2017

Pathways Linking Clinician Demographics to Mental Health Diagnostic Accuracy: An International Perspective

Julia Brechbiel
Virginia Commonwealth University

PATHWAYS LINKING CLINICIAN DEMOGRAPHICS TO MENTAL HEALTH DIAGNOSTIC ACCURACY: AN INTERNATIONAL PERSPECTIVE

A thesis submitted in partial fulfillment of the requirements of the degree of Master of Science at Virginia Commonwealth University

By: Julia K. Brechbiel

Bachelor of Science, The College of William & Mary, May 2014

Director: Jared Keeley, Ph.D.

Associate Professor of Psychology

Virginia Commonwealth University

Richmond, VA

Acknowledgement

There are several individuals I wish to thank for their help in completing this thesis. First, I want to thank my advisor and thesis director, Jared Keeley, for his guidance and support throughout this project. I am grateful to have worked with him on this project, and I look forward to our continued work together. Also, I would like to thank my committee members, Paul Perrin and Kia Bentley for their feedback and assistance with this project. Lastly, I would like to thank my family, friends, and my cohort for their continual support and encouragement over the past year.

Table of Contents

List of Tables

List of Figures

Abstract

PATHWAYS LINKING CLINICIAN DEMOGRAPHICS TO MENTAL HEALTH DIAGNOSTIC ACCURACY: AN INTERNATIONAL PERSPECTIVE

By Julia K. Brechbiel

A thesis in partial fulfillment of the requirements for the degree of Master of Science at Virginia Commonwealth University,

Virginia Commonwealth University, 2017

Director: Jared Keeley, Ph.D.
Associate Professor of Psychology
Department of Psychology

Significant research efforts have focused on examining the effect of patient factors on providing diagnoses across clinical settings; however, the influence of clinician demographics have received less attention. This study aimed to understand the impact of nonclinical factors such as clinician characteristics and response time on diagnostic accuracy. The study used data from a WHO field study of the ICD-11 development (n = 1822) that required clinicians to diagnose two case vignettes. Clinicians' slower response times had a significant positive impact on their rates of diagnostic accuracy. However, there was no evidence that clinicians' demographic features were directly related to their diagnostic accuracy. Rather, clinicians' age, years of experience, world region, and their clinical profession indirectly predicted accuracy through their overall response time. Contradictory to decision-making theories, older clinicians and clinicians with more years of experience had higher rates of diagnostic accuracy when they spent more time completing the study. Additionally, clinicians in South America with slower

response times had higher accuracy compared to clinicians in North America. Clinicians in Asia had faster response times that negatively impacted their accuracy rates compared to North America clinicians. These findings suggest differences in response time and the applicability of the ICD-11 across cultures. Consistent with previous research, medical professionals with quicker response times had the lowest rates of accuracy compared to psychologists and other clinical professionals. These findings highlight the need for researchers and clinicians to consider the role their dispositional features have in the diagnostic process. Moreover, it is crucial that future research into diagnostic decision-making and accuracy should consider additional mediating factors such as response style, culture, and experience.

Pathways Linking Clinician Demographics to Mental Health Diagnostic Accuracy: An International Perspective

Diagnostic decision-making is an important part of clinical practice because it informs client treatment, selection of interventions, and therapeutic outcomes. Inaccurate diagnostic decisions can lead clinicians to select inappropriate or ineffective treatment plans and inhibit clients' therapeutic progress. To minimize instances of misdiagnosis, it is important to understand clinicians' decision-making process and what factors influence the accuracy of their diagnostic decision. Two predominant explanations for how clinicians make diagnostic decisions come from the field of cognitive psychology: (a) the use of cognitive heuristics and biases and (b) dual-process theory. Cognitive psychology and decision theories have identified numerous heuristics or "mental short cuts" in which people unconsciously engage that influence their judgement and decision-making (Blumenthal-Barby & Krieger, 2015). For the most part, heuristics help individuals reach the correct conclusion quicker, conserving mental energy. However, reliance on these shortcuts can lead individuals to make incorrect judgements (Garb, 2005; Garb, 2010; Schwartz & Elstein, 2009). In contrast, dual-process theory builds on traditional decision-making models by proposing that individuals have two modes of thinking—intuitive and analytical reasoning—with biases and heuristics occurring in the intuitive mode (Croskerry & Nimno, 2011). The addition of the analytic mode acknowledges that on occasion individuals use scientific thinking and hypothesis testing during the decision-making process.

While these theories elucidate clinicians' thought processes, they may not account for everything that may influence the diagnostic decision. McKinlay, Potter and Feldman (1996) found that physicians' tendencies for diagnosis were not completely explained by traditional theories of clinical decision-making. Instead, past studies reported that non-clinical factors have an influence on diagnostic accuracy (Eisenberg, 1979; McKinlay et al., 1996). The clinical diagnostic process is influenced by a multitude of social, cultural, and psychological variables that may be overlooked when studying diagnostic accuracy. These factors are typically divided into three groups: (a) characteristics of the patient, (b) characteristics of the clinician, and (c) features of the clinical setting. Significant research efforts have focused on studying the influence the patient's demographics (e.g., gender, age, race/ethnicity) on psychodiagnostic accuracy in numerous settings (e.g., general practice, psychiatry, nursing, emergency care) while clinician characteristics have received less attention (Blumenthal-Barby & Krieger, 2015). For example, studies on diagnostic errors and clinician biases predominately focus on the patient's race, gender, and culture and rarely report the clinician's race, gender, or culture when making cross-cultural diagnoses (Solomon, 1992). Moreover, research about the relation between clinicians' level of experience and diagnostic decisions typically do not report demographic information about clinicians or include them in statistical analyses. This gap in the literature places focus on patient dispositional factors rather than researchers and clinicians examining the influence their own dispositional factors have on their work. The current study will use an existing dataset to investigate the effect of some clinician dispositional factors on diagnostic accuracy.

Theories of Decision Making

Heuristics and Cognitive Bias

Research on heuristics and cognitive biases began with Tversky and Kahneman in the 1970s. Since the field's inception, a myriad of heuristics and biases have been identified and definitions for these terms have evolved. Presently in the cognitive psychology literature, the term heuristic is used to describe unconscious, automatic cognitive short-cuts that assist individuals in judgements and decision-making (Keren & Teigen, 2004). Heuristics are not without consequences. While they allow individuals to make quicker decisions with less cognitive energy, they can result in inaccurate judgements due to taking short-cuts in reasoning. In clinical psychology, heuristics are used to describe how clinicians make diagnostic decisions and explore how to improve training and clinical practice (Garb, 2005). Clinicians use heuristics to reduce memory limitations and high cognitive loads; however, this can lead to diagnostic errors. For instance, reliance on heuristics can lead clinicians to focus on more salient symptoms, become overconfident, and may result in misdiagnosis (Brannon & Carson, 2003; Smith & Dumont, 1997). One strategy to reduce these errors is to be familiar with heuristics and learn how to avoid them. Three main heuristics referenced in the literature on accuracy of clinical diagnoses are the availability, representativeness, and affect heuristics. The utility and consequences of each of these heuristics is explored below.

Availability heuristic. Individuals use the accessibility with which an event comes to mind to estimate the probability of the event occurring as a shortcut to make quicker decisions (Tversky & Kahneman, 1974). When individuals rely on available information or engage in the availability heuristic, they typically judge things to be more likely or more frequent if they readily come to mind. This can be generalized to psychology and clinical work. For example, a clinician's recent experiences with a diagnosis may inflate the clinician's likelihood to select the diagnosis in the future. Additionally, information that a clinician can easily recall about a patient will have a greater influence on the clinician's judgment (Blumenthal-Barby & Krieger, 2015). Conversely, if patients rarely present with a diagnosis, then the diagnosis is less available and may be underdiagnosed (Croskerry, 2003).

Representativeness heuristic. Clinicians can also make diagnoses by judging how similar a patient's presentation is to a diagnostic category or a prototypical patient with a diagnosis (Blumenthal-Barby & Krieger, 2015; Garb, 1996). This strategy involves the clinician comparing a patient to a mental representation of a diagnosis and making a categorical, yes or no, decision (Tversky & Kahneman, 1974). While historically clinical diagnosis was viewed as a categorization process, reducing it to pattern-recognition can result in a misapplication of the representativeness heuristic and serious clinical errors (Croskerry, 2003). Nurses have been shown to rely primarily on the representativeness heuristic to determine whether a patient requires emergency assistance. Nurses reported coping with uncertainty in emergency situations by comparing patients to a reference group or a prototype that included past patients with similar presenting conditions (Cioffi, 2001). This is analogous to how mental health clinicians apply this heuristic to make diagnostic decisions. This heuristic can result in errors when the clinician's or nurse's prototypical representation of a diagnosis or emergency situation is not accurate, by focusing on similarities more than differences between the patient and prototype, or ignoring the prevalence rate of the proposed diagnosis in their clinical sample (Croskerry, 2003; Schwartz & Elstein, 2009). For example, Graber et al. (2005) investigated the reasons behind 100 diagnostic errors by internal medicine interns. Cognitive factors contributed to 74% of errors with faulty synthesis of information as the most common. Additionally, there were a few cases involving rare conditions that were initially mistaken for more common conditions with similar symptoms.

Affect heuristic. Individuals often make decisions based on intuition or "gut instinct" and evaluate the validity of the decision based on the emotion it evokes (Garb, 2010; Kahneman, 2003). Emotions are typically the focus of clinical work but they also can guide clinical judgement and decision-making. The affect a clinician associates with a patient either can have a positive or negative influence on clinical judgment. Clinicians' affect is thought to be related to clinical intuition and is often an accurate metric for determining ethical versus unethical behavior (Garb, 2010). Positive affect may be related to determining meaningful goals and treatment progress with patients. Isen, Rosenzweig, and Young (1991) found that medical students who received positive feedback to induce a positive affect were quicker and more accurate in determining which out of six hypothetical patients had lung cancer compared to medical students who did not receive feedback. Also, subjects who received positive feedback compared to subjects in the control condition appeared to integrate more information, expressed less confusion, and went beyond the scope of the task by providing diagnoses and treatment options for the other five hypothetical patients. The presence of affect can also yield negative clinical effects like racial or gender biases or diagnostic errors. For example, a patient that arouses negative feelings in a clinician (i.e., countertransference) might be given the wrong diagnosis of Borderline Personality Disorder simply because of the negative impression the patient creates. To counter their affective response to patients, clinicians are recommended to (a) attend to their feelings, (b) consider several alternatives, (c) consult with peers or supervisors, and (d) review diagnostic criteria more closely (Garb, 2005).

Dual-process model

Building from Tversky and Kahneman's early work, a more nuanced way to explain how clinicians make decisions is the dual-process model (Epstein, 1994; Kahneman, 2003). This model merges facets of various theories together to describe two distinct systems that operate in tandem during the decision-making process. The intuitive judgement system, designated System 1, describes how people typically engage in the decision-making process. As the name suggests, System 1 is the more unconscious, automatic of the two systems. This track includes canonical cognitive processes like pattern recognition and heuristics. While efficient, this system relies on gut reactions making it more susceptible to factors like fluctuating emotional states, affecting the decision-making process. In contrast, System 2, the analytic mode of judgement, is slower, deliberate, and hypothesis driven, resulting in more rational, fact-based decision-making. While this approach may yield a more precise decision, it requires more cognitive effort and may be unreliable under heavier cognitive load.

These two systems operate together to make decisions and can be applied to summarize clinicians' diagnostic process (Croskerry & Nimno, 2011; Schwartz & Elstein, 2009). The clinicians' level of recognition of their patient's presenting concern determines which system they use to process this information. When the person's symptoms are familiar or follow an expected pattern, this will trigger System 1, or the intuitive mode and a fast diagnostic decision is made. The alternative is relying on System 2, or analytic reasoning, when the patient's symptoms are unfamiliar. The clinician will engage in a slower, systematic, hypothesis testing process to determine the correct diagnosis.

Over time, if multiple patients with the same symptoms present and illicit the clinician's analytic mode, this repetition will develop into pattern recognition and subsequent incidents will engage System 1. This is sometimes referred to as expertise development (Croskerry & Nimno, 2011). There is evidence suggesting that clinicians with more experience do not depend on strategies found in System 2, e.g., hypothesis testing, and instead have mastery of content allowing most of their decision making to occur in System 1 (Schwartz & Elstein, 2009). In contrast, novice clinicians typically lack speed, efficiency, and accuracy, suggesting that they process decisions more analytically through System 2.

Another way these two processes interact is their ability to override the other's decision. System 2 can serve an executive function while reflecting on the decision made by System 1. For example, a clinician may make a fast diagnosis based on pattern recognition from past experiences and knowledge but after further consideration recognize his or her first impression was wrong. In doing so, the person engaged the analytic mode to acknowledge the mistake and find the correct solution. In turn, System 1 can supersede the rational decision made by System 2. The previous example of a clinician diagnosing an off-putting patient with Borderline Personality Disorder is a common example. In these instances, intuition is a powerful force and can result in an irrational or affect-based decision.

Factors that Influence Diagnostic Decisions

Ideally, the above cognitive theories would capture the clinical diagnostic decision process; however, there are factors unique to clinical practice not accounted for in most theories. The following section provides an overview of findings about the influence of patient and clinician dispositional factors as well as features of the clinical setting and the clinician's implicit biases.

Clinician Characteristics

Traditionally, patient characteristics are the primary focus in research about factors that affect diagnostic accuracy (Eisenberg, 1979). However, Callaghan (2012) conducted focus groups with general practitioners who ranked the characteristics of the general practitioner as being more important than patients' characteristics as potential biases in their decision-making. This may represent a shift in the field to investigate factors influencing both sides of the clinical relationship in recent research. Specifically, characteristics such as education, training, theoretical orientation, age, gender, and personality are thought to influence clinicians' decision-making process (Eisenberg, 1979; Garb, 2005). Woodward and colleagues (2009) investigated whether clinicians' gender, age, and theoretical orientation influenced the diagnosis given to case vignettes containing an equal number of symptoms for Posttraumatic Stress Disorder (PTSD) and Borderline Personality Disorder (BPD). While client gender, clinician gender and age had no influence, diagnosis did differ across theoretical orientation. Psychodynamic clinicians diagnosed BPD more often than PTSD compared to cognitive behavioral clinicians typically selecting PTSD more than BPD. These findings suggest clinicians are applying their theoretical perspective to interpreting diagnostic criteria and making diagnostic decisions. The diagnosis of BPD and the associated interventions could potentially be harmful for an individual experiencing PTSD and vice versa for a client experiencing BPD. It is concerning that clinicians' hypothetical conceptualization of their client's symptoms will impact the client's diagnosis, then future treatment plans and efficacy of interventions.

Culture. Research has provided evidence that clinicians' culture also influences clinical judgement because culture is a part of the context within which they engage in diagnostic decision-making. East Asians tend to take a holistic approach to judgment processes by gathering and holding various types of details, while Westerners typically employ an analytical approach and only hold details that assist in making categorical judgments (Nisbett, Peng, Choi, & Norenzayan, 2001). Additionally, culture can influence how individuals explain behavior with Westerners making more internal attributions compared to East Asians when exploring the interaction of internal attributes and situational factors for an explanation (Choi, Choi, & Norenzayan, 2004). These differences across culture have implications for mental health providers' diagnostic decision-making processes. For example, western clinicians may focus their questions to determine whether a patient's symptoms fit within the diagnostic criteria for a disorder. In comparison, an East Asian clinician may gather information about both the patient's internal symptoms and situational information to support more holistic assumptions about the client's mental health.

Additionally, the majority of psychiatrists and psychologists reported having difficulty applying existing diagnostic classification systems across cultures or when the patient is of a different culture or ethnic background than their own (Evans et al., 2013; Reed et al., 2011). Moreover, "two-thirds of global psychiatrists indicated that they prefer a system of flexible guidance that would allow for cultural variation and clinical judgment as opposed to a system of strict criteria" (Reed et al., 2011, p. 129). Despite major differences across cultures in decision-making and self-reports from clinicians that diagnostic classification systems are not always culturally appropriate, there are limited findings about how clinicians' culture impacts clinical judgment and diagnostic accuracy.

Professional Experience. Among professional psychologists, there is a perception that through continuous education, training and clinical experience their clinical judgement accuracy will improve. Out of 39 factors that potentially influence diagnostic decision-making, the previously mentioned focus group of general practitioners ranked clinical experience as 3rd most important (Callaghan, 2012). However, there is a lack of evidence supporting this perception. Instead, research has shown that practitioners with more training and experience only marginally perform better than novices in making diagnostic decisions (Ægisdóttir et al., 2006; Garb, 2005; Spengler et al., 2009; Strasser & Gruber, 2004; Witteman & Tollenaar, 2012; Witteman & Van den Bercken, 2007). Moreover, studies about clinicians with intermediate levels of experience revealed they tend to perform faster and worse than both novices and experienced clinicians (Witteman & Van den Bercken, 2007). Most concerning is the evidence showing that while clinicians recalled more details than undergraduate psychology students, they also reported significantly more false details about the case vignettes (Webb, Keeley & Eakin, 2016).

One possible explanation for this phenomenon is that it is difficult to learn and improve from clinical experience (Garb, 2005). After participating in graduate programs, clinicians rarely receive feedback about the quality and accuracy of their work. Without accurate feedback about their diagnostic decision-making, clinicians are unable to determine whether their judgement was accurate or not. Feedback about accuracy not only informs past decisions but also influences how future decisions are made. Heuristics and the dual-process model both rely on feedback to improve cognitive processes. However, clinicians rarely have confirmation of their patient's diagnosis or information about long-term outcomes.

Features of the Clinical Setting

Keller and colleagues (1986) evaluated the treatment quality of patients diagnosed with major depressive disorder across five medical centers. An unintended finding from their study was that at which medical center patients were located significantly predicted the amount and type of treatment (medicine, electroconvulsive therapy, psychotherapy) better than patients' characteristics and depression symptoms. These findings brought into question whether the location of patients' treatment along with patients' characteristics is influential on clinical judgment rather than solely which treatment is empirically supported.

Potential explanations for Keller et al.'s (1986) findings include clinician dispositional factors or the time pressures and heavy work load typically placed on clinicians at hospital psychiatric units. In stressful settings with time constraints, individuals tend to rely more on heuristics or the System 1, intuitive path of the dual-process model to free up needed cognitive capacities (Garb, 2010). However, overreliance on heuristics or System 1 can lead to acting on biases resulting in misdiagnoses and providing poor clinical care (Stepanikova, 2012). Additionally, time pressure can also result in individuals defaulting to their culture's pattern of attribution (i.e., internal or external) when required to make judgements under time pressure (Chiu et al., 2000). For example, North American clinicians would be more likely to rely on internal dispositional attributes when making clinical judgements while Chinese peers engage in a holistic examination of influential attributes. This could be detrimental to treatment and the therapeutic relationship when the patient's and therapist's culture are dissimilar and under time pressure.

Additionally, time constraints have an influence on clinicians' tendency to engage in biases (Muroff, Jackson, Mowbray, &, Himle, 2007; Witteman & Tollenaar, 2012). Muroff and colleague's (2007) study provided evidence supporting the idea that limited time increased the probability that psychiatrists may apply social stereotypes to the diagnostic decision-making process. In their study, psychiatrists were more likely to make a bipolar disorder diagnosis—the correct decision—than a depression disorder diagnosis, when there was no time restriction and low case load. Stepanikova (2012) experimentally manipulated time pressure to stimulate high patient load and limited time to make medical decisions. Physicians participating in the study reported higher stress levels and depleted cognitive resources. Findings from this study provided evidence that negative implicit bias influences medical decisions more under time pressure, putting racial minority patients at a disadvantage and potentially contributing to racial disparities in health care.

Patient Characteristics

The diagnostic decision-making process typically involves forming a hypothesis based on the patient's salient symptoms, characteristics, and contextual factors. While some of these factors are used to organize diagnoses, others, like patient characteristics, can influence clinicians to act on implicit biases (e.g., race bias, social class bias, gender bias, age bias) and over-pathologize the patient (Eisenberg, 1979; Garb, 2010; Solomon, 1992). This typically occurs when a symptom or characteristic is reinterpreted or discounted by the clinician so that the patient fits a specific prototype or representation of a disorder (Kirmayer, 2005).

Racial Bias. Research has shown that medical providers perceive African American patients to be less intelligent, less educated, less likeable, less friendly, and less likely to adhere to medical advice (Sabin et al., 2009; Stepanikova, 2012). In return, African Americans perceive discrimination in health care settings and express more mistrust when they felt their symptoms were discredited due to poor communication (Cuevas, O'Brien, & Saha, 2016). Over the years, Garb (1996, 2005, 2010) has provided and reviewed substantial evidence about the effect of racial bias in the diagnosis of mental disorders. Specifically, errors in diagnosis occur when mental health practitioners attend to patients' race rather than prevalence rates of disorders.

Racial bias is most evident in psychiatric settings and the prescription of psychotropic medication. For example, African-American and Hispanic patients are more likely than white patients to be diagnosed with schizophrenia and less likely to be diagnosed with affect disorders when psychopathology assessments do not support a schizophrenia diagnosis (Garb, 1996; Neighbors, Trierweiler, Ford & Muroff, 2003). Neighbors and colleagues (2003) found that inappropriate affect and auditory hallucinations increased the likelihood of a schizophrenia diagnosis for African American patients but not for white patients. Conversely, catatonic behaviors increased the likelihood of a bipolar diagnosis for white patients more often than African American patients. In 2010, Garb provided a review of studies that cumulatively showed that African American patients were less likely to receive necessary lithium, SSRIs, and secondgeneration antipsychotics than white patients even though both groups received a diagnosis of bipolar disorder or schizophrenia, respectively. Additionally, racial minority patients receive a significantly higher number of psychotropic medications, antipsychotic injections, and higher dosages compared to white patients in psychiatric settings after controlling for level of functioning, presenting symptoms, risk assessment, and medical history (Segal, Bola, & Watson, 1996). One explanation is that clinicians spend significantly less time with African American patients than other patients. When clinicians spent more time evaluating African American patients, differences across racial groups in antipsychotic prescriptions decreased.

Gender Bias. Like with race, research has shown that patients' gender influences diagnosis; however, there is mixed evidence about how gender bias influences diagnosis (Chiaramonte & Friend, 2006; Garb, 1997). Past research suggested that mental health professionals viewed female clients as less psychologically healthy than male clients (Garb, 2010; Widiger & Settle, 1987). More recent research refutes this idea and provides evidence that clinicians do not view male clients as more psychologically heathy than female clients (Kelley & Blashfield, 2009).

Mixed evidence comes from research on client-therapist gender matching and in the diagnosis of mood and personality disorders (Muroff et al., 2007; Woodward, Taft, Gorgon & Meis, 2009). For example, a study on differential diagnosis of PTSD and BPD found that female clinicians diagnosed PTSD rather than BPD for both male and female clients more than male clinicians. Differential diagnosis of personality disorders also lends itself to gender bias because gender is closely tied to stereotypes about personality disorder; specifically, women are diagnosed more commonly with histrionic personality disorder while men are more likely to receive the diagnosis of antisocial personality disorder (Garb, 2010). As for mood disorders, female patients are more likely to receive a major depressive disorder or bipolar disorder diagnosis compared to male patients who more often received a psychotic disorder diagnosis (Garb, 1997; Muroff et al., 2007).

Response latency

One of the most common ways to measure decision-making processes, especially those that include biases, is response latency. Measuring response latency, or how long it takes a respondent to make a decision, has been used as an indicator of unconscious cognitive processes in psychology research. Response latency is the computation of the time that elapses between stimulus presentation and response production (Lane, Banaji, Nosek, & Greenwald, 2007). The advantage of response latency is its ability to produce orderly data about a participant's underlying cognitive process by circumnavigating introspection (Lane et al., 2007). Time as a variable of unconscious processes is the basis of several cognitive tasks that examine how individuals form constructs and process information: the Stroop task, episodic or repetition priming, semantic priming, evaluation priming, and notably the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998). The response latency literature also states that the easier a mental task is, the quicker one reaches a decision point with fewer errors.

Response latency can provide information about whether individuals draw conclusions from information provided or only summarize the information without making judgements. In a series of experiments, half of participants were given directions to develop attitudes towards novel objects (Fazio, 1990). Afterwards, they were asked to complete questionnaires about the novel objects. Response times showed that the participants who were asked to form an opinion earlier responded more quickly to questions about their attitudes than participants who were not asked to form an opinion. These findings suggest that individuals who have previously consolidated available information into a judgment respond quicker to future questions about their attitudes compared to individuals who have not consolidated the information, thus needing more time to develop their opinion to answer the questions. These results can be generalized to describe clinicians' diagnostic decision-making process. The level of exposure a clinician has with individuals with a diagnosis and experience making the diagnosis would influence how much time elapsed during the diagnostic process, with more familiarity leading to clinicians making quicker diagnoses. Clinicians who lack exposure to a diagnosis would be less likely to rely on judgments and instead would engage in a hypothesis testing approach to diagnostic decision-making.

Response time has been shown to be an efficient way to assess social judgement or implicit associations. For example, the IAT measures strengths of associations between words with positive or negative connotations and individual characteristics (e.g., race, gender, age, weight) by comparing response times in two combined discrimination tasks. The basic assumption of the IAT is that, if two concepts are highly associated (i.e., positive words and young people), the sorting tasking will be easier and response times faster (Nosek, Greenwald, Banaji, 2006; Greenwald et al., 1998). It could be said that one has an implicit preference for younger people relative to older adults if the respondent was faster to categorize negative words with older adults than pairing positive words with older adults (Greenwald et al., 1998).

IAT tasks have been used to assess the unconscious biases of mental health and medical service providers. Peris, Teachmen, and Nosek (2008) examined the presence of explicit and implicit biases about mental illness among individuals with different levels of mental health training using the IAT and a scale overtly measuring attitudes toward mental illness. Findings suggest that people with more clinical experience have more positive implicit and explicit evaluations of individuals with mental illness. Additionally, implicit negative biases towards people with mental illness significantly predicted over-diagnosis while explicit negative biases predicted negative patient prognoses. These findings suggest that both explicit and implicit biases have negative consequences for the quality of clinical care individuals receive.

Limitations

While the field of cognitive psychology has extensively studied how individuals engage in the decision-making process, there are limitations of studies researching clinician diagnostic decision-making. For the most part, samples of clinicians are predominantly female and do not report additional demographic information (e.g., clinicians' ethnicity, culture, training, or clinical experience with diagnostic manuals) that could be influential in the process of making a diagnostic decision (Ægisdóttir et al., 2006; Callaghan, 2012; Garb, 2005; Spengler et al., 2009; Strasser & Gruber, 2004; Witteman & Tollenaar, 2012; Witteman & Van den Bercken, 2007).

These studies are not capturing representative samples of the clinician population and instead are studying participants that more closely represent people from Western, Educated, Industrialized, Rich and Democratic (WEIRD) societies (Henrich, Heine, & Norenzayan, 2010). These WEIRD participants are typical subjects and make up the bulk of participants in psychology research and research about cognitive decision-making. However, they are "outliers within an outlier population;" they are the exception when American psychologists and psychiatrists comprise only an estimated 14.3% and 2.4% of the global clinician population (Evans et al., 2013; Heinrich et al., 2010, p. 18; Reed et al., 2011). It cannot be assumed that diagnostic decision-making and accuracy of WEIRD clinicians is generalizable across the clinician population, especially given the anticipated effect of culture on diagnostic decision-making. Additionally, lack of information about participating clinicians brings into question the generalizability of findings on the impact of dispositional factors on psychodiagnostic decisions. It is possible that the previously discussed association between level of experience and accurate diagnostic decision-making differs across training experiences, especially when comparing clinicians who completed training in different countries which may have alternative requirements for licensure or independent practice.

Another problem that arises from not reporting clinician demographic information is that researchers are not considering the influence of clinician characteristics. Thus, there are limited findings discussing the influence clinicians have on their own diagnostic accuracy. As previously stated, this gap in the literature places onus on patient dispositional factors affecting diagnostic accuracy rather clinicians examining the influence their own dispositional factors have on their work and the interaction between clinician and client dispositional factors.

Last, the field of clinical cognition and diagnostic accuracy heavily studies errors made by both experts and novices (Elstein & Schwarz, 2002). Despite this focus on mistakes, the prevalence rate of diagnostic errors among clinicians is unknown and instead there may be an overestimation of the rate at which clinicians make mistakes. While errors provide a wealth of information about an individual’s cognitive process, these studies do not accurately represent how skilled clinicians are at determining diagnoses. From a training and clinical practice point of view, in addition to identifying errors and the cognitive processes that may lead to them, it would be beneficial to provide feedback about what clinicians are doing correctly during the diagnostic process.

Current Study

Previous research has focused on understanding how medical professionals make diagnostic decisions and whether medical professionals make biased diagnostic decisions based on patient characteristics. To date, there is limited information about the relation between clinician characteristics and diagnostic accuracy. For example, Sabin et al. (2009) examined physicians' implicit racial attitudes considering physicians' race and gender using the IAT and found that African American physicians showed less implicit preference for either Blacks or Whites and women showed less implicit bias than men. However, the study did not explore whether these implicit attitudes about race affect the diagnosis these physicians provided.

Additionally, most findings presented in this review lacked important demographic information about clinicians and/or its impact on diagnostic decision-making. This study aimed to expand on past research by using existing data with a sample that is more inclusive of the global clinician population, specifically regarding gender and ethnicity. This more diverse sample will allow for examining the influence of various clinician characteristics (e.g., gender, experience, world region).

The current study aimed to elucidate the relation between clinician characteristics, response time, and diagnostic accuracy. Because various clinician characteristics are associated with implicit attitudes and diagnostic decisions, response time and diagnostic accuracy were explored as outcome variables. In addition to the influence of clinician characteristics on response time and diagnostic accuracy, the current study explored how the combination of response time and clinician characteristics influenced diagnostic accuracy. This was accomplished by using secondary data analysis on an existing data set that included clinician demographics and time elapsed while completing an internet survey asking participants to diagnose case vignettes (Keeley et al., 2016).

Specific Aims

This study examined the relation between clinician demographics (i.e., gender, years of training and experience, profession, language, and world region), response time, and degree of accuracy diagnosing case vignettes. The following paragraphs further outline the specific aims of the study.

Aim 1. The first aim of the study is to investigate whether clinician characteristics and response time predict diagnostic accuracy. To predict the correctness of diagnostic decisions, ordinal logistic regressions will be used. Based on past research, it was hypothesized that:

1a. clinician gender will have no influence on diagnostic accuracy,

1b. clinician experience will predict diagnostic accuracy,

1c. the interaction between patient's and clinician's gender will predict clinicians' accuracy,

1d. response time will have a bimodal relationship with diagnostic accuracy with moderate response times being the least accurate, longer response times being the most accurate, and quicker response times somewhere in-between.

Last, it is unclear what influence clinicians' WHO region, primary language, and diagnostic classification system will have on diagnostic accuracy. We treated these hypotheses as exploratory analyses.

Aim 2. The second aim of the current study is to determine whether clinician characteristics predict clinicians' response time. The predictive ability of clinician characteristics will be assessed using simple linear regression. Based on past research:

2a. clinician gender alone will not influence response times,

2b. congruent clinician gender and case vignette gender will be related to responding faster than when gender is discordant,

2c. due to differences in internet connectivity and local infrastructure, clinicians' WHO region may impact clinicians' response rate.

2d. clinicians' experience level is expected to have an inverse relationship with their response time with more experienced clinicians responding faster,

2e. there is limited evidence about the influence of age, language, clinical profession, and amount of experience with either the DSM-5 or ICD-10; these factors may also influence response time.

Aim 3. Based on the results of the first two aims, significant predictors will be incorporated into more sophisticated models of mediation to determine how clinician characteristics, response time and diagnostic accuracy interact. These models will be assessed using Hayes' PROCESS macro program for SPSS (Hayes, 2013). For example, if match between clinician and case vignette genders significantly predicted response time and response time predicted diagnostic accuracy, a mediation model would be used to assess how these variables interacted.

Method

Design

A secondary data analysis will be performed using within-subject data from the World Health Organization (WHO) Department of Mental Health and Substance Abuse field trials of Schizophrenia and Other Primary Psychotic Disorders to potentially be included in the Eleventh Revision of Internal Classification of Diseases and Related Health Problems (ICD-11). This dataset included a sample of global clinicians who responded to an online questionnaire requiring them to read two vignettes and make diagnostic decisions using either the ICD-10 or ICD-11 as described in the following section (see Keeley et al., 2016 for additional details).

Participants

Clinicians were drawn from the Global Clinical Practice Network (GCPN), a worldwide network of mental health professionals, to participate in the ICD-11 field studies. Mental health professionals were invited to join the GCPN through professional listservs; national and regional professional associations; international and national conferences in psychology, psychiatry, and related disciplines; and professional word-of-mouth. The dataset used in this study consisted of 2,330 mental health professionals representing 100 different nationalities who completed the survey in Chinese, English, French, Japanese, Russian, or Spanish. When the study launched, 9,323 GCPN members qualified for the study and were invited to participate. Of those individuals, 6,637 did not respond to or finish the study. Additionally, participants were required to have provided diagnoses for two vignettes and have no missing or extreme timing variables (i.e., within three standard deviations). Based on these inclusion criteria, we removed a total of 863 participants from the dataset. The final sample for this study included 1,822 clinicians. Demographic information can be found in Table 1. Clinicians also reported the degree (e.g., routinely, often, sometimes) to which they used various diagnostic classification systems (see Table 2). The majority of participants (55.5%) stated that they routinely use some version of the ICD.

Table 1. Participant demographics

	f (%)
Region	
AFRO	36 (2.0)
AMRO-North	208 (11.4)
AMRO-South	229 (12.6)
EMRO	39 (2.1)
EURO	796 (43.7)

SEARO	92 (5.0)
WPRO-Asia	367 (20.1)
WPRO-Oceania	45 (2.5)
Other	10 (0.5)
Income Level	
Low/Lower-middle	166 (9.1)
Upper-middle	488 (26.8)
High	1158 (63.6)
Primary Language	
Arabic	7 (0.4)
Chinese	223 (12.2)
English	747 (41.0)
French	156 (8.6)
German	57 (3.1)
Japanese	137 (7.5)
Russian	245 (13.4)
Spanish	233 (12.8)
Portuguese	17 (0.9)
Gender	
Male	1015 (55.7)
Female	807 (44.3)
Profession	
Counseling	68 (3.7)
Medicine	1114 (61.1)
Nursing	23 (1.3)
Psychology	521 (28.6)
Social work	16 (0.9)
Sex Therapy	10 (0.5)
Speech Therapy	4 (0.2)
Other	9 (0.5)
	M (*SD*)
Age	44.92 (11.52)
Years of Training	7.3 (5.45)
Years of Experience	15.16 (10.55)
Total N	1822

Note : APRO: African Regional Office, AMRO-North: North American Regional Office, AMPRO-South: South American Regional Office, EMRO: Eastern Mediterranean Regional Office, EURO: European

Regional Office, SEARO: South-East Asia Regional Office, WPRO-Asia: Asian Western Pacific Regional Office, WPRO-Oceania: Oceania Western Pacific Regional Office.

Table 2. Diagnostic classification frequency of use

	ICD *f* (%)	DSM *f* (%)	Other *f* (%)
Routinely	1011 (55.5)	615 (33.7)	27 (1.5)
Often	239 (13.1)	305 (16.7)	26 (1.4)
Sometimes	198 (10.9)	312 (17.1)	20 (1.1)
Rarely	107 (5.9)	154 (8.4.)	6 (0.3)
Never	94 (5.2)	102 (5.6)	127 (7.0)
Total (*n*)	1649 (90.5)	1488 (81.4)	206 (11.3)

Procedure

In the ICD-11 electronic field trial for psychotic disorders, participants received an email invitation to complete a survey which was administered through Qualtrics, a web-based survey program. This survey and GCPN registration collected participants' demographic information—including clinical profession, experience level, and area(s) of expertise—the final diagnostic decisions for two vignettes about clients who potentially had psychotic disorders, the presence or absence of each diagnostic guideline for the diagnosis they selected, and the amount of time participants took to make these diagnoses. The current study will use data about clinicians' demographics and clinical background, response time, and the accuracy of diagnoses from this dataset to investigate the aims of the current study.

Measures

Clinician Characteristics. Demographics and information about participants' professional and clinical work were collected as part of their registration with the GCPN. The primary variables of interest for the present study include clinicians' self-reported gender, WHO region, years of experience, primary diagnostic classification system (DCS; e.g., ICD-10, DSM-IV), and primary language. Additionally, the pairing of clinicians' gender and the gender of the case vignettes will be explored.

Response Latency. Electronic administration through Qualtrics collected information about how long participants spent on each page (e.g., viewing the diagnostic guidelines, reading the vignettes, and determining a diagnosis) by capturing at what time they first clicked on the page, when they last clicked, and when they submitted the page. For this study, total response time was measured by adding how long participants spent on each page of the survey, in seconds. How much time participants spent providing a diagnosis was also measured as a percentage of their total time completing the survey. Response time could be influenced both by speed of Internet connectivity and individual differences in average response latency (Nosek et al., 2006; Reed et al., 2011). As such, we used the income level of participants' self-reported country as a control variable to determine if there are significant differences in response latency across regions.

Diagnostic Accuracy. After viewing a vignette, participants gave a diagnosis for a vignette and then answered inquiries about the presence or absence of diagnostic guidelines for the diagnosis they selected. Then, clinicians were given the option of choosing a different final diagnosis. If their final diagnosis was incorrect, participants received an additional differential diagnosis question, inquiring to why the person did not assign the correct diagnosis. To determine diagnostic accuracy, participants' final diagnoses were categorized as either correct or incorrect for each of the two vignettes. Correct diagnoses for the vignettes were determined by the study investigators and confirmed by an independent panel of content experts. Participants' diagnostic accuracy for the two vignettes were combined to measure clinicians' overall degree of diagnostic accuracy (such that they could have 0, 1, or 2 diagnoses correct). When independent variables were not specific to either vignette (i.e., overall response time and clinician demographic variables), we used this combined diagnostic accuracy variable to reflect participants' overall accuracy across the study. When independent variables were specific to one diagnosis or another (i.e., time spent on diagnosis, gender interaction with the vignette), we used the individual diagnostic accuracy for that vignette (either correct [1] or incorrect [0]).

Results

Data Preparation and cleaning

We used SPSS 24.0 for all data analyses. Existing variables in the dataset were used to derive clinician demographics, diagnostic accuracy, and response time. Descriptive statistics were calculated prior to testing study hypotheses to verify that data met the assumptions of the planned analyses.

A review was conducted to assess skewness, kurtosis, and outliers for all variables of interest. Skewness and kurtosis values for age, country's income level, years of experience, primary language, diagnostic accuracy on vignette one, gender interaction on vignette two, overall diagnostic accuracy, and diagnostic classification frequency of use were close to or below an absolute value of 1, indicating these variables were approximately normally distributed. Diagnostic accuracy for vignette two, participants' gender, and gender interaction on vignette one were platykurtic. Years of training, participants' profession, overall response time, time spent diagnosing vignette one and vignette two, along with the percentage of the total time spent diagnosing each vignette were positively skewed and were leptokurtic. All response time variables were transformed using the standard logarithm. We used these transformed response time variables for correlational analyses, regression analyses, and follow-up analyses.

Demographic information does not contain transformed values. Due to low sample size, we consolidated participants' clinical profession into three groups: medicine (61%), psychology (29%), and other (10%). Additionally, due to small sample sizes, participants who did not report their country or region (n = 10) or reported their primary language as either Arabic (n = 7) or Portuguese (n = 17) were excluded from analysis using region or language as a predictor variable. We explored whether the income level of participants' country might be a covariate for participants' response time; due to small sample size, we combined low and lower-middle income levels into one group and excluded participants (n = 10) who did not report their country from analyses. Country income level did not significantly predict clinicians' overall response time suggesting the infrastructure and internet connective at the level of participants' country did not significantly impact their response times during the study (b = -.013, $t(1810) = -1.079$, p = .281).

In addition to a review of skewness and kurtosis, we assessed assumptions of independence, multicollinearity, and homoscedasticity. All variables sufficiently met these assumptions except minimally significant correlations between predictor and criterion variables (Table 3). As expected there were substantial, significant correlations among time variables; however, none were used conjointly in the following analyses. All other significant correlations were less than r = .10 and were considered minimal. These significant correlations likely resulted from the large sample size. Analyses were conducted as planned because multicollinearity was minimal.

Table 3. Correlation coefficients among predictor and criterion variables

	Overall Diagnostic Accuracy	Diagnosis 1 Accuracy	Diagnosis 2 Accuracy	Total Time	Percent of Time for Diagnosis 1	Percent of Time for Diagnosis 2
Time on Diagnosis 1	-.016	.004	-.025	.285**	.809**	.077**
Time on Diagnosis 2	.035	.026	.024	.307**	.090**	.755**
Total Time	.081**	.001	.072**	1.00	-.333**	-.392**
% of Total Time for Diagnosis 1	-.066**	-.020	-.069**	--	1.00	.316**
% of Total time for Diagnosis 2	-.022	-.002	-.027	--	--	1.00

Gender	-.033	-.024	-.022	.007	.024	-.038
Interaction 1	-.032	-.020	-.025	-.009	.029	.016
Interaction 2	-.081**	-.042	-.069**	.015	-.045	-.008
Age	-.032	-.008	-.035	.052*	.004	-.012
Region	.013	-.004	.020	-.120**	.048*	.098*
Primary Language	-.013	-.012	-.006	.025	.049*	.063**
Years of Training	-.038	-.004	-.047*	-.013	-.001	.045
Years of Experience	-.012	-.012	-.022	.054*	.003	.004
Profession	-.048*	-.042	-.025	.075**	-.027	-.023
ICD	<.001	-.002	.001	-.006	.033	.012
DSM	.045	.011	.05	.040	-.004	-.018
Other DCS	.115	.074	.083	-.086	-.042	-.125

Note. * $p < .05$, ** $p < .01$

Aim 1.

A series of ordinal logistic regressions were performed to ascertain whether clinicians' demographics predicted the likelihood that participants would correctly diagnose the vignettes. Gender Interaction 1 and Gender Interaction 2 were used as predictor variables in separate logistic regression models for diagnostic accuracy for vignette one and vignette two, respectively. All other clinician demographics were used as predictors of participants' combined diagnostic accuracy on both vignettes. Each predictor variable was entered into its own model.

The gender interaction for the second vignette was a statistically significant predictor of clinicians accurately diagnosing the second vignette, $\chi^2(1) = 8.863$, $p = .003$. Congruent clinician-vignette gender led to correctly diagnosing the vignette 7.7% more often than incongruent clinician-vignette pairs (Table 4). The gender interaction accounted for 0.7% of the variability in diagnostic accuracy ($b = .314$, 95% CI [.106, .521]). Other clinician characteristics did not significantly predict diagnostic accuracy (Table 5).

Table 4. Vignette two cross tabulation of gender interaction and diagnostic accuracy

	Diagnostic accuracy		
	Incorrect	Correct	Total

Gender	Congruent	196 (36.9%)	335 (63.2%)	531
Interaction	Incongruent	574 (44.5%)	717 (55.5%)	1291
	Total	770	1052	1822

Table 5. Logistic regressions of diagnostic accuracy with clinician demographics

	-2 Log Likelihood	χ^2	*df*	*p*	Nagelkerke R^2
Gender	30.91	1.314	1	.252	.001
Interaction 1*	14.03	.709	1	.400	.001
Interaction 2**	14.26	8.86	1	**.003**	.007
Age	454.98	1.33	1	.248	.001
Region	91.75	2.57	7	.922	.002
Primary Language	80.76	2.73	6	.842	.002
Years of Training	263.30	2.60	1	.107	002
Years of Experience	384.99	.768	1	.381	<.000
Profession	37.99	3.63	2	.163	.002
DCS					
ICD	54.55	6.42	4	.170	.005
DSM	62.01	3.19	4	.527	.003
Other	37.67	5.98	4	.200	.034

Note. *Accuracy diagnosing the first vignette was the dependent variable rather than overall diagnostic accuracy, ** Accuracy diagnosing the second vignette was the dependent variable rather than overall diagnostic accuracy, DCS = diagnostic classification system.

To assess whether clinicians' overall response time and the time elapsed while providing a diagnosis predicted their degree of diagnostic accuracy, a series of ordinal and binary logistic regressions were conducted. Total time elapsed was used as a predictor of participants' combined diagnostic accuracy on both vignettes. The percentage of time that was spent diagnosing each vignette was used to predict diagnostic accuracy on vignettes one and two separately. Each predictor variable was entered into its own model. The total time clinicians spent participating in the study significantly predicted their overall diagnostic accuracy, $\chi^2(1) = 12.07$, $p = .001$.

Participants' overall response time accounted for 0.8% of the variability in diagnostic accuracy (Nagelkerke R^2 = .008). Tukey HSD analysis revealed that participants who correctly diagnosed both vignettes (*M* = 1,954.12, *SD* = 1,662.87) spent significantly more time completing the study than participants who correctly diagnosed one (*M* = 1,884.52, *SD* = 1,898.14, *p* = .020) or neither of the vignettes (*M* = 1,719.73, *SD* = 1,659.67, *p* = .008). However, time spent on either individual diagnosis was not directly predictive of its respective accuracy variable (Table 6).

Table 6. Logistic regressions of diagnosis accuracy with time elapsed in seconds

	M	*SD*	-2 Log Likelihood	χ^2	*df*	*p*	Nagelkerke R^2
Total Elapsed	1,896.16	1,778.43	3474.29	12.07	1	**.001**	.008
Percent of Time for Diagnosis 1*	4.27%	6.04%	2104.62	.726	1	.394	.001
Percent of Time for Diagnosis 2**	5.06%	6.28%	2480.67	1.34	1	.248	.001

Note. *Accuracy diagnosing the first vignette was the dependent variable rather than overall diagnostic accuracy, ** Accuracy diagnosing the second vignette was the dependent variable rather than overall diagnostic accuracy.

An additional chi-square analysis assessed whether total response time had a bimodal relationship with diagnostic accuracy (Aim 1d). Total time elapsed was divided into three categories—fast, moderate and slow responders—to determine whether moderate response times would be less accurate than longer or quicker response times. Total time elapsed was divided into thirds, with fast responders making up the lower third of response time, slow responders make up the top third of respond time and with moderates falling between slow and fast responders. Categorical response time was a significant predictor of overall diagnostic accuracy, $\chi^2(4) = 20.51$, $p < .001$. Although the relationship between diagnostic accuracy and responders' speed was linear rather than bimodal, there was a significant interaction effect. Cross tabulation showed that fast responders were more likely to incorrectly diagnose both vignettes than slow or moderate responders (Table 7). There was no significant difference across responders' speed in their rates of correctly diagnosing both or one vignette.

Table 7. Overall diagnostic accuracy by level of response time

		Diagnostic accuracy			
		Both Incorrect	One Correct	Both Correct	Total
Speed of Responders	Fast	73 (12.2%)	320 (53.3%)	207 (34.5%)	600
	Moderate	62 (10.3%)	269 (44.8%)	270 (44.9%)	601
	Slow	59 (9.5%)	275 (44.3%)	287 (46.2%)	621
	Total	194	864	764	1822

Aim 2.

A series of simple linear regressions were performed to examine the relation between clinician characteristics and response time. Each predictor variable was entered into its own model. Clinicians' age (b = .002, $t(1820) = -2.21$, $p = .028$) and years of experience (b = .002, $t(1820) = 2.30$, $p = .022$) were statistically significant predictors of total response time. The amount of time participants spent completing the study increased as both age and years of experience increased explaining small portions of the variance in response time (age $R^2 = .003$ and experience $R^2 = .003$). Additionally, participants' global region and clinical profession significantly predicted their total response time, respectively $b = -.024$, $t(1810) = -5.15$, $p < .001$ and $b = .028$, $t(1820) = 3.21$, $p = .001$. Participants' region and profession explained small but significant portions of the variance in time elapsed completing the study (region $R^2 = .014$ and profession $R^2 = .006$). Post-hoc tests using Tukey HSD revealed that participants in the WPRO-Asia region (*M* = 1,512.14, *SD* = 1,442.30) spent significantly less time than those in AMRO-North (*M* = 1,988.87, *SD* = 1,977.42, $p = .001$), AMRO-South (*M* = 2,227.09, *SD* = 1,902.83, $p < .001$), and EURO regions (*M* = 1,960.68, *SD* = 1,806.25, $p < .001$). There were no other significant differences between regions in total time spent responding to the survey. Tukey HSD also showed that participants practicing medicine (*M* = 1,796.56, *SD* =1,673.02) were significantly quicker than those practicing psychology (*M* = 2,117.85, *SD* = 2,083.33, $p = .002$).

There was not enough evidence to support that other clinician demographics significantly predicted their total response time (Table 8).

Table 8. Simple linear regressions of total response time with clinician demographics

	F	*df*	*p*	Variance (R^2)	*b*
Gender	.100	1, 1820	.751	<.001	.005
Interaction 1	.145	1, 1820	.703	<.001	-.006
Interaction 2	.414	1, 1820	.520	<.001	.011
Age	4.86	1, 1820	**.028**	.003	.002
Region	26.51	1, 1810	**<.001**	.014	-.024
Years in Training	.306	1, 1820	.580	<.001	-.001
Primary Language	1.09	1, 1796	.297	.001	.003
Years of Experience	5.29	1, 1820	**.022**	.003	.002
Profession	10.29	1, 1820	**.001**	.006	.028
DCS					
ICD	.055	1, 1820	.815	<.001	-.002
DSM	2.36	1, 1820	.125	.002	.011
Other	1.52	1, 1820	.219	.007	-.019

Note: DCS = diagnostic classification system

Further simple linear regressions were conducted on the percentage of the total time participants spent diagnosing each vignette. Of the total time participants spent completing the study, on average they spent 4.27% and 5.06% diagnosing the first and second vignette, respectively. Participants' reported region and primary language significantly predicted the portion of time they spent diagnosing the first vignette, respectively $b = .016$, $t(1820) = 2.05$, $p < .041$ and $b = .011$, $t(1820) = 2.09$, $p = .036$. Region and language accounted for a small percentage of the variance in time diagnosing vignette one (region $R^2 = .002$ and language $R^2 = .002$). After correcting for family-wise error rates using Tukey HSD, there were not any significant differences across regions in participants' percent of total time diagnosing vignette 1.

In other words, the differences across regions leading to the significant overall effect were too small to be statistically reliable. Participants who reported their primary language as English (*M* = 3.95, *SD* = 6.14) spent a smaller percentage of their total time diagnosing the first vignette than those who primarily used French (*M* = 5.13, *SD* = 6.89, *p* = .010) and Japanese (*M* = 5.45, *SD* = 6.00, *p* < .001). Participants' whose primary language was Russian (*M* = 4.55, *SD* = 6.59) also spent a significantly smaller percentage of their total time diagnosing their first vignette than who primarily used Japanese (*M* = 5.45, *SD* = 6.00, *p* = .046). Other clinician demographics did not significantly predict the amount of time participants spent diagnosing the first vignette (Table 9).

Table 9. Simple linear regressions of portion of time spent diagnosing vignette one

	F	*df*	*p*	Variance (R^2)	*b*
Gender	1.01	1, 1820	.316	.001	.026
Interaction 1	1.53	1, 1820	.216	<.001	.032
Age	.036	1, 1820	.849	<.001	<.001
Region	4.19	1, 1811	**.041**	.002	.016
Years in Training	.003	1, 1820	.957	<.001	<.001
Primary Language	.4.39	1, 1796	**.036**	.002	.011
Years of Experience	.012	1, 1820	.912	<.001	<.001
Profession	1.35	1, 1820	.246	.001	-.022
DCS					
ICD	.1.83	1, 1646	.176	.001	.015
DSM	.026	1, 1485	.873	<.001	-.002
Other	.352	1, 204	.554	.002	-.015

As with the first vignette, clinicians' region and language were significant predictors of the portion of time participants spent diagnosing the second vignette, respectively *b* = .028, *t*(1820) = 4.17, *p* < .001 and *b* = .013, *t*(1820) = 2.68, *p* = .007. Region and language accounted for a small percentage of the variance in diagnostic response time (region R^2 = .010 and language R^2 = .004). Post-hoc analyses using Tukey HSD were used to determine which regions and languages significantly differed in percent of time spent diagnosing vignette two. Participants from AMRO-North (*M* = 3.57, *SD* = 3.69) spent less time diagnosing vignette two than participants from AMRO-South (*M* = 5.31, *SD* = .483, *p* = .004). Participants from WPRO-Asia (*M* = 6.70, *SD* = 8.33) spent significantly longer diagnosing the second vignette than participants from EURO (*M* = 4.63, *SD* = 5.54, *p* < .001), EMRO (*M* = 3.05, *SD* = 3.82, *p* = .003), and AMRO-North (*M* = 3.57, *SD* = 3.69, *p* < .001).

As with vignette one, English speaking participants (*M* = 4.17, *SD* = 5.37) spent a significantly smaller percent of their total time providing a diagnosis for vignette two compared to Chinese (*M* = 6.01, *SD* = 6.72, *p* < .001), French (*M* = 5.05, *SD* = 5.07, *p* = .035), Japanese (*M* = 8.07, *SD* = 10.45, *p* < .001), and Spanish speaking participants (*M* = 5.34, *SD* = .15, *p* = .005). Additionally, participants who primarily speak Russian (*M* = 5.01, *SD* = 5.83) spent significantly less time diagnosing vignette two than Japanese speaking participants (*M* = 8.07, *SD* = 10.45, *p* = .024). Additional clinician demographics did not significantly predict their percent of time diagnosing the second vignette (Table 10).

Table 10. Simple linear regressions of portion of time spent diagnosing vignette two

	F	*df*	*p*	Variance (R^2)	*b*
Gender	2.65	1, 1820	.104	.001	-.038
Interaction 2	.113	1, 1820	.737	<.001	-.009
Age	.254	1, 1820	.614	<.001	-.001
Region	17.41	1, 1811	**<.001**	.010	.028
Years in Training	3.63	1, 1820	.057	.002	.004
Language	7.17	1, 1796	**.007**	.004	.013

Years of Experience	.037	1, 1820	.848	<.001	<.001
Profession	.932	1, 1820	.335	.001	-.017
DCS					
ICD	.224	1, 1646	.636	<.001	.005
DSM	.500	1, 1485	.480	<.001	-.007
Other	3.21	1, 204	.074	.016	-.037

Aim 3.

We conducted bootstrapping analysis to examine the indirect effect of clinician demographics on overall diagnostic accuracy via total response time using model 4 in PROCESS (Hayes, 2012). We used a mediation model, through which clinician demographics (i.e., age, years of experience, profession, and WHO region) were modeled to impact overall diagnostic accuracy through participants' response time. A separate model was run for each clinician demographic variable. We included these variables because previous analyses found that clinicians' age, years of experience, global region, and profession were a significant predictor of overall response time, which in turn significantly predicted diagnostic accuracy. We prepared the data for analyses by centering each value around the variable's mean, as a result the values reported below will look different. Participant age indirectly influenced their overall diagnostic accuracy through the positive effect it had on time spent completing the study. As shown in Table 11 and Figure 1, older participants were more likely than younger participants to take longer finishing the survey (b = .0015) and slow responders were more likely to accurately diagnose both vignettes (b = .1599). A percentile bootstrap confidence interval for the indirect effect (b = .0002) based on 5,000 bootstrap samples was above zero, 95% CI [.00002, .00052], indicating that clinician age exerts an indirect influence upon diagnostic accuracy via response time.

We conducted a second mediation model to examine the influence of clinical experience on diagnostic accuracy through total response time. Table 11 and Figure 2 show that individuals with more experience were more likely to take longer completing the study (*b* = .0017) and as with the previous model slow responders were more likely to provide correct diagnoses (*b* = .1599). A percentile bootstrap confidence interval for the indirect effect (*b* = .0003) based on 5,000 bootstrap samples was above zero, 95% CI [.00004, .00614]. As expected, age and years of experience had a significant positive strong correlation ($r = .873$, $p < .001$), with older participants reporting having more experience. Thus, the first two mediational models likely represent the same effect, given the inherent relationship between age and years of experience.

We conducted categorical mediation models to evaluate whether participants' region or clinical profession indirectly influenced their accuracy through total response time. Table 11 and Figure 3 show that participants from AMRO-South (*b* = .0660) were more likely to be slower responders while participants from WPRO-Asia (*b* = -.1178) were more likely to be faster responders in comparison to clinicians in North America. As with the previous mediation models, slower responders were more likely to accurately diagnoses the vignettes (*b* = .1508). No other regions significantly predicted response time. A percentile bootstrap confidence interval for the omnibus indirect effect of participants' region (*b* = .0037) based on 5,000 bootstrap samples was above zero, 95% CI [.0014, .0084]. More specifically, clinicians in the AMRO-South region had an indirect effect (*b* = .0100) based on 5,000 bootstrap samples that was above zero, 95% CI [.0004, .0237], meaning that participants in South America were more likely to have slower response times and thus have higher diagnostic accuracy compared to North American clinicians. Additionally, clinicians in WPRO-Asia region had an indirect effect (*b* = -.0178) based on 5,000 bootstrap samples that was below zero, 95% CI [-.0339, -.0058], meaning that participants in WPRO-Asia were more likely to have fast responses and thus have lower rates of diagnostic accuracy compared to clinicians in North America.

Table 11 and Figure 4 show that in comparison to medical professionals, other clinical professions had slower respond times (*b* = .0536) and psychologists were the slowest responders (*b* = .0613). As before, slower responders provide accurate diagnoses more often (*b* = .1648). Percentile bootstrap confidence intervals for the indirect effect of clinical profession (*b* = .001) based on 5,000 bootstrap samples was above zero, 95% CI [.0001, .0029]. More specifically, the indirect effect of psychologist (*b* = .0101, 95% CI [.0032, .0187]), and other clinical professionals (*b* = .0088, 95% CI [.0009, .0029]) based on 5,000 bootstrap samples were above zero. Psychologists' and other clinical professionals' slower response time contributed to more accurate diagnostic decision making.

Table 11. Model coefficients for all mediation models predicting diagnostic accuracy

	Consequent					
	M (Response Time)			*Y* (Diagnostic Accuracy)*		
Antecedent	Coeff.	SE	*p*	Coeff.	SE	*p*
Age (*X*)	.0015	.0007	**.0276**	-.0021	.0013	.1197
Years of Experience (*X*)	.0017	.0008	**.0216**	-.0018	.0015	.2081
Response Time (*M*)	--	--	--	.1599	.0449	**.0004**
Region (*X*)						
X_1 AMRO-North**	--	--	--	--	--	--
X_2 AMRO-South	.0660	.0322	**.0404**	-.0391	.0626	.5329
X_3 AFRO	-.0551	.0607	.3645	-.1146	.1179	.3313
X_4 EMRO	.0310	.0587	.5981	-.0912	.1139	.4234
X_5 EURO	.0017	.0262	.9479	.0078	.0508	.8779
X_6 SEARO	-.0054	.0421	.8987	.0205	.0818	.8024
X_7 WPRO-Asia	-.1178	.0292	**.0001**	.0029	.0569	.9593
X_8 WPRO-Oceania	-.0825	.0553	.1357	.0285	.1074	.7910
M Response Time	--	--	--	.1508	.0457	**.0010**
Profession (*X*)						
X_1 Medicine**	--	--	**--**	--	--	--
X_2 Psychology	.0613	.0180	**.0007**	-.0621	.0347	.0739
X_3 Other	.0536	.0268	**.0451**	-.0955	.0514	.0637
M Response Time	--	--	--	.1648	.0450	**.0003**

Note. *These coefficients represent the direct effect of the antecedents on diagnostic accuracy.

** X_1 under Region and Profession served as the reference group for the categorical predictor in the mediation analyses.

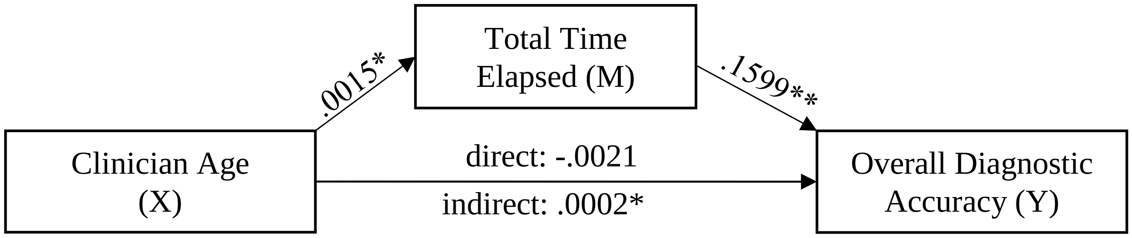

Figure 1. Simple mediation model for the association between clinicians' age and their diagnostic accuracy as mediated by total time elapsed.
*p <.05, **p < .001.

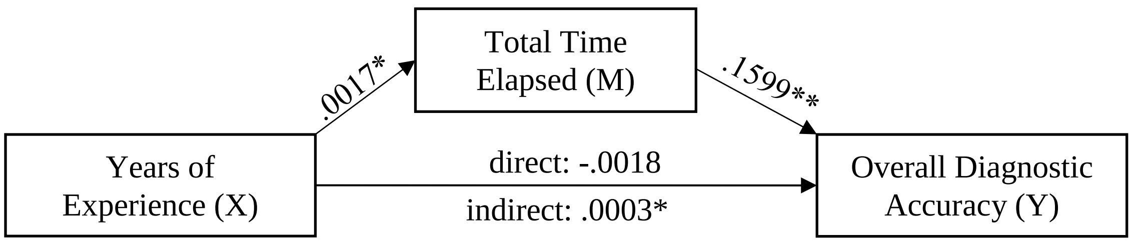

Figure 2. Simple mediation model for the association between years of experience and diagnostic accuracy as mediated by total time elapsed.
*p <.05, **p < .001.

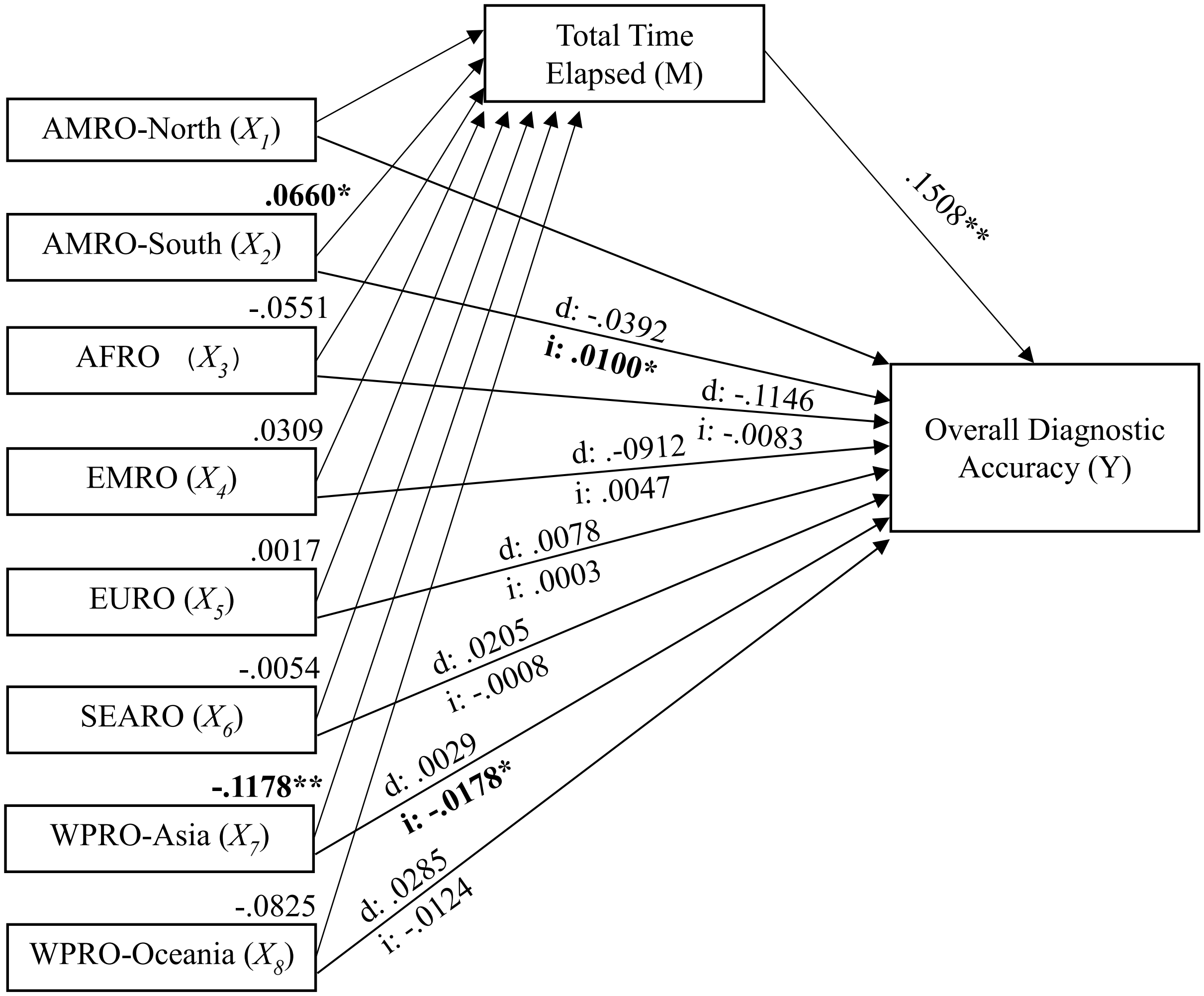

Figure 3. Simple mediation model for the association between region and diagnostic accuracy as mediated by total time elapsed. X_1 served as the reference group for analyses, therefore no statistical results are reported.
*p <.05, **p < .001.

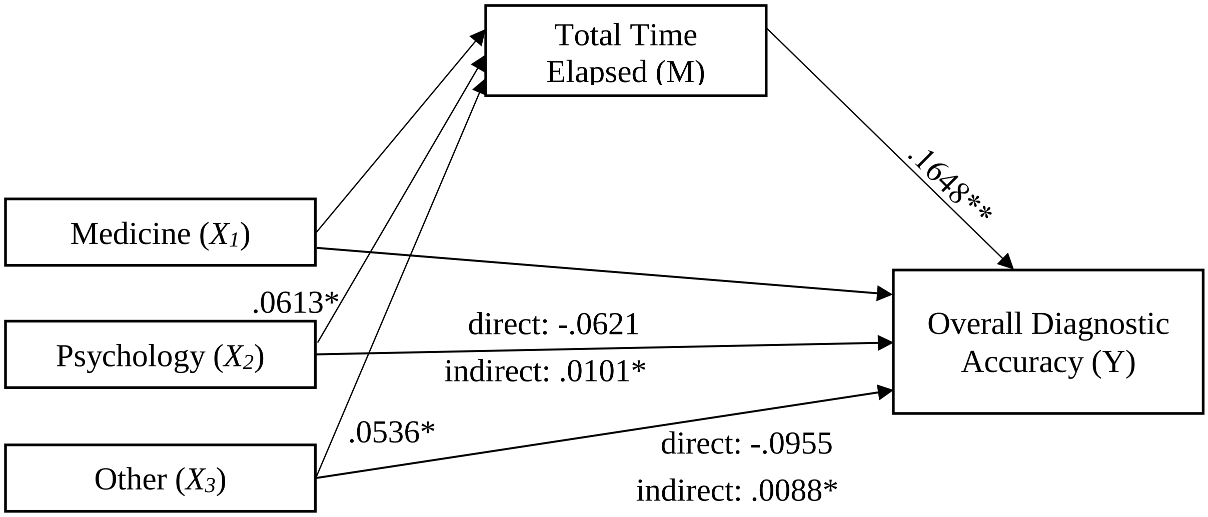

Figure 4. Simple mediation model for the association between clinical profession and diagnostic accuracy as mediated by total time elapsed. X_1 served as the reference group for analyses, therefore no statistical results are reported.
*p <.05, **p < .001.

Discussion

This study aimed to understand the impact of nonclinical factors such as clinician characteristics and response time on diagnostic accuracy. The current study did not find evidence that clinicians' demographic features were directly related to their diagnostic accuracy. Rather, participants' age, years of experience, world region, and their clinical profession indirectly predicted accuracy through their overall response time. Significant research efforts have focused on examining the effect of patient factors on providing diagnoses across clinical settings; however, the influence of clinician demographics have received less attention (Blumenthal-Barby & Krieger, 2015). The findings from the current study highlight the need for researchers and clinicians to consider the role mental health providers' dispositional features have in the diagnostic process. Although the primary aim of the study resulted in non-significant associations, the results from mediation analyses revealed an indirect effect of clinician's demographic features on diagnostic accuracy. Research on decision-making focused on the process and response time without considering accuracy of results while research on accuracy rates across years of experience neglected to consider response time (Ægisdóttir et al., 2006; Blumenthal-Barby & Krieger, 2015; Garber et al., 2005; Muroff et al., 2007; Webb et al., 2016). The findings of this study support that to better understand differences across clinicians in diagnostic accuracy, it is crucial to consider demographics and response styles.

Regarding the aim to investigate the association between clinician demographics and diagnostic accuracy, the main pattern of the findings suggest that demographic features do not directly contribute to whether clinicians provide accurate diagnoses. In one instance, the gender interaction between the clinician and vignette did significantly predict diagnostic accuracy for only the second vignette. Specifically, incongruent clinician-vignette gender led to higher rates of incorrect diagnoses for the second vignette than matching clinician-vignette genders. While this finding supports the study's hypothesis, it is unlikely that a gender interaction would impact accuracy for the second vignette but not the first. This finding adds to the mixed evidence suggesting that clinician-patient gender matching influences the diagnoses provided (Garb, 1997; Muroff et al., 2007; Woodward et al., 2009). Additionally, the gender of the second vignette has a small effect on diagnostic accuracy of the vignette, with gender accounting for less than one percent of the variability in diagnostic accuracy. One alternative explanation for this pattern is that the second vignette typically presented a more complex case, thus being more difficult to diagnose. Similarly, participants could have experienced responder fatigue, with half spending 30 minutes or longer completing the study, such that their diagnosis on the second vignette would become less accurate. Additionally, there were relatively few vignettes about women, and they were typically the second vignette presented. As such, the interaction could have only emerged for the second vignette. These other factors may have acted as confounds which obscure our understanding of the impact of the clinician-patient gender interaction on the diagnostic process. Muroff and colleagues (2007) found that when diagnosing female patients but not male patients, clinicians' patient load, time, and cognitive resources impacted their ability to differentiate bipolar disorder from other mood disorders. Overall, they found that clinicians with heavy patient loads, fewer cognitive resources, and those who spent less time with female patients resulted in them applying gender stereotypes in diagnostic decision-making, e.g., female patients were depressed rather than bipolar. This study suggests that availability of cognitive resources and time influences how the clinician-patient gender interaction impacts diagnostic decision-making. It is possible that by the second vignette, clinicians had less time and cognitive resources to devote to decision-making and relied more heavily on gender stereotypes, heuristics, or intuition.

To further understand how time influenced clinicians' diagnostic decision-making, participants were categorized as either slow, moderate or fast responders based on their overall response time. Fast responders, with the lowest response times, were found to incorrectly diagnose both vignettes more often than moderate, average speed responders, or slow responders, those with the longest response times. This finding suggests that participants who spent less time on the study were missing crucial details in the diagnostic guidelines or vignettes necessary to provide an accurate diagnosis. Alternatively, quick responders who made incorrect diagnoses might have relied incorrectly on heuristics or their intuitive pathway of the dual process model of decision-making without assessing whether these decision-making tactics were appropriate. While heuristics and the intuitive pathway are advantageous, they cannot be applied reliably to new information, i.e., the ICD-11 condition of the study (Schwartz & Elstein, 2009). In contrast, moderate and slow responders demonstrated that lack of speed favorably impacted accuracy. It is plausible that they relied on their analytic reasoning abilities and engaged in systematic, hypothesis testing which typically requires more time but increases their odds of determining the correct diagnosis (Croskerry & Nimno, 2011; Schwartz & Elstein, 2009). While these theoretical explanations aid in understanding why fast responders were less accurate diagnosing vignettes than moderate and slow responders, more research is needed to understand which heuristics clinicians use and the impact of individual differences in memory and processing speed on diagnostic accuracy.

In addition to individual differences in response time, this study examined whether other participant demographics would predict differences in the amount of time they spent completing the study. Age and years of experience were statistically significant predictors of overall response time. Specifically, older participants and individuals with more years of experience were more likely to spend more time completing the study. While older individuals are typically slower responders, the association between years of experience and response time contradicted findings in the literature on decision-making and psychodiagnostic accuracy (Faust, Balota, Spieler, & Ferraro, 1999). Research on cognitive heuristics and the dual-process theory both challenge the idea that individuals with more expertise would spend more time processing their diagnostic decision. Instead, these theories state that more experience is related to having established more heuristics or higher levels of recognition to engage the intuitive pathway for decision-making (Croskerry & Nimno, 2011; Keren & Teigen, 2004; Schwartz & Elstein, 2009). These cognitive processes then would allow experienced individuals to respond more quickly and accurately. Despite decision-making research supporting the notion that more experience is related to having more cognitive resources for quick and accurate decision-making, past research in diagnostic accuracy did not find significant differences between novice and experienced clinicians (Ægisdóttir et al.,2006; Spengler et al., 2009; Strasser & Gruber, 2004; Witteman &

Tollensaar, 2012; Witteman & Van den Bercken, 2007). Prior to this study, it was unclear why more expert clinicians were not out performing early career clinicians. A mediation was conducted to elucidate how diagnostic accuracy was impacted when more experienced clinicians responded slower than less experienced clinicians. The mediation analysis revealed that more experienced clinicians indirectly increased their likelihood of making accurate diagnoses when they spent more time on the study. Specifically, experienced participants engaged in slower, more deliberate, systematic hypothesis testing rather than relying on intuition and cognitive aids which allowed them to be more accurate. This indirect effect does not support decision-making theories that expect experienced individuals to utilize heuristics or intuition and respond quickly. Moreover, this study does not replicate findings that novice and experience clinicians do not differ in diagnostic accuracy. The indirect effect of years of experience through response speed counted for less than one percent of the variability of diagnostic accuracy. While there was a significant indirect effect, it only captures a subtle influence on diagnostic accuracy. Therefore, researchers should re-examine the notion that clinicians with more clinical experience are only marginally better than early career clinicians by considering mediating factors such as response style and cognitive decision-making processes.

Global region was another demographic feature that was associated with the amount of time participants required to complete the study. Specifically, participants in Asia were significantly quicker than those in Europe followed by North America, and South America who took the longest to finish. Additional analyses reduced concerns that these differences might be due the level of infrastructure in the country and internet connectivity, by using countries' income level as a proxy which was found not to be associated with response time. Unfortunately, there is a dearth of cross-cultural research examining differences in response style. Some research in decision-making suggests that under time pressure, participants will rely on cultural attributions or expectations to make a quick decision (Chiu et al., 2000). A mediation analysis was used to assess whether differences in response time across regions impacted diagnostic accuracy compared to clinicians in North America. There was a significant indirect effect of region; South American participants were more likely to be more accurate when they responded slower. Additionally, Asian participants were more likely to respond quickly resulting in less accurate diagnoses. It is possible that cultural and professional differences are linked to differences in response time and diagnostic accuracy across regions. One possible explanation is that clinicians in South America needed to spend more time than clinicians in other regions to make accurate diagnoses because approximately half of Latin American psychologists expressed a need for a national DCS (Evans et al., 2013; Reed et al., 2011). South American clinicians might have been slower to be more careful while using a DCS they do not feel is culturally applicable. As a result, these clinicians were able to provide the correct diagnosis. In comparison, clinicians in Asia were significantly less likely to view applying the ICD-10 cross-culturally as problematic (Evans et al., 2013; Reed et al., 2011). Asian clinicians possibly felt more comfortable using the ICD-10, therefore felt safe responding more quickly than clinicians in other regions. However, this result suggests that Asian clinicians did not allot the appropriate time necessary to provide an accurate diagnosis. Unfortunately, the meaning of this indirect effect is not clear due to gaps in the literature. Additionally, these indirect effect of region through response time accounted for less than one percent of the variability in diagnostic accuracy. This small effect size suggests region has a subtle influence on accuracy rates. More research on the psychodiagnostic process using diverse samples that include measures assessing culture is needed to better understand the implications for this finding.

Last, clinical profession was associated with differences in response style. Medical professionals were significantly faster at completing the study than psychologists. Other mental health providers did not significantly differ from either group of professionals in their diagnostic accuracy. This finding is congruent with the abundance of evidence that mental health providers in medical settings are required to make quick diagnostic decisions under time pressure due to larger caseloads resulting in less time to devote to each patient (Muroff et al., 2007; Stepanikova, 2012). A mediation analysis was conducted to determine whether these differences in response time across clinicians' clinical settings impacted diagnostic accuracy. In comparison to medical professionals, psychologists and other mental health providers were more likely to accurately diagnose vignettes when they spent more time completing the study. The indirect effect of type of profession through response time accounted for less than one percent of the variability in diagnostic accuracy. This small effect size suggests that clinical profession has a subtle influence on accuracy rates. This indirect effect supports past research that time constraints increased the probability of medical professionals providing inaccurate diagnoses (Muroff et al., 2007; Witteman & Tollenaar, 2012). Additionally, these findings suggest that medical professionals might be employing incorrect decision-making processes (e.g., heuristics, pattern recognition, and the intuitive judgment pathway) to compensate for limited cognitive resources, time pressure, and heavy caseloads. If medical professionals allotted more time for the diagnostic decision-making process, they could evaluate the efficacy of the decision process and accuracy of the resulting diagnosis. The results of this study suggest that medical professionals who are fast responders incorrectly categorized vignettes and relied on a quicker intuitive mode of judgment (e.g., an incorrect heuristic) rather than a slower, effortful analytic method of decision-making (Croskerry & Nimno, 2011; Schwartz & Elstein, 2009).

Implications

Differences in diagnostic accuracy across mental health providers have implications for how medical professionals are trained to think about their patients and the decision-making process for mental health diagnoses. The above findings suggest that medical professionals are relying on cognitive aids to respond faster but less accurately. Fast responding might also result in clinicians engaging implicit biases. Overall, the findings of this study suggests that diagnostic accuracy rates would improve if mental health provider allotted more time for the diagnostic decision-making process. To avoid engaging negative implicit biases when selecting a diagnosis, medical professionals should receive training on how to appropriately handle large caseloads, time pressure, and limited cognitive resources without engaging in inaccurate or biased judgment processes (Garber et al., 2012; Stone & Moskowitz, 2011). Additionally, it is important to combat the impact of implicit biases and cognitive errors in diagnostic decision-making because they might be perpetuating health care disparities (Chapman, Kaatz, & Carnes, 2013). Stone and Moskowitz (2011) suggest that cultural competence training that focuses on inhibiting or suppressing stereotypes and attitudes are ineffective training modalities for reducing unconscious biases. Instead, they suggested that medical providers receive multicultural competence training that also incorporates the current findings about the impact of implicit biases and participation in activities that help clinicians in medical settings identify their own implicit biases (e.g., IAT). After these activities, medical professionals would also learn four strategies that have the potential to reduce their implicit biases: (a) pursuing egalitarian goals through engaging in metacognition about their decision-making process; (b) identifying common identities; (c) counter-stereotyping; (d) and perspective taking (Croskerry, 2003; Stone & Moskwotiz, 2011).

Last, increased understanding about clinicians' diagnostic accuracy has implications for the type of treatment and clinical interventions clinicians implement and therapeutic outcomes of the patients. Receiving a mental health diagnosis can either be stigmatizing, discordant with how a patient views their self, or confirming of the patient's experience depending on how their culture views mental health and the quality of interaction with mental health professionals (Kirmayer, 2005; Shanttell, McAllister, Hogan, & Thomas, 2006). On the other hand, a patient learning they were incorrectly diagnosed with a mental health disorder could increase stigma towards mental health providers and negatively impact their perceptions of themselves (Corrigan, 2007). The findings of this study and past research suggest that taking more time to assess one's decision-making process could diminish rates of diagnostic errors and the negative impact of providing a patient a biased or inaccurate diagnosis (Croskerry, 2003; Garber et al., 2012).

Strengths and Limitations

Although the study does contain limitations, several strengths must be acknowledged. First is the overall diversity of clinicians in the current sample. While a majority of the sample was from North America or Europe, the remaining sample was from other global regions creating a more representative and heterogeneous sample rather than solely studying clinicians that more closely represent people from Western, Educated, Industrialized, Rich and Democratic (WEIRD) societies. In doing so, the findings of this study will be more applicable to mental health providers of different cultures and clinical professions. Additionally, prior research either failed to report or include pertinent demographics about clinicians creating a paucity of information about the influence these individual characteristics have on diagnostic accuracy. This study was one of the first to report specific demographic information about a heterogeneous sample of clinicians and use these data in statistical analyses. This study provided much needed information about the global clinician population and the indirect effects of their characteristics on diagnostic accuracy.

In addition to sample strengths, this is the first study to explore mediating factors of diagnostic accuracy beyond decision-making processes and patient characteristics. Specifically, this study was novel in empirically measuring clinicians' response time during the diagnostic decision-making process. Additionally, the use of vignettes provided a precise means of determining whether or not participants provided an accurate diagnosis. These measurement strengths allowed this study to gather data about the prevalence rate of diagnostic accuracy in a large, heterogeneous sample of clinicians. Prior research in the fields of clinical cognition and diagnostic accuracy has heavily focused on errors, which might have lead researchers to overestimate the amount of errors clinicians make (Schwartz & Elstein, 2009). Instead, this study provided beneficial insight and feedback about what clinicians are doing correctly during the diagnostic process (e.g., slower response times and analytic modes of decision-making).

Although the current study has many strengths, several limitations must be addressed. First, it utilized secondary data analysis. These data provided a unique opportunity to analyze the decision-making process of a diverse group of clinicians; however, the study was limited by the measures and methodology used in the original study. Therefore, proxy variables were used to measure response latency, culture, and impact of country infrastructure and internet connectively on response time. While these proxy variables do not diminish the significant findings of the study, they do limit the generalizability to research on the implicit biases clinicians hold towards their patients. Additionally, the methodology of the study did not allow for exploration or control of possible covariates. The effect sizes of all the analysis in the study were quite small suggesting that we are only discussing subtle influences upon diagnostic accuracy. Future research is needed to explore what other factors may confound or contribute to predicting clinicians' diagnostic accuracy. The project was also limited in its ability to determine causality due to lack of manipulation and the use of a convenience sample. For true experimental design, new methodologies should be explored that allow researchers to manipulate the type of decision-making processes clinicians use to determine a diagnosis. Last, the use of patient vignettes to assess clinicians' rate of diagnostic accuracy does not provide ideal ecological validity. The following section provides suggestions for how future research can improve upon these limitations.

Future Directions

The current study explored the relationship between clinician demographics, response time and diagnostic accuracy in a sample of international clinicians. Overall, four indirect pathways linked clinician age, years of experience, region, and clinical profession to participants' rates of diagnostic accuracy. These four indirect pathways suggest that diagnostic accuracy significantly improved when participants slowed down and took longer to review diagnostic guidelines and provide a diagnosis. In addition to response time, it is possible that other indirect pathways might be relevant to explore within the association between diagnostic accuracy and clinician demographics. Based on the findings of this study, future research should account for other factors that might indirectly impact diagnostic accuracy. First, prior work about clinicians' level of experience found only marginal differences across novice, intermediate, and expert clinicians. These findings should be revisited to explore the indirect pathway between experience and accuracy. Future studies should account for response style for replicability but also other factors of the decision-making process or therapeutic process that might impact diagnostic accuracy. The current study did not directly explore whether implicit biases, memory, individual's decision-making process, or their therapeutic approach impacted the diagnosis they provided. These factors have been shown to influence diagnostic accuracy and treatment outcomes; however, it is unclear whether they differ across experience level, cultures, and clinical profession (Webb et al., 2016; Chapman et al., 2013; Croskerry, 2003; Woodward et al., 2009). While this study was one of the first studies to use a heterogeneous sample of global clinicians to explore clinicians' diagnostic accuracy, the study was limited in how it measured culture, race, and ethnicity. Therefore, future research should directly measure clinicians' culture and racial identities and expand on the indirect effect of region on diagnostic accuracy found in this study. It is also important to consider future work that could be done with these results to reduce rates of diagnostic errors. Effective interventions and training should focus on the underlying mechanisms that typically cause cognitive and diagnostics errors: implicit biases and stigma towards mental health diagnoses (Chapman et al., 2013; Corrigan, 2007; Croskerry, 2003; Garber et al., 2002; Stone & Moskowitz, 2011).

Conclusion

The current study found evidence supporting indirect effects of clinician demographics (e.g., age, years of clinical experience, global region, and profession) positively impacting diagnostic accuracy when clinicians spent more time responding to the vignettes and reviewing diagnostic guidelines. Most importantly, taking into account differences in response style resolved conflicting perspectives about whether clinical experience is a precise predictor of diagnostic accuracy. Rather than expert clinicians narrowly doing better than novice clinicians, as clinicians gained years of experience and allotted more time for the study, their diagnostic accuracy improved. Additionally, differences in diagnostic accuracy across years of experience and age contradicted theories about how clinicians engage different methods of decision-making. The dual-process theory posits that individuals with less experience will rely on the slower, more deliberate analytic mode to arrive at a solution, while someone with more experience or pattern recognition would engage the intuitive mode more often. However, an indirect effect showed that more experienced clinicians took longer but provided the correct diagnosis more often. This result suggests that experienced clinicians utilize the analytic mode more often contrary to the dual-process theory. Moreover, differences in response time across regions and the impact on diagnostic accuracy suggests that there are cultural differences in response style and approach to decision-making which resulted in differences in accuracy rates. These findings on the indirect impact of region, age, and years of experience have implications for continued education and training for clinicians, and future research should assess the efficacy of clinical training to provide adequate training about modes of diagnostic decision-making across cultures (Spengler et al., 2009).

References

Ægisdóttir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S., & Nichols, C. N…., Rush, J. D. (2006). The meta-analysis of clinical judgment project: Fifty-sex years of accumulated research on clinical versus statistical prediction. *The Counseling Psychologist, 34,* 341-382.

Blumenthal-Barby, J. S., & Krieger, H. (2015). Cognitive biases and heuristics in medical decision making: A critical review using a systematic search strategy. *Medical Decision Making, 35,* 539-57.

Brannon, L. A., & Carson, K. L. (2003). The representativeness heuristic: Influence on nurses' decision making. *Applied Nursing Research, 16,* 201-204.

Callaghan, K. S. N. (2012). Factors influencing diagnostic decision-making. *Journal of Primary Health Care, 4,* 223-230.

Chapman, E. N., Kaatz, A., & Carnes, M. (2013). Physicians and implicit bias: How doctors may unwittingly perpetuate health care disparities. *Journal of Internal Medicine, 28,* 1504-1510.

Chiaramonte, G. R., & Friend, R. (2006). Medical students' and residents' gender bias in the diagnosis, treatment and interpretation of coronary heart disease symptoms. *Health Psychology, 25,* 255-266.

Choi, I., Choi, J. A., & Norenzayan, A., (2004). Culture and decisions. In D. J. Koehler & N. Harvey (Eds.), *Blackwell handbook of judgement and decision making* (pp. 504-524). Malden, MA.: Blackwell Publishing Ltd.

Chiu, C., Morris, M. W., Hong, Y., & Menon, T. (2000). Motivated cultural cognition: The impact of implicit cultural theories on dispositional attribution varies as a function of need for closure. *Journal of Personality and Social Psychology, 78,* 247-259.

Cioffi, J. (2001). A study of the use of past experiences in clinical decision making in emergency situations. *International Journal of Nursing Studies, 38,* 591-599.

Corrigan, P. W. (2007). How clinical diagnosis might exacerbate the stigma of mental illness. *Social Work*, *52*(1), 31-39.

Croskerry, P. (2003). The importance of cognitive errors in diagnosis and strategies to minimize them. *Academic Medicine*, *78,* 775-780.

Croskerry, P., & Nimno, G. R. (2011). Better clinical decision making and reducing diagnostic error. *Journal of the Royal College of Physicians of Edinburgh*, *41*, 155-162.

Cuevas, A. G., O'Brien, K., & Saha, S. (2016). African American experiences in healthcare: "I always feel like I'm getting skipped over". *Health Psychology, 35,* 987-995.

Eisenberg, J. M. (1979). Sociologic influences on decision-making by clinicians. *Annals of Internal Medicine*, *90*, 957-965.

Elstein, A. S., & Schwarz, A. (2002). Evidence base of clinical diagnosis. Clinician problem solving and diagnostic decision making: Selective review of the cognitive literature. *The BMJ, 324*, 729-732.

Epstein, S. (1994). Integration of the cognitive and psychodynamic unconscious. *American Psychologist*, *49*, 709-724.

Evans, S. C., Reed, G. M., Roberts, M. C., Esparza, P., Watts, A. D., Correia, J. M., & Ritchie, P. …, Saxena, S. (2013). Psychologists' perspective on the diagnostic classification of mental disorders: Results from the WHO-IUPsys global survey. *International Journal of Psychology, 48,* 177-193.

Faust, M. E., Balota, D. A., Spieler, D. H., & Ferraro, F. R. (1999). Individual differences in information-processing rate and amount: Implication for group differences in response latency. *Psychological Bulletin, 125,* 777-799.

Fazio, R. H. (1990). A practical guide to the use of response latency in social psychology research. In C. Hendrick & M. S. Clark (Eds.), *Research methods in personality and social psychology* (pp. 74-97). Newbury Park, CA.: Sage Publication.

Garb, H. N. (1996). The representativeness and past-behavior heuristics in clinical judgment. *Professional Psychology: Research and Practice, 27,* 272-277.

Garb, H. N. (1997). Race bias, social class bias, and gender bias in clinical judgment. *Clinical Psychology: Science and Practice, 4,* 99-120.

Garb, H. N. (2005). Clinical judgment and decision making. *Annual Review of Clinical Psychology, 1,* 67-89.

Garb, H. N. (2010). The social psychology of clinical judgment. In J. E. Maddux, & J. P. Tangney (Eds.), *Social psychological foundations of clinical psychology* (pp. 297–311). New York, NY.: Guilford Press.

Graber, M. L., Frnaklin, N., & Gordon, R., (2005). Diagnostic error in internal medicine. *Archive of Internal Medicine, 165,* 1493-11499.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual different in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74,* 1464-1480.

Hayes, A. F. (2013). *An introduction to mediation, moderation, and conditional process analysis*. New York: The Guilford Press.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33* (2/3), 1-75. doi:10.1017/S0140525X0999152X

Isen, A. M., Rosenzweig, A. S., and Young, M. J. (1991). The influence of positive affect on clinical problem solving. *Medical Decision Making, 11,* 221-227.

Kahneman, D. (2003). Maps of bounded rationality: A perspective on intuitive judgment and choice. In T. Frangsmyr (Ed.), *Les prix Nobel. The Nobel prizes,* Stockholm: Almqvist & Wiksell International.

Keeley, J. W., Reed, G. M., Roberts, M. C., Evans, S. C., Medina-Mora, E. M., Rebello, T., & Sharan, P. …, Saxena, S. (2016). Developing a science of clinical utility in diagnostic classification systems field study strategies for ICD-11 mental and behavioral disorders. *American Psychologist, 71*(1), 3-16.

Keller, M. B., Lavori, P. W, Klerman, G. L., Andreasen, N. C., Endicott, J., Coryell, W., & Fawcett, J., …, Hirschfel, R. M. A. (1986). Low levels and lack of predictors of somatotherapy and psychotherapy received by depressed patients. *Archive of General Psychiatry, 43,* 458-466.

Kelley, L. P., & Blashfield, R. K. (2009). An example of psychological science’s failure to self-correct. *Review of General Psychology, 13,* 122-129.

Keren, G. & Teigen, K. H. (2004). Yet another look at the heuristics and biases approach. In D. J. Koehler & N. Harvey (Eds.), *Blackwell handbook of judgement and decision making* (pp. 89-109). Malden, MA.: Blackwell Publishing Ltd.

Kirmayer, L. J. (2005). Culture, context and experience in psychiatric diagnosis. *Psychopathology, 38*, 192-196.

Lane, K. A., Banaji, M. R., Nosek, B. A, & Greenwald, A. G. (2007). Understanding and using the implicit association test: IV: What we know (so far) about the method. In B. Wittenbrink & N. Schwarz (Eds.), *Implicit measures of attitudes* (pp. 59-102). New York, NY.: The Guilford Press.

McKinlay, J. B., Potter, D. A., & Feldman, H. A. (1996). Non-medical influences on medical decision-making. *Social Science Medicine*, *42*, 769-776.

Muroff, J. R., Jackson, J. S., Mowbray, C. T., &, Himle, J. A. (2007). The influence of gender, patient volume, and time on clinical diagnostic decision making in psychiatric emergency services. *General Hospital Psychiatry*, *29*, 481-488.

Neighbors, H. W., Trierweiler, S. J., Ford, B. C., & Muroff, J. R. (2003). Racial differences in DSM diagnosis using a semi-structured instrument: The importance of clinical judgement in the diagnosis of African Americans. *Journal of Health and Social Behavior, 43,* 237-256.

Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychology Review*, *108*, 291-310. doi:10.1037//0033-295X.108.2.291

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2006). The implicit association test at age 7: A methodological and conceptual review. In J. A. Bargh (Ed.), *Social psychology and the unconscious: The automaticity of higher mental processes* (pp. 265-292). Philadelphia, PA: Psychology Press.

Peris, T. S., Teachman, B. A., & Nosek, B. A. (2008). Implicit and explicit stigma of mental illness: Links to clinical care. *The Journal of Nervous and Mental Disease, 196,* 752-760.

Reed, G. M., Correia, J. M., Esparza, P., Saxena, S., & Maj, M. (2011). The WPA-WHO global survey of psychiatrists' attitudes towards mental disorders classification. *World Psychiatry, 10,* 118-131.

Sabin, J. A., Nosek, B. A., Greenwald, A. G., & Rivara, F. P. (2009). Physicians' implicit and explicit attitudes about race by MD race, ethnicity, and gender. *Journal of Health Care for the Poor and Underserved, 20,* 896-913.

Schwartz, A. & Elstein, A. S. (2009). Clinical problem solving and diagnostic decision making: A selective review of the cognitive research literature. In J. A. Knottnerus & F. Buntinx (Eds.), *The evidence base of clinical diagnosis: Theory and methods of diagnostic research* (pp. 237-255). Malden, MA.: Blackwell Publishing Ltd.

Segal, S. P., Bola, J. R., & Watson, M. A. (1996). Race, quality of care, and antipsychotic prescribing practices in psychiatric emergency services. *Psychiatric Services, 47,* 282-286.

Shanttell, M. M., McAllister, S., Hogan, B., & Thomas, S. P. (2006). "She took the time to make sure she understood": Mental health patients' experiences of being understood. *Archives of Psychiatric Nursing, 20,* 234-241.

Smith, D. & Dumont, F. (1997). Eliminating overconfidence in psychodiagnosis: Strategies for training and practice. *Clinical Psychology: Science and Practice, 4,* 335-345.

Solomon, A. (1992). Clinical diagnosis among diverse populations: A multicultural perspective. *Families in Society, 73*, 371-377.

Spengler, P. M., White, M. J., Ægisdóttir, S., Maugherman, A. S., Anderson, L. A., Cook, R. S., Nichols, C. N…&, Rush, J. D. (2009). The meta-analysis of clinical judgement project: Effects of experience on judgment accuracy. *The Counseling Psychologist, 37*, 350-399.

Stepanikova, I. (2012). Racial-ethnic biases, time pressure, and medical decisions. *Journal of Health and Social Behavior, 53i,* 329-343.

Stone, J. & Moskowitz, G. B. (2011). Non-conscious bias in medical decision making: What can be done to reduce it? *Medical Education, 45,* 768-776.

Strasser, J. &, Gruber, H. (2004). The role of experience in professional training and development of psychological counselors. In H. P. A. Boshuizen, R. Bromme, & H. Gruber (Eds.), *Professional learning: Gaps and transitions on the way from novice to expert* (pp. 11-27). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Tversky, A. & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science, 185,* 1124-1131.

Webb, C. A., Keeley, J. W., & Eakin, D. K. (2016). Are clinicians better than lay judges at recall case details? An evaluation of expert memory. *Journal of Clinical Psychology, 72,* 384-400. doi:10.1002/jclp.22256

Widiger, T. A. & Settle, S. A. (1987). Broverman et al. revisited: An artifactual sex bias. *Journal of Personality and Social Psychology, 53,* 463-469.

Witteman, C. L. M. &, Tollenaar, M. S. (2012). Remembering and diagnosing clients: Does experience matter? *Memory, 20,* 266-276.

Witteman, C. L. M. &. Van den Bercken, J. H. L. (2007). Intermediate effects in psychodiagnostic classification. *European Journal of Psychological Assessment*, *23*, 56-61.

Woodward, H. E., Taft, C. T., Gorgon, R. A., & Meis, L. A. (2009). Clinician bias in the diagnosis of posttraumatic stress disorder and borderline personality disorder. *Psychological Trauma: Theory, Research, Practice, and Policy, 1*, 282-290.

Vita

Julia Keller Brechbiel was born on December 30, 1991, in Washington D.C. and is an American citizen. She graduated from St. Stephen's & St. Agnes School, Virginia in 2010. She obtained her Bachelor of Science degree in Psychology from the College of William & Mary in Williamsburg, Virginia in 2014. She subsequently worked as a hotline crisis counselor providing services and support for women who experienced domestic or sexual violence for a year. She began her graduate work in Counseling Psychology doctoral program at Virginia Commonwealth University in August 2015 under the mentorship of Jared Keeley, Ph.D.