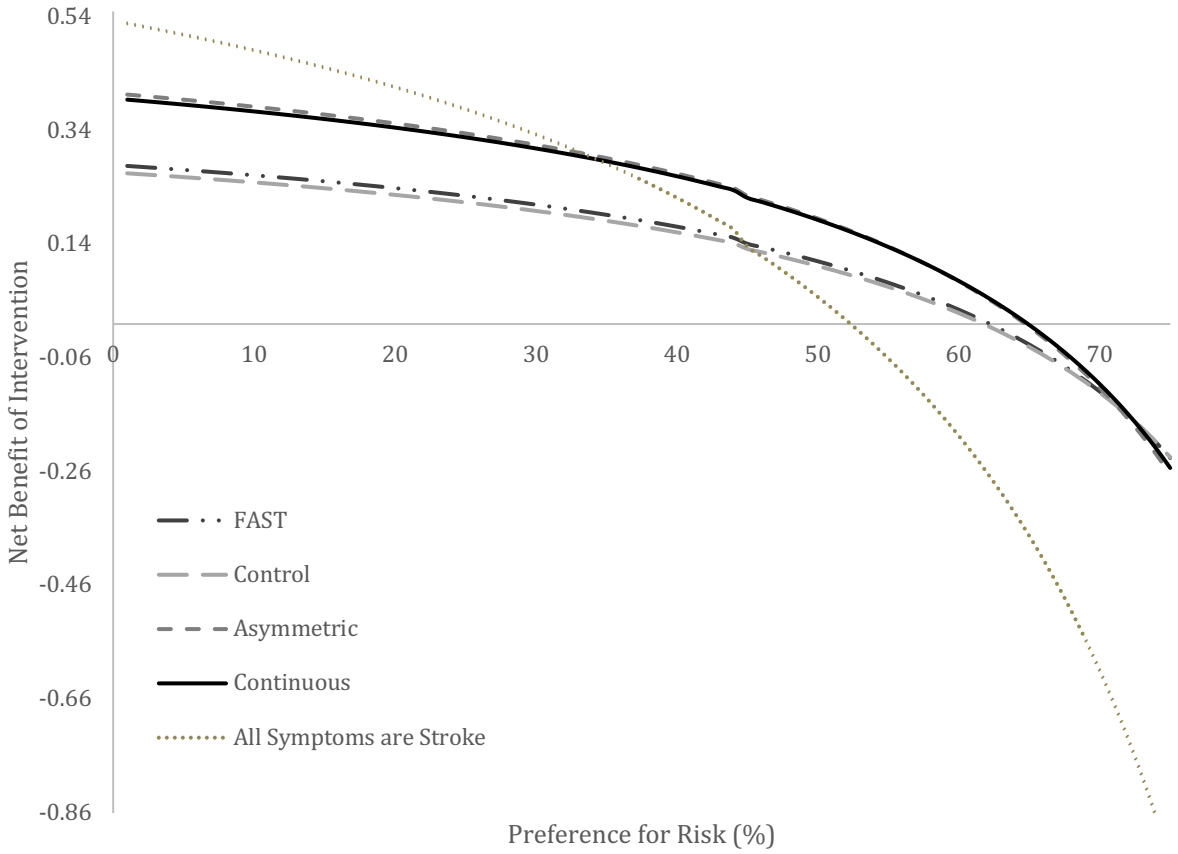
Bias and Discriminability Based on FAST Reminder Presses



Note. This figure shows the mean discriminability score of participants who viewed the FAST reminder at least once or not at all (top panel) and the mean bias score of participants who viewed the FAST reminder at least once or not at all (bottom panel).

Figure 8

Decision Curve Analysis Within the Context of the Experiment

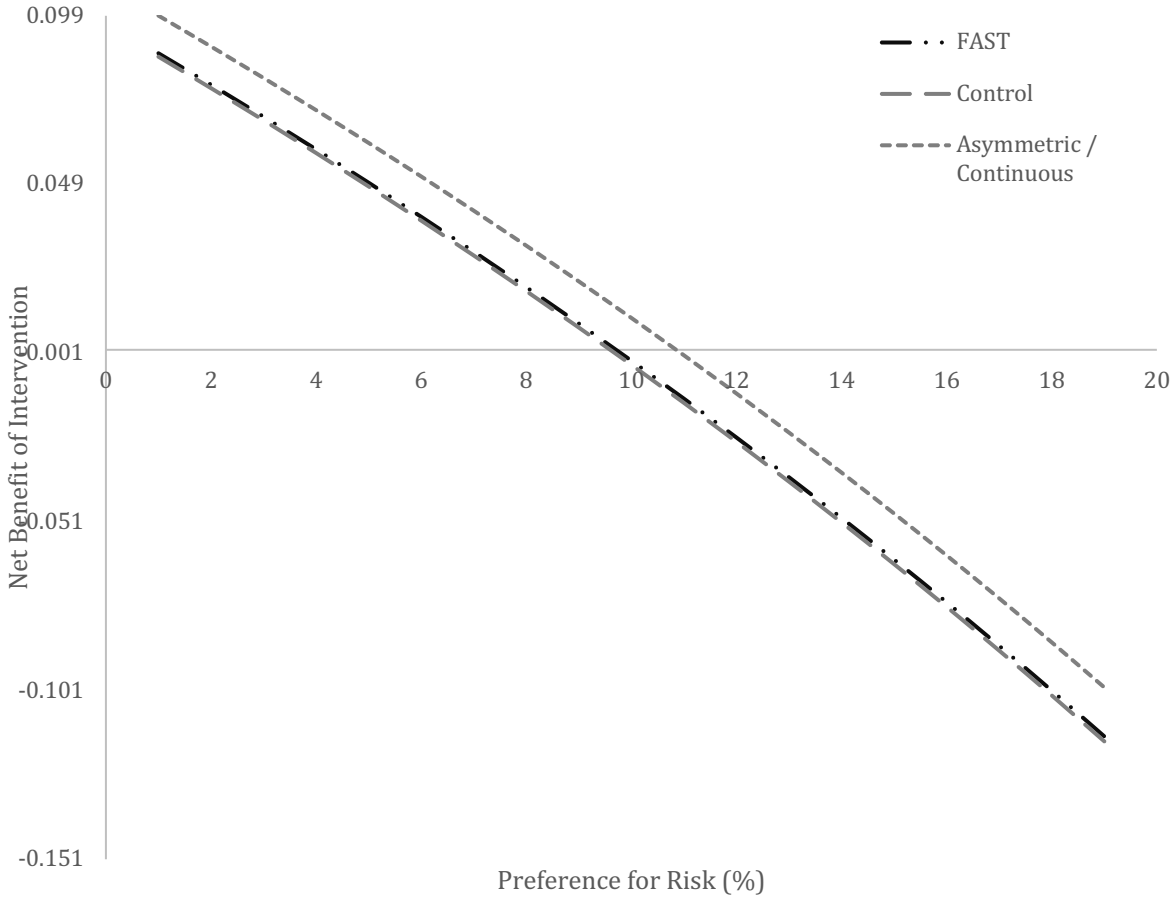


Note. This figure shows the net benefit of the interventions and control at various levels of risk of stroke. Net benefit is can be interpreted as the percentage of people that the intervention will help to identify stroke in.

*The net benefit here is affected by the stimulus presentation ratio which is close to 1:1, while the actual probability of stroke is much lower.

Figure 9

Decision Curve Analysis Considering Prevalence of Stroke



Note. This figure shows the net benefit of the interventions and control at various levels of risk of stroke. Net benefit is can be interpreted as the percentage of people that the intervention will help to identify stroke in. Asymmetric and continuous values were nearly identical so they are represented as a single curve.

APPENDICES

Appendix A

Stroke Intervention Social Validity Survey

Start of Block: Default Question Block

Principal Investigator: Dr. Jonathan C. Baker, PhD., BCBA-D
Student Investigator: Jordan D. Bailey, MA, BCBA
Title of Study: Applying Signal Detection Methods to Potential Stroke Recognition
STUDY SUMMARY: This consent form is part of an informed consent process for a research study and it will provide information that will help you decide whether you want to take part in this study. Participation in this study is completely voluntary. The purpose of the research is to: determine how good a signal detection task involving short video clips is for training people to recognize potential strokes. This project will serve as Jordan Bailey's dissertation for the requirements of the doctor of philosophy in psychology degree. You will be asked to rate how effective a signal detection task appears to be at helping others to identify strokes. Your time in the study will be approximately 15-30 minutes. Possible risk and costs to you for taking part in the study consist of the time involved. There are no direct benefits from taking part in this study. You are invited to participate in this research project titled Applying Signal Detection Methods to Potential Stroke Recognition and the following information in this consent form will provide more detail about the research study. Please ask any questions if you need more clarification and to assist you in deciding if you wish to participate in the research study. You are not giving up any of your legal rights by agreeing to take part in this research or by signing this consent form. After all of your questions have been answered and the consent document reviewed, if you decide to participate in this study, you will be asked to sign this consent form.
What are we trying to find out in this study? We would like to find out how socially valid a signal detection task is for training undergraduates to recognize potential strokes or transient ischemic attacks.
Who can participate in this study? A medical professional. We are specifically seeking responses of physicians and nurses.
Where will this study take place? The survey will be e-mailed to you directly. You can complete them wherever you choose.
What is the time commitment for participating in this study? The study will take between 15 and 30 minutes to complete.
What will you be asked to do if you choose to participate in this study? In this study you will be asked to provide a response to several survey questions. The survey will consist of questions that ask about the social validity of a signal detection task for teaching stroke

recognition. Social validity is your opinion about the acceptability, effectiveness, and satisfaction of the procedures and results of a study.

What information is being measured during the study? This section will describe the measurements that we are going to take during your participation in the study. The measurements that will be collected are your responses to questions about a previous experiment. Specifically we would like to know how well you understood the experiment, whether you think the samples used in the experiment were appropriate, your assessment of how much the results of the study matter, how much you like the task, how likely you would be to recommend it to a patient, and the degree to which you think patients should be conservative with their decision to seek medical treatment.

What are the risks of participating in this study and how will these risks be minimized? There are no risks for participating in this study.

What are the benefits of participating in this study? There are no known benefits to participating in this study.

Are there any costs associated with participating in this study? There are no costs associated with participating in this study.

Is there any compensation for participating in this study? There is no compensation for participating in this study.

Who will have access to the information collected during this study? Your responses will not be connected with you in any way as the program does not store any kind of personal information. All data will be stored in a protected server, or external hard drive in a locked room and no one will be able to identify you, your responses, or identifying information such as your name. No one will know you participated in the study. You will be asked to make up a unique participant identification number at the beginning of the study. Data may be presented at a conference or in a manuscript that is submitted for publication, but it will not be linked with you in any way.

What will happen to my information or biospecimens collected for this research after the study is over? The information collected about you for this research will not be used by or distributed to investigators for other research.

What if you want to stop participating in this study? You can choose to stop participating in the study at anytime for any reason. You will not suffer any prejudice or penalty by your decision to stop your participation. You will experience NO consequences either academically or personally if you choose to withdraw from this study. The investigator can also decide to stop your participation in the study without your consent. Should you have any questions prior to or during the study, you can contact the primary investigator, Dr. Jonathan Baker at (269) 387-4355 or jonathan.c.baker@wmich.edu. You may also contact the Chair, Institutional Review Board at 269-387-8293 or the Vice President for Research at 269-387-8298 if questions arise during the course of the study. This study was approved by the Western Michigan University Institutional Review Board (WMU IRB) on (5-4-20). Participating in this survey online indicates your consent for use of the answers you supply

Page Break

Info This questionnaire details the effects and goals of a study that was conducted on stroke identification. The goals of the study were to look at the commonly used Facial Droop, Arm Weakness, Speech, Time to Call 911 (FAST) warning signs intervention in comparison with condition s that involved watching videos, identifying whether or not a stroke was occurring, and receiving one of two types of feedback about performance accuracy. The study used 60 video clips that were approximately 1 minute long.

There was little difference between treatments in ability to recognize symptoms, but the groups receiving the feedback intervention were more likely to say that a stroke was occurring in conditions of uncertainty (i.e., symptoms could not be identified).

We are interested in your impressions of how valuable it is for people to respond as if a stroke is occurring in conditions of uncertainty, and your reaction to different methods of teaching symptom recognition.

Q23 Please enter details about your profession and experience with stroke

_____

Q1 If an individual is uncertain about the symptoms they are experiencing or caregiver is uncertain about the symptoms they observe, **and the individual with the symptoms is at significant risk of stroke (10-year probability > 10%)**, the individual or caregiver should assume a stroke is occurring.

	Strongly Disagree				Neutral					Strongly Agree
	0	1	2	3	4	5	6	7	8	9	10
0: Strongly Disagree, 10: Strongly Agree ()											

Q2 If an individual is uncertain about the symptoms they are experiencing or caregiver is uncertain about the symptoms they observe, **and the individual with the symptoms is at**

moderate risk of stroke (10-year probability 5-8%), the individual or caregiver should assume a stroke is occurring.

Strongly Disagree Neutral Strongly Agree

0 1 2 3 4 5 6 7 8 9 10

0: Strongly Disagree, 10: Strongly Agree ()

Q3 If an individual is uncertain about the symptoms they are experiencing or caregiver is uncertain about the symptoms they observe, **and the individual with the symptoms is not necessarily at risk of stroke (10-year probability 3-4%)**, the individual or caregiver should assume a stroke is occurring and seek medical care.

Strongly Disagree Neutral Strongly Agree

0 1 2 3 4 5 6 7 8 9 10

0: Strongly Disagree, 10: Strongly Agree ()

Q4 How likely are you to recommend a 1-2 hour video-based training program with feedback for those who are significantly at-risk of stroke (10-year probability > 10%) compared to letting them know about FAST?

Very Unlikely Neutral Very Likely

0 1 2 3 4 5 6 7 8 9 10

0: Very Unlikely, 10: Very Likely ()

Q5 How likely are you to recommend a 1-2 hour video-based training program with feedback for those who are at moderate risk of stroke (10-year probability 5% - 8%) compared to letting them know about FAST?

Very Unlikely Neutral Very Likely

0 1 2 3 4 5 6 7 8 9 10

0: Very Unlikely, 10: Very Likely ()	

Q6 How likely are you to recommend a 1-2 hour video-based training program with feedback for those who are at moderate risk of stroke (10-year probability 3% - 4%) compared to letting them know about FAST?

Very Unlikely Neutral Very Likely

0 1 2 3 4 5 6 7 8 9 10

0: Very Unlikely, 10: Very Likely ()	

Q7 How effective do you believe FAST is at causing patients to seek help when they are having stroke symptoms?

Extremely Ineffective | Moderately Ineffective | Slightly Ineffective | Neutral | Slightly Effective | Moderately Effective | Extremely Effective

0 1 2 3 4 5 6 7 8 9 10

0: Very ineffective, 10: Very effective ()	

Q8

Please watch the following video clip. Rate the degree to which you believe that the video represents symptom(s) that patients or caregivers should seek medical help for STROKE.

	A stroke is not occurring	It's highly unlikely that s stroke is occurring	It's unlikely that a stroke is occurring	Neutral	It's likely that a stroke is occurring	It's highly likely that a stroke is occurring	A stroke is occurring
	0	1	2	3	4	5	6	7	8	9	10
0: Do not believe a stroke is occurring, 10: Strongly believe a stroke is occurring ()											

Q9

Please watch the following video clip. Rate the degree to which you believe that the video represents symptom(s) that patients or caregivers should seek medical help for STROKE.

	A stroke is not occurring	It's highly unlikely that s stroke is occurring	It's unlikely that a stroke is occurring	Neutral	It's likely that a stroke is occurring	It's highly likely that a stroke is occurring	A stroke is occurring
	0	1	2	3	4	5	6	7	8	9	10
0: Do not believe a stroke is occurring, 10: Strongly believe a stroke is occurring ()											

Q10 Are there any conditions under which a bias toward seeking medical attention for stroke in conditions of uncertainty would NOT be indicated?

_____

Q11 Assuming the effects of the feedback interventions were not long-lasting, how often would it be feasible to provide those at-risk of stroke with a booster/reminder intervention?

Also consider - How likely do you think patients would be to use this video and feedback training method?

The training was web-based and took 1-2 hours to complete. Staff did not need to be present while patients/families completed the training. This video training could be accessed at home, in the office, or anywhere using a mobile device.

_____

Q12 How much do you like the video and feedback procedures described here?

Dislike a great deal	Dislike a moderate amount	Dislike a little	Neither like nor dislike	Like a little	Like a moderate amount	Like a great deal

0 1 2 3 4 5 6 7 8 9 10

0 = Dislike a great deal, 10 = Like a great deal	
()	

Q13

The video shows the manner in which feedback was delivered. How much do you like the feedback delivery?

Dislike a great deal	Dislike a moderate amount	Dislike a little	Neither like nor dislike	Like a little	Like a moderate amount	Like a great deal

0 1 2 3 4 5 6 7 8 9 10

0 = Dislike a great deal, 10 = Like a great deal	
()	

Q14 Participants could select one of two buttons to declare whether they thought a stroke was occurring or not. They were given animated feedback after the trials. How much do you like the task?

	Dislike a great deal	Dislike a moderate amount	Dislike a little	Neither like nor dislike	Like a little	Like a moderate amount	Like a great deal
	0	1	2	3	4	5	6	7	8	9	10
0 = Dislike a great deal, 10 = Like a great deal ()											

Q15 Please respond to the following questions.

	0	10	20	30	40	50	60	70	80	90	100
What percentage of stroke symptoms should someone be able to accurately identify after intervention? ()											
How many false-alarms are acceptable? (Stroke symptoms did not occur but they were identified) ()											

Q16 If you would like to enter the drawing for a $50 gift card please enter your e-mail address here.

_____

Appendix B

HSIRB Approval

WESTERN MICHIGAN UNIVERSITY

Institutional Review Board
FWA00007042
IRB00000254

Date: November 19, 2019

To: Jonathan Baker, Principal Investigator
Jordan Bailey, Student Investigator for dissertation

From: Amy Naugle, Ph.D., Chair

Re: IRB Project Number 19-10-02

This letter will serve as confirmation that your research project titled "Applying Signal Detection Methods to Video-Based Potential Stroke Recognition" has been **approved** under the **expedited** category of review by the Western Michigan University Institutional Review Board (IRB). The conditions and duration of this approval are specified in the policies of Western Michigan University. You may now begin to implement the research as described in the application.

Please note: This research may **only** be conducted exactly in the form it was approved. You must seek specific board approval for any changes to this project (e.g., ***add an investigator, increase number of subjects beyond the number stated in your application, etc.***). Failure to obtain approval for changes will result in a protocol deviation.

In addition, if there are any unanticipated adverse reactions or unanticipated events associated with the conduct of this research, you should immediately suspend the project and contact the Chair of the IRB for consultation.

The Board wishes you success in the pursuit of your research goals.

A status report is required on or prior to (no more than 30 days) November 18, 2020 and each year thereafter until closing of the study.

When this study closes, submit the required Final Report found at https://wmich.edu/research/forms.

Note: All research data must be kept in a secure location on the WMU campus for at least three (3) years after the study closes.

Office of the Vice President for Research
Western Michigan University
1903 W. Michigan Ave., Kalamazoo, MI 49008-5456
PHONE: (269) 387-8293 FAX: (269) 387-8276
WEBSITE: wmich.edu/research/compliance/hsirb

CAMPUS SITE: Room 251 W. Walwood Hall