



CENTRO UNIVERSITÁRIO AUGUSTO MOTTA – UNISUAM
PROGRAMA DE PÓS-GRADUAÇÃO STRICTU SENSU EM CIÊNCIAS DA
REABILITAÇÃO

**CONCORDÂNCIA INTEREXAMINADORES E ACURÁCIA
DIAGNÓSTICA PARA DIFERENCIAÇÃO DE PADRÕES E
PRESCRIÇÃO DE ACUPONTOS**

INGRID JARDIM DE AZEREDO SOUZA OLIVEIRA

RIO DE JANEIRO

2015

**CONCORDÂNCIA INTEREXAMINADORES E ACURÁCIA
DIAGNÓSTICA PARA DIFERENCIAÇÃO DE PADRÕES E
PRESCRIÇÃO DE ACUPONTOS**

INGRID JARDIM DE AZEREDO SOUZA OLIVEIRA

Dissertação de mestrado apresentado ao programa de pós-graduação em ciências da reabilitação, do Centro Universitário Augusto Motta, como parte dos requisitos para obtenção do título de Mestre em Ciências da Reabilitação.

Orientador: Prof. Dr. ARTHUR DE SÁ FERREIRA

RIO DE JANEIRO

2015

FICHA CATALOGRÁFICA
Elaborada pelo Sistema de bibliotecas e
Informação – SBI – UNISUAM

615.892 O48e	Oliveira, Ingrid Jardim de Azeredo Souza Concordância interexaminadores e acurácia diagnóstica para diferenciação de padrões e prescrição de acupontos / Ingrid Jardim de Azeredo Souza Oliveira. Rio de Janeiro, 2015. 119 p. Dissertação (Mestrado em Ciências da Reabilitação). Centro Universitário Augusto Motta, 2015. 1. Diagnóstico. 2. Acupontos. 3. Medicina Tradicional Chinesa. 4. Reabilitação I. Título
-----------------	--

CONCORDÂNCIA INTEREXAMINADORES E ACURÁCIA DIAGNÓSTICA PARA DIFERENCIAÇÃO DE PADRÕES E PRESCRIÇÃO DE ACUPONTOS

INGRID JARDIM DE AZEREDO SOUZA OLIVEIRA

Dissertação de mestrado apresentado ao programa de pós-graduação em ciências da reabilitação, do Centro Universitário Augusto Motta, como parte dos requisitos para obtenção do título de Mestre em Ciências da Reabilitação.

Aprovado em: _____.

BANCA EXAMINADORA:

Prof. Dr. Arthur de Sá Ferreira, DSc
Centro Universitário Augusto Motta – UNISUAM

Prof. Dr. Fernando Silva Guimarães, DSc
Centro Universitário Augusto Motta – UNISUAM

Prof. Dr. Antonio Guilherme Fonseca Pacheco, DSc
Centro Universitário Augusto Motta – UNISUAM

Prof. Dr. Márcio Antonio Babinski, DSc
Universidade Federal Fluminense – UFF

Rio de Janeiro
2015

DEDICATÓRIA

À Deus pela graça da criação ter-me permitido concluir mais esta etapa em minha vida.

Aos meus pais Isaias de Azeredo Souza (*In memoriam*) e Nancy Jardim de Azeredo Souza, por todo esforço e dedicação, que muitas vezes abriram mão das próprias vidas para investir na minha formação e educação, vocês são responsáveis pelo o que sou hoje!

Ao meu esposo José Tadeu Madeira de Oliveira, meu amor, amigo, conselheiro, companheiro de todas as horas e pai dos meus amados filhos!

E aos meus filhos Gabriel Jardim Madeira de Oliveira (“Magrila”) e Gustavo Jardim Madeira de Oliveira (“Cara-de-Bico”), meus maiores adversários no cumprimento de minhas obrigações junto ao mestrado! Vocês são a razão do meu viver e o motivo pelo qual me faz querer ser melhor a cada dia!

AGRADECIMENTOS

Sirvo-me deste momento para prefaciар esta dissertação com os louros agradecimentos àqueles que tiveram importância fundamental no decorrer dos meus estudos.

Ao meu orientador, Professor Dr. Arthur de Sá Ferreira, todo o meu respeito, admiração e a certeza de que cumpristes a missão valiosa de um mestre na concepção verdadeira da palavra. Criastes um discípulo que mesmo não te imitando quanto a perfeição, sente-se capaz de dizer que começo a ser uma pesquisadora de verdade. O meu muito obrigado, e um até breve...

Aos abnegados mestres que mesclaram-se no aperfeiçoamento de minha formação inicial de pesquisadora, começo uma nova etapa de minha vida e comigo ficou a seiva vivificante dos saberes ensinados por vós. O meu muito obrigado aos professores: Agnaldo Lopes, Alex Souto, Anke Bergmann, Antonio Guilherme Fonseca, Érika de Carvalho, Fernando Guimarães, Julio Silva, Laura Oliveira, Lilian Felicio, Patrícia Vigário e Sara Menezes.

Agradeço aos profissionais que colaboraram com a pesquisa por mim realizada. Obrigada pela coragem, atenção e paciência que culminaram de forma ímpar para a conclusão desta dissertação.

Ao Centro Universitário UNIABEU e a Universidade Iguaçu UNIG, onde exerço a função de docente, agradeço ao apoio à minha capacitação, e a certeza de que todo trabalho é vão, quando não se tem amor, e graças a Deus tenho amor pelo que faço nestas Instituições.

A todos que, de alguma forma contribuíram para a elaboração de minha dissertação, principalmente, aos colegas do Curso de Mestrado em Ciências da Reabilitação do Centro Universitário Augusto Motta - UNISUAM.

A Érica Mendes e Dayene Gama Terra, secretária do curso do Programa de Pós-Graduação em Ciências da Reabilitação do UNISUAM, pela simplicidade e prestimosidade.

Resumo

Introdução: A medicina tradicional chinesa (MTC) vem se desenvolvendo há pelo menos 3.000 anos por uma coletividade geográfica, social e culturalmente distante da ocidental. A fisiologia da MTC é baseada nas substâncias vitais, nos órgãos internos e nos canais e colaterais de acupontos. As manifestações clínicas de um indivíduo são obtidas durante a sua avaliação por meio da inspeção (Ip), ausculta-olfação (AO), interrogação (Iq) e palpação (P) – coletivamente conhecidos como os quatro exames, onde os sinais e sintomas são interpretados como relacionados às substâncias vitais, órgãos internos e canais e colaterais, constituindo assim os padrões patológicos que indicam o estado de saúde daquele indivíduo em um dado momento. **Objetivo:** Investigar a concordância interexaminadores e desempenho diagnóstico para diferenciação de padrões e prescrição de acupontos sob condições de diagnóstico correto e incorreto. **Métodos:** Este estudo simples-cego, observacional de corte-transversal, comparou a concordância entre examinadores, especialistas em Acupuntura para a diferenciação de padrão e prescrição acupontos entre dois grupos de pacientes humanos simulados com os resultados de diagnóstico diferentes. Um conjunto de dados com 509 manifestações de padrões *zangfu* foi organizado. Os pacientes humanos foram simulados e separados em grupos ($n = 30$ cada um) de acordo com o resultado de diagnóstico determinado por um modelo computacional: correto e incorreto. Um questionário em papel com 90 questões que representam cada paciente simulado foi entregue face-a-face para 6 especialistas em MTC (4 do sexo masculino, de 4 anos a mínima de experiência clínica) para indicar um único padrão (entre 73) e até 8 pontos de acupuntura (entre 378). **Resultados:** A concordância foi maior para o padrão de diferenciação do que para a prescrição de acupontos. A concordância no padrão de diferenciação foi leve para ambos os grupos com diagnóstico correto (κ de Light = 0,167 [0,108; 0,254]) e desfecho diagnóstico incorreto (κ de Light = 0,190 [0,120; 0,286]). A concordância sobre a prescrição de acupontos também foi pequena para os dois grupos de diagnóstico correto ($\iota = 0,029$ [0,015; 0,057]) e diagnóstico incorreto ($\iota = 0,040$ [0,023; 0,058], $p = 0,075$). Análise do desempenho diagnóstico dos avaliadores rendeu acurácia = 60,9%, sensibilidade = 21,7% e especificidade de 100%. Uma melhoria global da concordância entre e acurácia diagnóstica foi observada quando os dados foram analisados utilizando códigos para os sistemas internos em vez de padrões rotulados. **Conclusão:** A concordância foi leve para diferenciar os padrões e a prescrição de acupontos em ambos os casos de diagnóstico correto e incorreto. A concordância foi melhor para a combinação de diferenciação, mas ainda leve para a prescrição de pontos de acupuntura com base nos órgãos internos selecionados. A acurácia diagnóstica dos examinadores em MTC para diferenciação dos padrões *zangfu*, não foi a melhor possibilidade para a maioria, embora fosse melhor para diferenciar os órgãos internos afetados, em especial, quando um método automatizado não conseguiu apresentar o diagnóstico correto.

Descritores: Diagnóstico, acupontos, medicina tradicional chinesa, reabilitação.

Abstract

Introduction: Traditional Chinese medicine (TCM) has been developed for at least 3,000 years for a geographic community, socially and culturally distant from the West. The physiology of TCM is based on the vital substances in the internal organs and channels and collateral acupoints. Clinical manifestations of an individual are obtained during their evaluation through inspection (Ip), auscultation, olfaction (AO), inquiry (Iq) and palpation (P) - collectively known as the four tests, where the signs and symptoms are interpreted as related to vital substances, internal organs and channels and collaterals, thus constituting the pathological patterns that indicate the state of health of that individual at a given time. **Aim:** To investigate the interrater agreement and diagnostic performance for pattern differentiation and acupoint prescription under correct diagnosis and misdiagnosis. **Methods:** This single-blinded, observational cross sectional study compared the interrater agreement of Acupuncture experts for pattern differentiation and acupoints prescription between two groups of human patients simulated with different diagnostic outcomes. A dataset with 509 manifestations of zangfu patterns was organized. Human patients were simulated and separated into groups (n=30 each) according to the diagnostic outcome determined by a computational model: correct and incorrect. A hardcopy questionnaire with 90 questions representing each simulated patient was delivered face-to-face to 6 TCM experts (4 males, 4-year minimal clinic experience) for indicating a single pattern (among 73) and up to 8 acupoints (among 378). **Results:** Interrater agreement was higher for pattern differentiation than for acupuncture prescription. Interrater agreement on pattern differentiation was slight for both groups with correct diagnosis (Light's kappa=0.167 [0.108; 0.254]) and incorrect diagnosis outcome (Light's kappa=0.190 [0.120; 0.286]). Interrater agreement on acupuncture prescription was also slight for both groups of correct (iota=0.029 [0.015; 0.057]) and incorrect diagnosis (iota=0.040 [0.023; 0.058], p=0.075). Analysis of diagnostic performance of raters yielded accuracy=60.9%, sensitivity=21.7%, and specificity=100%. An overall improvement in the interrater agreement and diagnostic accuracy was observed when the data were analyzed using codes for the internal systems instead of the pattern's labels. **Conclusion:** Interrater agreement was slight for differentiating zangfu patterns and prescribing acupoints, regardless of whether the diagnosis is accurate or not. Interrater agreement is better for differentiating combinations but yet slight for acupuncture prescription based on the selected internal organs. Diagnostic accuracy of TCM raters for differentiating zangfu patterns was not better than chance for most raters, although it was better for differentiating the affected internal organs in particular when an automated method failed to provide the correct diagnosis.

Keywords: Diagnosis, acupuncture points, traditional chinese medicine, rehabilitation.

Sumário

	Pág.
Capítulo 1 Introdução	
1.1 Medicina tradicional chinesa.....	9
1.1.1 Contextualização.....	9
1.1.2 Fundamentos filosófico-sistemáticos de fisiologia da MTC.....	10
1.1.3 Fisiopatologia e diferenciação de padrões da MTC.....	10
1.1.4 Prescrição de intervenções terapêuticas baseada na diferenciação de padrões.....	13
1.2 Medidas de desempenho estatístico de métodos diagnósticos.....	15
1.3 Revisão da Literatura.....	19
1.3.1 Acurácia e erros na diferenciação de padrões.....	19
1.3.2 Concordância interexaminadores na diferenciação de padrões e na prescrição de acupontos.....	21
1.3.3 Utilização de questionários e casos simulados em estudos diagnósticos.....	25
1.4 Justificativas.....	25
1.5 Problemas.....	26
1.6 Hipóteses.....	27
1.7 Objetivos.....	27
1.7.1 Geral.....	27
1.7.2 Específicos.....	27
Capítulo 2 Métodos	
2.1 Delineamento do estudo.....	28
2.2 Amostra.....	28
2.2.1 Local de realização do estudo.....	28
2.2.2 Cálculo e justificativa do tamanho amostral.....	28
2.3 Critérios de inclusão e exclusão.....	28
2.4 Procedimentos.....	29
2.5 Análise dos dados.....	31
2.5.1 Variáveis de controle.....	32
2.5.2 Variáveis de desfecho.....	32
2.5.3 Análise estatística.....	32
2.6 Aspectos éticos.....	33
Capítulo 3 Resultados	
3.1 Manuscrito – <i>Journal of Integrative Medicine</i>	37
3.2 Manuscrito – <i>Evidence-based Complementary and Alternative Medicine</i>	58
Capítulo 4 Considerações Finais	90
Referências	92
Anexo 1 – Carta de aprovação do comitê de ética em pesquisa	99
Apêndice 1 – Questionário para caracterização da amostra	102
Apêndice 2 – Caderno de casos clínicos	103
Apêndice 3 – Formulário de respostas	113
Apêndice 4 – Tabela de padrões	116
Apêndice 5 – Tabela de acupontos	117
Apêndice 6 – Termo de consentimento livre e esclarecido	119
Apêndice 7 – Comprovantes de submissão dos manuscritos	121

Capítulo 1 Introdução

1.1 Medicina tradicional chinesa

1.1.1 Contextualização

A Organização Mundial da Saúde (OMS) define medicina tradicional como “a soma total do conhecimento, habilidades e práticas baseadas nas teorias, crenças e experiências de diferentes culturas, explicáveis ou não, e usadas na manutenção da saúde, bem como na prevenção, diagnóstico, tratamento ou melhoria de doenças físicas e mentais” (WHO, 2000). Práticas da medicina tradicional variam de região para região, uma vez que são influenciadas por fatores como a cultura, a história, as atitudes pessoais e a filosofia, e em muitos casos, a sua teoria e aplicação são muito diferentes daquelas relativas à medicina ocidental (também referida como convencional) (WHO, 2000).

Dentre as medicinas tradicionais orientais, a medicina tradicional chinesa (MTC), vem se desenvolvendo há pelo menos 3.000 anos por uma coletividade geográfica, e social culturalmente distante da ocidental. Achados arqueológicos datados de cerca de 1.000 a.C. incluem agulhas de acupuntura e ossos entalhados com registros de discussões sobre problemas de saúde. No período de 206 a.C. a 220 d.C., as bases teóricas e práticas da MTC estavam mais desenvolvidas em comparação à épocas anteriores, com uma abordagem prática relativamente sofisticada. Entre os séculos III e IX, avanços na MTC foram observadas, marcadamente a publicação em forma de livro das prescrições para o tratamento de diversas condições mórbidas. Os livros de MTC que estabeleceram os seus fundamentos foram datados do século IV, sendo o mais importante desses preservado até a presente data, o *Huangdi Neijing* – Clássico Interno do Imperador Amarelo (T’AO, 1955; LEE et al., 1962).

As medicinas tradicionais orientais adentraram o Brasil após 1960 no contexto do movimento internacional de contracultura, o qual sugeria um novo estilo de vida que incorporasse tendências naturalistas e afins com as civilizações do oriente (QUEIROZ, 2006). Desde seu reconhecimento pela OMS na declaração Alma-Ata (WHO, 1978), as medicinas tradicionais têm sido cada vez mais objeto de pesquisa científica teórica, experimental e principalmente clínica. No Brasil, o acesso gratuito da população às técnicas orientais de medicina pelo Sistema Único de Saúde também tem crescido de forma surpreendente nos

últimos anos (FRAGOSO, 2010), principalmente devido à implantação da Política Nacional de Práticas Integrativas e Complementares (BRASIL, 2006). Tal política recomenda a implantação e implementação de ações e serviços derivados das medicinas tradicionais no SUS, com o objetivo de garantir a prevenção de agravos, a promoção e a recuperação da saúde, além de propor o cuidado continuado, humanizado e integral em saúde, contribuindo com o aumento da resolubilidade do sistema, com qualidade, eficácia, eficiência, segurança, sustentabilidade, controle e participação social (BRASIL, 2006).

1.1.2 Fundamentos filosófico-sistemáticos de fisiologia da MTC

A MTC possui raízes filosóficas e sistemáticas baseadas nas relações entre a natureza e a humanidade (O'CONNOR & BENSKY, 1987). A fisiologia da MTC é baseada nas substâncias vitais, nos órgãos internos e nos canais e colaterais (meridianos) de pontos de acupuntura (acupontos). A teoria dos sistemas internos se refere aos mesmos representados pelos aspectos funcional e orgânico mas não necessariamente anatômico e descreve as relações envolvidas nas funções fisiológicas do organismo (YAMAMURA, 2001). Embora tal teoria seja baseada no conhecimento limitado de anatomia da época antiga, o seu desenvolvimento ocorreu por meio da observação dos relatos dos enfermos daquela época, com base no princípio de que “o exterior do corpo reflete o seu interior” (O'CONNOR & BENSKY, 1987).

Os antigos praticantes de MTC mapearam regiões do corpo que, quando estimuladas, aparentemente apresentaram repercussões terapêuticas no organismo – acupuntura, do latim *acus* e *pungere* para ‘agulha’ e ‘puncionar’, respectivamente. A união desses acupontos em sequências ao longo do corpo configura os canais longitudinais e colaterais horizontais, sendo que sua principal função é conectar todo o corpo, de maneira que se mantenham saudáveis todas as estruturas. Embora a quantidade de acupontos tenha aumentado ao longo do desenvolvimento da MTC, atualmente 361 acupontos nos canais e colaterais são mapeados e utilizados (WHO, 1993), além de outros muitos acupontos extra-canais (O'CONNOR & BENSKY, 1987).

1.1.3 Fisiopatologia e diferenciação de padrões da MTC

As manifestações clínicas de um indivíduo são obtidas durante a sua avaliação por meio da inspeção (Ip), ausculta-olfação (AO), interrogação (Iq) e palpação (P) – coletivamente conhecidos como os quatro exames (O'CONNOR & BENSKY, 1987). Os sinais e sintomas são interpretados como desequilíbrios relacionados às substâncias vitais, órgãos internos e

canais e colaterais, constituindo assim os padrões patológicos que indicam o estado de saúde daquele indivíduo em um dado momento (MACIOCIA, 1996; CINTRA & PEREIRA, 2012). Também importa para o diagnóstico outros fatores, tais como os aspectos emocionais, os hábitos alimentares, sexuais e de atividade física (YAMAMURA, 2006).

Respeitando as diferenças conceituais entre a medicina oriental e a medicina convencional (ocidental), o termo ‘padrão’ (*zheng*) representa uma alteração de saúde de um indivíduo. O padrão é descrito por manifestações clínicas que identificam simultaneamente a origem, a localização e a natureza da condição mórbida (MACIOCIA, 1996). A diferenciação de padrões é o procedimento da MTC análogo ao diagnóstico na medicina convencional, sendo realizada com a intenção de reconhecer e rotular padrões (cartesianos) para que se possa reproduzir a intervenção apropriada, a fim de restaurar a saúde ou minimizar a deficiência permanente (LU & CHEN, 2009). A diferenciação de padrões portanto não é focada no reconhecimento de uma doença no sentido biomédico, mas sim do padrão. Em MTC, para uma identificação acurada do padrão, torna-se fundamental apresentar um entendimento etiológico e patológico do indivíduo. O principal ponto na diferenciação de padrões envolve a identificação do padrão primário; o padrão primário compreende aquele que mais representa a doença principal do indivíduo (SHÛ-DÉ, 2006). Uma mesma doença pode apresentar diferentes padrões (O’CONNOR & BENSKY, 1987; YANG et al., 1998; MACIOCIA, 1996), e portanto, diferentes padrões são possíveis para uma mesma doença; esse argumento suporta a teoria de que diferentes tratamentos em MTC também podem ser dados para a mesma doença.

A diferenciação de padrões pode ser estabelecida com base em diversas teorias pautadas na interpretação da Natureza (meio ambiente) e sua relação com a humanidade (ZASLAWSKI, 2003; CINTRA & PEREIRA, 2012). O rótulo de um padrão depende então da combinação sistemática entre os mecanismos fisiopatológicos, as substâncias vitais e sistemas internos envolvidos (Tabela 1); por exemplo, um padrão pode assumir o rótulo “deficiência de *qi* do *fei*”, mostrando a possibilidade de interação entre os aspectos considerados para diagnóstico. O rótulo do padrão sugere todo o seu curso de alterações patológicas, determinando a morbidade de uma doença num certo estágio (CAO et al., 2012). Reitera-se que não é necessário que o indivíduo apresente todas as manifestações descritas pela literatura para o referido padrão; um subconjunto pode ser suficiente para o estabelecimento acurado do diagnóstico (MACIOCIA, 1996).

A literatura não explicita claramente as teorias para o raciocínio diagnóstico na MTC (YANG & LI, 1993; YANG & CHACE, 2004), mas este está refletido na tomada de decisões terapêuticas (ZASLAWSKI, 2003). Entretanto, pode-se notar que o raciocínio lógico da MTC

depende de um julgamento simbólico e intuitivo, com raciocínio abstrato integrado ao processo. Embora careça de medição quantitativa, presta-se atenção à análise da causa do padrão e sua evolução para obter discernimento sobre a sua natureza (ZHOU et al., 2013).

Tabela 1: Aspectos considerados para rotulação do padrão.

Mecanismo fisiopatológico	Excesso
	Deficiência
	Submersão
	Colapso
	Rebelião
	Estagnação, estase
	Secura
	Umidade
	Calor
	Frio
Substâncias vitais	<i>Qi</i>
	<i>Yin</i>
	<i>Yang</i>
	Essência (<i>jing</i>)
	Mente, espírito (<i>shen</i>)
	Sangue (<i>xue</i>)
Órgãos e vísceras	Fluidos corporais (<i>jinye</i>)
	Coração (<i>xin</i>)
	Fígado (<i>gan</i>)
	Baço (<i>pi</i>)
	Pulmão (<i>fei</i>)
	Rim (<i>shen</i>)
	Pericárdio (<i>xinbao</i>)
	Intestino delgado (<i>xiaochang</i>)
	Vesícula biliar (<i>dan</i>)
	Estômago (<i>wei</i>)
	Intestino grosso (<i>dachang</i>)
	Bexiga (<i>pangguang</i>)
	Tripo aquecedor (<i>sanjiao</i>)

OBS: Os padrões dos órgãos internos podem ser descritos por um ou mais atributos simultaneamente de cada aspecto considerado para a rotulação do padrão.

Uma análise do procedimento de diferenciação de padrões de órgãos (órgãos e vísceras – *zangfu*) – e suas consequências terapêuticas – revela desfechos possíveis (SÁ FERREIRA, 2012). Um primeiro desfecho possível é o diagnóstico correto, que consiste na identificação do padrão verdadeiro do indivíduo. Neste caso, o tratamento imediato pode ser realizado com métodos terapêuticos da MTC apropriados àquele padrão. Um segundo desfecho possível é o diagnóstico incorreto, no qual o padrão identificado não é o verdadeiro padrão do indivíduo,

mas qualquer outro padrão. Neste caso, o tratamento imediato segundo a MTC ainda pode ser realizado, mas talvez não seja o mais adequado ao padrão verdadeiro. Outro desfecho possível compreende o diagnóstico ausente, no qual não é possível identificar o padrão subjacente com base nas manifestações apresentadas pelo indivíduo. O diagnóstico ausente pode resultar em atraso no diagnóstico e no tratamento, pois contraria o princípio da MTC de “tratar aqueles que ainda não estão doentes” (YANG & LI, 1993; YANG & CHACE, 2004; FLAWS, 2004). Os dois últimos desfechos – diagnóstico incorreto e ausente – compreendem erros de diferenciação de padrões e podem ter repercussões terapêuticas desfavoráveis.

1.1.4 Prescrição de intervenções terapêuticas baseada na diferenciação de padrões

A MTC possui diversos métodos e técnicas com aplicações aos níveis primários, secundários e terciários de prevenção (FROOM & BENBASSAT, 2000); em parte, isso explica porque a MTC é uma prática multidisciplinar para a promoção da saúde, diagnóstico precoce e intervenção, bem como reabilitação (BRASIL, 2006; FERREIRA & LOPES, 2011). Embora a acupuntura e a moxabustão sejam as modalidades terapêuticas mais reconhecidas da MTC, também estão disponíveis para os profissionais de saúde a terapia herbal e alimentar, massagem terapêutica, exercícios físicos e exercícios de respiração e meditação (DE SÁ FERREIRA, 2013). Frequentemente, essas modalidades são utilizadas em combinação, de acordo com a natureza holística da MTC (GUANG, 2001). Dentre esses, a acupuntura e moxabustão merecem destaque, pois têm se mostrado eficaz e segura para uma vasta gama de enfermidades, inclusive doenças do sistema nervoso, musculoesquelético e cardiovascular (OMS, 2003; DE SÁ FERREIRA, 2013).

O termo chinês *zhen jiu* – literalmente ‘espetar’ e ‘queimar’ – compreende de forma mais completa o método de tratamento por acupuntura e/ou moxabustão. A acupuntura é uma das técnicas mais antigas e comumente utilizada da MTC, sendo uma terapia baseada na teoria dos canais e colaterais. As agulhas são inseridas em pontos do corpo específicos denominados acupontos, cada um dos quais com distintas ações terapêuticas. Por sua vez, a moxabustão é uma técnica que consiste em aquecer os pontos de acupuntura pela queima de ervas medicinais (YAMAMURA, 2001). Após a identificação dos padrões e a elaboração de um plano de tratamento, o acupunturista seleciona um ou mais acupontos baseando-se em critérios tais como sua ação regional (próxima e à distância) e específica, bem como a ação combinada dos

acupontos de acordo com a dinâmica dos canais e colaterais (O'CONNOR & BENSKY, 1987; MACIOCIA, 1996).

Embora existam vários estudos sobre a eficácia de um acuponto como intervenção (FIREMAN et al., 2001; SCHAECHTER & NEUSTEIN, 2008), raramente o praticante usa somente um acuponto como tratamento; igualmente rara é a utilização dos mesmos acupontos para todos os pacientes com o mesmo diagnóstico (CHENG, 1987; MACIOCIA, 1996). A seleção de acupontos para os padrões da MTC é prescrita com base em textos antigos e livros didáticos contemporâneos de MTC – os quais apresenta ampla variedade de acupontos sugeridos – bem como na experiência clínica dos praticantes. Porém, ressalta-se que a maioria dos conceitos e terapias da MTC ainda não foi comprovada pelo método científico (CAO et al., 2012), embora se acredite que o tratamento baseado em padrões irá proporcionar uma melhor eficácia. De fato, estudos sobre o benefício adicional da seleção de acupontos utilizando a diferenciação de padrão em MTC são escassos (YEUNG et al., 2012).

O desenvolvimento da acupuntura se deu na China, com seus muitos dialetos, bem como nos países vizinhos, onde línguas como o japonês, coreano e vietnamita são faladas, dando origem a um grande número de diferenças de nomenclatura. Alguns acupontos são referidos por nomes diferentes e uma variedade de traduções e transliterações dificultam sua comparação entre estudos. Entretanto, esforços para desenvolver uma nomenclatura uniforme já vem acontecendo há algum tempo. Em 1965, o *Japan Meridian* e o Comitê de Pontos foram estabelecidos, recomendando um nome provisório padrão japonês para cada ponto de acupuntura, e um sistema de numeração internacional. Na China, a Sociedade de Acupuntura e Moxabustão criou uma comissão que desenvolveu uma nomenclatura padrão. Desde então, vários outros países formaram comitês nacionais de nomenclatura. Após acordo em nível regional, um grupo científico adotou uma Nomenclatura Padrão Internacional em Outubro-Novembro de 1989. O grupo de trabalho concordou que há um total de 361 pontos de acupuntura clássica e que a ordem dos canais e colaterais seria baseada no padrão de circulação dos canais como hoje percebidos na China, Japão, República da Coreia e do Vietnã. O grupo propôs que a nomenclatura padrão deveria ser composta por três elementos essenciais, a saber: (1) código alfanumérico; (2) o alfabeto fonético (*pinyin*) do nome chinês; e (3) os caracteres *han* (chinês) do canal e do acuponto, cujo código alfanumérico facilita a troca internacional embora careça de sentido a partir de um ponto de vista terapêutico (WHO, 1993). Assim, o rótulo de um acuponto pode assumir o formato LU-1 *zhongfu*, indicando ser o 1º acuponto do canal do pulmão (*fei*) que assume o nome *zhongfu*.

1.2 Medidas de desempenho estatístico de métodos diagnósticos

Erros de diagnóstico não são incomuns na prática médica; o diagnóstico de uma doença pode apresentar vários vieses dependentes de fatores relativos ao examinador, tais como a sua formação, experiência, práticas e história pessoal, além do erro humano inevitável. Além desses fatores, existem muitos casos em que a dificuldade no diagnóstico reside na diferenciação entre dois diagnósticos muito intimamente relacionados ou similares, mas que podem ter prognósticos muito diferentes (RASHID & JAKOBIEC, 2014). Finalmente, há o erro atrelado a qualquer procedimento diagnóstico, mesmo sem a interferência do erro humano: qualquer teste é passível de erro, seja clínico ou laboratorial. Portanto, o estudo do desempenho estatístico de um modelo diagnóstico representa um passo importante na avaliação de sua utilidade clínica (GUYATT et al., 1986).

Os parâmetros de desempenho estatístico de um teste diagnóstico são geralmente obtidos a partir de dados organizados em uma tabela de contingência 2×2 (Tabela 2). Nessa apresentação, os dados são distribuídos de acordo com o resultado do teste que determina a presença ou não de uma condição – denominado padrão-ouro ou referência – e com o resultado do teste em avaliação.

Tabela 2: Tabela de contingência 2×2 para avaliação do desempenho de um teste diagnóstico.

Resultado do teste	Presença da condição real (ou teste de referência)		
		<i>Presente (+)</i>	<i>Ausente (-)</i>
	<i>Positivo (+)</i>	Verdadeiro-positivo (VP)	Falso-positivo (FP)
	<i>Negativo (-)</i>	Falso-negativo (FN)	Verdadeiro-Negativo (VN)

Conforme exibido na Tabela 2, o cruzamento entre a condição real e o resultado do teste resultam nos seguintes dados:

- Verdadeiro-positivo (VP):** condição real *presente* e corretamente identificada pelo teste como *presente*;
- Falso-negativo (FN):** condição real *presente* mas erroneamente identificada pelo teste como *ausente*;
- Falso-positivo (FP):** condição real *ausente* mas erroneamente identificada pelo teste como *presente*;

- d) **Verdadeiro-negativo (VN):** condição real *ausente* e corretamente identificada pelo teste como *ausente*.

Os dados anteriormente descritos – *VP*, *FN*, *FP* e *VN* – são utilizados para calcular os parâmetros de desempenho diagnóstico de um teste. Dentre os parâmetros possíveis, merecem destaque a acurácia (*ACC*), a sensibilidade (*SEN*), a especificidade (*ESP*), o valor preditivo positivo (*VP+*) e o valor preditivo negativo (*VP-*).

A *ACC* (equação 1) pode ser definida como a proporção de casos de acerto do teste (*VP* e *VN*) em relação a todos os casos analisados (*VP*, *VN*, *FP* e *FN*):

$$(1) \quad ACC = \frac{VP+VN}{VP+VN+FP+FN} \times 100\%$$

A *SEN* (equação 2) é calculada pela proporção de casos corretamente identificados pelo teste como presentes (*VP*) dada a presença da condição (*VP* e *FN*). A *SEN* representa a probabilidade condicional de um teste em detectar a presença de uma condição dado que essa condição de fato está presente (ALTMAN & BLAND, 1994a):

$$(2) \quad SEN = \frac{VP}{VP+FN} \times 100\%$$

A *ESP* (equação 3) é calculada pela proporção de casos corretamente identificados pelo teste como ausentes (*VN*) dada a ausência da condição (*VN* e *FP*). A *ESP* reflete a probabilidade condicional capacidade de um teste em detectar a ausência de uma condição dado que essa condição de fato não está presente (ALTMAN & BLAND, 1994a):

$$(3) \quad ESP = \frac{VN}{VN+FP} \times 100\%$$

De modo complementar, no caso de testes diagnósticos, é necessário conhecer a probabilidade de o teste fornecer o diagnóstico correto em casos futuros. O *VP+* (equação 4) é calculado pela proporção de casos corretamente identificados pelo teste como presentes (*VP*) dada a presença da condição identificada pelo teste (*VP* e *FP*). Por outro lado, o *VP-* (equação 5) é calculado pela proporção de casos corretamente identificados pelo teste como ausentes (*VN*) dada a ausência da condição identificada pelo teste (*VN* e *FN*) (ALTMAN & BLAND, 1994b).

$$(4) \quad VP+ = \frac{VP}{VP+FP} \times 100\%$$

$$(5) \quad VP- = \frac{VN}{VN+FN} \times 100\%$$

Ressalta-se que tais parâmetros – *ACC*, *SEN*, *ESP*, *VP+* e *VP-* – são dependentes da prevalência da condição-alvo e, portanto, há fórmulas mais detalhadas que capturam essa dependência.

Além dos parâmetros obtidos da tabela de contingência, outras medidas servem de suporte para a recomendação de um dado teste com fins diagnósticos, dentre elas a concordância. A concordância pode ser definida como “o grau em que as pontuações ou classificações são idênticas” (KOTNNER et al., 2011). A comparação pode ser feita baseada no mesmo examinador classificando os mesmos itens em momentos repetidos (intraexaminador) ou entre dois ou mais examinadores classificando os mesmos itens (interexaminador), em ambos os casos utilizando-se da mesma escala, classificação, instrumento ou procedimento (KOTNNER et al., 2011).

Métodos estatísticos para medir a concordância entre os resultados categóricos estão bem estabelecidos. Cohen (1960) desenvolveu a estatística κ (kappa) como um índice de concordância univariado para variáveis binárias. Mais tarde, Cohen (1968) generalizou o κ original para o coeficiente κ ponderado para resultados discretos ordinais. Desde o seu desenvolvimento, κ com suas extensões (COHEN, 1960; COHEN, 1968; FLEISS, 1971; LIGHT, 1971; KRAEMER, 1980) têm sido bem estudados na literatura, sendo o mais utilizado para avaliar a concordância entre dois ou mais examinadores em variáveis nominais (THOMPSON & WALTER, 1988). Genericamente, o coeficiente κ pode ser calculado como (equação 6):

$$(6) \quad \kappa = \frac{\text{Proporção observada} - \text{Proporção esperada em concordância}}{1 - \text{Proporção esperada em concordância}}$$

O cálculo do κ de Light (1971) pode ser obtido de tabelas de classificação cujo tamanho variam conforme a quantidade de examinadores e de categorias da variável nominal (Tabela 3).

Tabela 3: Tabela de classificação para comparação entre n examinadores avaliando N sujeitos.

Sujeitos avaliados	Examinadores				
	1	2	3	...	n
1	$d_{1,1}$	$d_{1,2}$	$d_{1,3}$	...	$d_{1,n}$
2	$d_{2,1}$	$d_{2,2}$	$d_{2,3}$	...	$d_{2,n}$
3	$d_{3,1}$	$d_{3,2}$	$d_{3,3}$	...	$d_{3,n}$
...	...	...	...	...	...
N	$d_{N,1}$	$d_{N,2}$	$d_{N,3}$	...	$d_{N,n}$

$d_{N,k}$: dado categórico de cada sujeito informado por cada examinador.

No caso da análise de concordância interexaminadores para duas ou mais variáveis (por exemplo: uma lista de acupontos para tratamento), o κ já não é mais adequado. Quando dois ou mais examinadores são utilizados para classificar os mesmos itens em duas ou mais variáveis, a estatística ι (iota) se torna uma opção válida por ser um estimador de concordância interexaminador corrigido para o acaso entre classificações multivariadas de vários examinadores. Portanto, enquanto o κ só poderia ser utilizado para estudar a concordância interexaminadores para um único acuponto por vez (univariado), o ι pode ser utilizado para estudar a concordância interexaminadores para dois ou mais acupontos por vez (multivariado). A estatística ι pode ainda ser interpretada como uma extensão do κ para vários examinadores no caso univariado (JANSEN & OLSSON, 2001).

κ pode ser interpretado qualitativamente como concordância pobre ($<0,00$), leve ($0,00$ a $0,20$), razoável ($0,21$ a $0,40$), moderada ($0,41$ a $0,60$), substancial ($0,61$ a $0,80$) ou quase perfeita ($0,81$ a $1,00$) (LANDIS & KOCH, 1977). ι também tem sido classificado pela mesma escala qualitativa utilizada para o κ (HSIEH et al., 2013). A maioria dos estatísticos consideram valores $>0,6$ para reivindicar um bom nível de concordância, uma vez que a hipótese nula $\kappa = 0$ ou $\iota = 0$ não se aplica, pois além de a própria literatura já conter resultados prévios para esta comparação, a comparação com valor nulo somente mostra diferença para concordância ao acaso. A utilização do κ utilização juntamente com outras medidas de confiabilidade e de acurácia diagnóstica pode mostrar a especificidade da população estudada.

1.3 Revisão da literatura

1.3.1 Acurácia e erros na diferenciação de padrões

A literatura tem questionado a utilização do diagnóstico dado por praticante de MTC como método padrão-ouro, uma vez que é extremamente dependente da interpretação das manifestações subjetivas pelo examinador. De modo a diminuir a subjetividade dos examinadores praticantes de MTC, diversos métodos computacionais têm sido desenvolvidos e testados em casos simulados ou reais e em diferentes modelos de doenças (ZHENG & WU, 2004; WANG et al., 2004; HUANG & CHEN, 2007; LIU et al., 2010). Nesse sentido, nosso grupo de pesquisa iniciou em 2008 uma série de estudos para desenvolvimento do *pattern differentiation algorithm* (PDA); os principais resultados seguem abaixo.

Ferreira (2008) desenvolveu métodos computacionais para simulação de pacientes e diferenciação de padrões. Sessenta e nove padrões foram catalogados pela teoria dos órgãos internos a partir da literatura (MACIOCIA, 1996). Cem perfis de manifestações verdadeiros e falsos foram simulados para cada padrão considerando as seguintes combinações de exame: Ip; AO; Iq; Pa; Ip+AO; Ip+AO+Iq; Ip+AO+Iq+Pa. Os perfis foram simulados utilizando o Algoritmo de Simulação de Perfis de Manifestação (*Manifestation Profiles Simulation Algorithm*, MPSA), também proposto nesse estudo. O PDA realiza uma busca pelas manifestações (tanto como termos livres e expressões entre aspas) para classificação dos padrões candidatos; um padrão era considerado como um candidato caso apresentasse pelo menos uma manifestação coletada no exame ($F\% > 0\%$), ou seja, o $F\%$ representa a proporção entre a quantidade de manifestações clínicas de um dado padrão e o total de manifestações apresentadas pelo paciente no exame. Os padrões candidatos eram classificados por ordem decrescente de $F\%$ para compor a lista de hipóteses diagnósticas, sendo identificado como o diagnóstico o padrão com maior e único $F\%$. Caso não houvesse um único padrão candidato com valores mais altos e não repetidos de $F\%$, o padrão não era identificado (diagnóstico ausente). A diferenciação de padrões por meio dos quatro exames mostrou que maiores valores para acurácia foram obtidos pela interrogação (98,7%), seguido da inspeção (70,7%), da ausculta-olfação (59,9%) e da palpação (56,1%), respectivamente. Quando utilizados os quatro exames combinados, a maior acurácia foi observada, igual a 93,2% ($SEN = 86,5\%$, $ESP = 99,9\%$) (FERREIRA, 2008).

Em sequência, Ferreira (2009) avaliou o impacto da quantidade de informação disponível para diferenciação de padrões – representada pela quantidade de manifestações no perfil simulado, $N\%$ – na acurácia diagnóstica do PDA e propôs a utilização de $N\%$ como um critério secundário para diferenciação de padrões, aplicado após a ordenação dos padrões candidatos pelo $F\%$. Esse critério $N\%$ representa a proporção entre a quantidade de manifestações clínicas de um dado padrão reportadas pelo paciente e o total de manifestações que descrevem o mesmo padrão no banco de dados. Os mesmos 69 padrões dos órgãos internos provenientes do estudo anterior foram utilizados. Os perfis de manifestações foram gerados pelo MPSA também aprimorado a partir do estudo anterior para simular perfis de padrões falsos mais realísticos. Com perfis verdadeiros e falsos foram simulados por padrão para cada combinação de métodos de exame: Ip; Ip+AO; Ip+AO+Iq; Ip+AO+Iq+Pa. A combinação dos quatro exames resultou no melhor desempenho estatístico global com os critérios $F\%$ e $N\%$ ($ACC = 80,1\%$; $SEN = 80,2\%$; $ESP = 79,5\%$), embora menor do que utilizando somente o $F\%$. Identificou-se que uma quantidade ótima de manifestações ($N\% = 28,5\%$) está relacionada com a maior acurácia para identificação dos padrões, uma vez que valores maiores e menores de $N\%$ se mostraram menos acurados para diferenciação de padrões; utilizando-se os dois critérios $F\%$ e $N\%-28.5$, para diferenciação de padrões com os quatro exames, observou-se um aumento estatisticamente significativo no desempenho ($P < 0.001$) do PDA, obtendo-se acurácia maior do que a do estudo anterior ($ACC = 94,7\%$; $SEN = 89,8\%$; $ESP = 99,5\%$) (FERREIRA, 2009).

Mais recentemente, Sá Ferreira (2011) testou o efeito da similaridade de padrões e métodos de análise sobre os resultados do PDA utilizando os dois critérios $F\%$ e $N\%-28.5$. O resultado incorreto da diferenciação de padrões foi separado em dois desfechos: diagnóstico incorreto ou ausente. Para essa finalidade, a tabela de contingência 2×2 foi expandida para 3×2 a partir da divisão do desfecho ‘padrão não identificado’ (Tabela 4) (SÁ FERREIRA, 2012). O banco de dados foi expandido para 73 padrões dos órgãos internos, com um total de 539 manifestações. As manifestações que co-ocorreram em dois ou mais padrões foram atribuídos com o mesmo termo ou expressão (para aumentar a precisão da acurácia na busca do algoritmo exato). Com perfis de manifestações verdadeiros e falsos foram gerados para cada um dos padrões nas seguintes combinações dos quatro exames: Ip; Ip+AO; Ip+AO+Iq; Ip+AO+Iq+Pa. Os resultados sugeriram que a utilização dos quatro exames para diferenciação de padrões resultou em menor taxa de erros de diagnóstico e casos não diagnosticados (6,0% e 1,4%, respectivamente), seguindo-se três (6,6% e 1,3%, respectivamente), dois (14,4% e 24,2%, respectivamente) e com um único método (14,5% e 33%, respectivamente) (SÁ FERREIRA, 2011).

Tabela 4: Tabela de contingência 3×2 para avaliação do desempenho de um teste diagnóstico em estudos de simulação de dados adaptada para separação dos tipos de erro diagnóstico.

Resultado do teste para o padrão	Tipo de simulação (Referência)			
	Interpretação			
			Padrão verdadeiro	Padrão falso
Positivo		<i>Simulado = Identificado</i>	<i>VP</i>	<i>FP</i>
Negativo		<i>Simulado ≠ Identificado</i>	<i>FN</i>	<i>VN</i>
	Negativo	<i>Identificado = []</i>	<i>FN</i>	<i>VN</i>

VP: verdadeiro-positivo; VN: verdadeiro-negativo; FP: falso-positivo; FN: falso-negativo.

O PDA foi utilizado em pesquisas com pacientes reais, nos quais um banco de dados com manifestações de padrões dos órgãos internos relacionados à hipertensão arterial sistêmica foi compilado e o PDA foi utilizado para determinar o padrão dos pacientes e sua relação com lesões de órgãos-alvo (LUIZ et al., 2011). O PDA também foi utilizado para diferenciação do padrão de indivíduos saudáveis e pacientes com hemiparesia crônica após acidente vascular cerebral (FRAGOSO & FERREIRA, 2012). Atualmente, o PDA e o MPSA estão inseridos em um projeto de desenvolvimento de um conjunto de métodos computacionais aplicados à MTC utilizando-se dos métodos epidemiológico e estatístico para sua implementação e validação (DE SÁ FERREIRA, 2013).

1.3.2 Concordância interexaminadores na diferenciação de padrões e na prescrição de acupontos

A variabilidade da diferenciação de padrões e de prescrição de acupontos tem sido reportada com origem tanto na identificação quanto na interpretação das manifestações clínicas, resultando em ampla variação e baixa concordância dos diagnósticos feitos por especialistas em MTC. A seguir são apresentados um resumo dos principais estudos nesse tópico. Quando especificados na literatura original, os parâmetros estatísticos de concordância foram explicitamente informados (exemplo: Cohen's κ).

Sung et al. (2004) estudaram o consenso diagnóstico e para prescrição de fitoterapia para pacientes com Síndrome do Intestino Irritável (SII). Quatro praticantes da MTC de diferentes escolas na China com mais de cinco anos de formação entrevistaram os pacientes de forma independente. O estudo foi dividido em três fases: (I) avaliação cega individual, onde os pacientes (n=39) passaram pelos quatro praticantes da MTC individualmente com anamnese e exame do pulso e da língua, na qual cada praticante MTC foi então solicitado a fornecer um

diagnóstico MTC, princípio do tratamento e prescrição de ervas para cada paciente de acordo com os quatro exames; (II) discussão para chegar a um consenso sobre o diagnóstico e o tratamento, onde os quatro praticantes da MTC foram convidados a ver os pacientes ($n=65$) e pacientes com outras causas de diarreia ($n=17$, ‘controles’), seguidos pelo gastroenterologista (não envolvido na discussão); e (III) avaliação cega, individual com base em critérios diagnósticos consensuais, de modo similar à fase I, exceto que os pacientes com outras causas orgânicas de diarreia foram incluídos como controle. Foram calculados a concordâncias percentual e valor de κ entre todos os pares de praticantes da MTC, e apresentadas as médias com intervalo desses dados; os resultados da fase I e III foram comparados com o teste U de Mann-Whitney. A média dos índices de concordância no diagnóstico, princípio de tratamento e regime do tratamento na fase I foram 57%, 58% e 52%, e 80%, 81% e 80% na fase III ($P = 0,002$). A melhora entre as fases I-II foi estatisticamente significativa nos valores médios de κ para diagnóstico ($\kappa = 0,11$ e $\kappa = 0,34$, respectivamente, $P = 0,015$) e princípio de tratamento ($\kappa = 0,16$ e $\kappa = 0,37$, respectivamente, $P = 0,002$), mas não no regime do tratamento. O estudo sugere que as variações no diagnóstico e princípios de tratamento existem entre os praticantes da MTC, e que a concordância no diagnóstico pode ser aumentada por meio da convergência de opiniões entre os praticantes da MTC.

Zhang et al. (2005) investigaram a concordância diagnóstica e de prescrição de fitoterapia da MTC dentre três examinadores – com no mínimo cinco anos de experiência – para 40 pacientes com artrite reumatoide. Para concordância, considerou-se tanto o padrão específico relacionado à artrite reumatoide quanto os atributos mais genéricos dos oito princípios para diagnóstico na MTC. A concordância média para a diferenciação de padrões foi de 31,7%. A análise estatística descreve o cálculo do κ , embora seus valores não tenham sido apresentados; foram discutidos como sendo abaixo da concordância pelo acaso (seis casos dentre os 40 pacientes foram concordantes entre os três examinadores). A concordância entre prescrição de fitoterapia e o diagnóstico foi alta: 91,7% na média dos três examinadores. Mais tarde, os mesmos autores mostraram um aumento da concordância para diagnóstico após treinamento (ZHANG et al., 2008), com valores de κ entre 0,49 e 0,76 (média dos três pares de examinadores: $\kappa = 0,56$).

Coeytaux et al. (2006) compararam os padrões diagnosticados pela MTC por três acupunturistas licenciados e a seleção de acupontos para 37 pessoas com cefaleia. Os acupunturistas identificaram os fatores envolvidos no padrão conforme organizados na Tabela 1. Para cada padrão de cada participante foi solicitada a indicação de 8 acupontos. Todos os

acupunturistas identificaram o mesmo sistema interno em 27% dos participantes, enquanto que dois acupunturistas identificaram o mesmo sistema interno (*gan*) em 53% dos participantes. Trinta e dois participantes (89%) foram identificados por pelo menos um acupunturista com padrão relacionado à mesma substância vital (*qi*). Pelo menos um acupunturista prescreveu o mesmo ponto (F-3 *taichong* ou IG-4 *hegu*) para quase todos os participantes. Outros pontos foram frequentemente prescritos para muitos participantes (ex.: VG-20 *baihui*), mas houve pouca concordância para a maioria dos outros pontos prescritos. Apesar dos resultados mostrarem a variabilidade do diagnóstico e da prescrição de acupontos, não foram utilizadas medidas estatísticas de concordância para tal avaliação, assim como não foi estudada a relação entre a variabilidade do diagnóstico e dos acupontos prescritos.

Mist et al. (2009) investigaram o quanto um processo de treinamento focado em um questionário baseado no diagnóstico da MTC com 10 praticantes da MTC (licenciados há cinco anos ou mais) melhoraria o consenso na diferenciação do padrão. Quinze participantes foram entrevistados por quatro pares de praticantes com distúrbios na articulação temporomandibular. Foram conduzidos treinamentos com o uso de questionário baseado no diagnóstico chinês, seguido de três sessões diagnósticas nas quais os praticantes discutiam os resultados prévios, com foco na redução da variabilidade diagnóstica. Os resultados apontaram que usando o diagnóstico primário em cada uma das três rodadas, a concordância aumentou de $\kappa = 0,112$ para $\kappa = 0,149$ e $\kappa = 0,318$, respectivamente. Esse estudo sugeriu que a concordância entre os avaliadores do diagnóstico de MTC pode ser melhorada através de um processo de formação e um processo de diagnóstico baseado em questionário. Entretanto, não foi investigada a acurácia diagnóstica nem a sua relação com a concordância da prescrição de acupontos.

O'Brien et al. (2009) investigaram a concordância das observações de diagnóstico de acordo com os órgãos internos (*zangfu*) e os oito princípios (*bagan*) utilizados em MTC. Três praticantes da MTC examinaram 45 adultos com hipercolesterolemia leve a moderada. Todos os praticantes já haviam concluído licenciaturas de 5 anos em MTC e eram registrados em fitoterapia chinesa e acupuntura. Em relação ao diagnóstico pelos oito princípios (mais genérico), a concordância entre os três examinadores variou entre leve para a natureza da desordem ($\kappa = 0,15$ [0,00-0,29]) e quase perfeita para a localização da desordem ($\kappa = 0,87$ [0,75-0,98]). A concordância entre os examinadores para os diagnósticos pela teoria dos órgãos internos (mais específica) resultou em concordância máxima de 24%, mas nenhum valor de κ foi apresentado para esta análise. O estudo enfatizou a natureza subjetiva de aspectos

particulares do diagnóstico da MTC e sugeriu que seja dada mais atenção na definição ou redefinição do processo de diagnóstico utilizado pelos praticantes da MTC.

Grant et al. (2013) investigaram a concordância na diferenciação de padrões em 27 pacientes com pré-diabetes avaliados por dois praticantes utilizando um instrumento previamente validado (TEAMSI-TCM). Os dois praticantes concordaram com diagnóstico primário em 70% dos participantes; houve concordância exata em ambos os padrões de diagnóstico em 41% dos participantes. Três padrões de diagnóstico da MTC para as pessoas com pré-diabetes foram identificados, sendo distribuídos de forma equilibrada dentre os participantes. A concordância entre os praticantes foi κ [IC95%] = 0,56 [0,25-0,81] ($P < 0,001$). Os autores sugerem que ao usar o modelo dos oito princípios na diferenciação do padrão da MTC, os praticantes de diagnóstico de pacientes com pré-diabetes apresentaram confiabilidade moderada. Entretanto, não foi investigada a variabilidade da prescrição de acupontos e por conseguinte sua relação com a variabilidade do diagnóstico.

Birkeflet et al. (2014) realizaram um estudo cujo objetivo foi investigar a concordância da diferenciação de padrões a fim de explorar o impacto de fatores demográficos na conclusão diagnóstica. O método contou com oito avaliadores acupunturistas independentes que avaliaram 25 casos (15 mulheres férteis e 10 inférteis) apresentados na forma escrita. Os resultados apontaram que houve baixa confiabilidade entre os avaliadores sobre os padrões de MTC ($\kappa < 0,20$) e foi encontrada uma grande variação no número de padrões. Foi apontado que sexo, tempo de prática e escolaridade tiveram efeito altamente significativo ($P < 0.001$) e as horas de trabalho apresentaram um efeito significativo ($P = 0.029$) no uso dos padrões da MTC. Houve considerável variação dos avaliadores no uso de sintomas para se chegar a um diagnóstico. Os resultados mostraram ampla variação e pouca confiabilidade entre avaliadores em diagnósticos dos padrões, e sugeriu que variáveis demográficas influenciam a frequência de diagnósticos. Entretanto, não foi investigada a variabilidade da prescrição de acupontos e por conseguinte sua relação com a variabilidade do diagnóstico.

Xu et al. (2014) analisaram a concordância diagnóstica de especialistas em MTC. Um total de 2218 casos com doenças cardiovasculares de diferentes hospitais na China tiveram seus perfis de manifestações obtidos utilizando um questionário padronizado baseado nos quatro exames tradicionais. Os especialistas foram previamente treinados na escala aplicada para garantir sua familiarização com os questionários de avaliação e maior consistência nos procedimentos. O padrão de cada caso foi identificado por dois especialistas em MTC, sendo considerado acurado caso ambos os especialistas concordassem com o mesmo padrão. A

concordância foi quantificada utilizando-se o Cohen's κ e variou de 0,005 a 0,801 dependendo do padrão analisado. As síndromes com boa concordância de diagnóstico foram as síndromes de deficiência do *qi* do coração, deficiência do *yang* do coração, a deficiência do *yin* do coração, fleuma, estagnação de sangue e estagnação de *qi* (Kappa de Cohen $\geq 0,75$). As síndromes com baixa concordância foram deficiência de sangue do coração e do fígado (Kappa de Cohen $\leq 0,40$). As outras síndromes mostraram consistência intermediária. Os autores não discutem, entretanto, a acurácia diagnóstica dos especialistas ou o possível efeito do erro de diagnóstico na confiabilidade interexaminadores.

1.3.3 Utilização de questionários e casos simulados em estudos diagnósticos

As entrevistas face-a-face têm sido consideradas ao longo do tempo como a abordagem de escolha para a coleta de dados em estudos epidemiológicos (SIEMIATYCKI & CAMPBELL, 1994). Em contrapartida, existem algumas limitações na abordagem por envio postal dos questionários: a impossibilidade de garantir que as respostas tenham sido dadas pelo indivíduo selecionado; menos respostas do que as obtidas com entrevistas face a face; pessoas com nível socioeconômico mais baixo poderem ser sub representadas nas entrevistas; e os participantes podem responder de modo diferente as mesmas questões, por estarem, ou não, na presença de entrevistadores. Os entrevistados podem se sentir mais à vontade em responder a questões embaraçosas num questionário, ou serem mais motivados a responder determinadas questões pelo estímulo dos entrevistadores.

O uso de simuladores teve sua origem em outros ramos de atividade como a indústria da aviação e nuclear, em que os operadores devem estar atentos e treinados para enfrentar qualquer intercorrência. A reprodutibilidade dessas intercorrências, sem trazer consequências reais, só é possível com o uso da simulação (DOYLE & ARELLANO, 1995). Assim, o questionário torna-se uma ferramenta importante no estudo de acurácia diagnóstica quando associada a uma simulação computacional.

1.4 Justificativas

A aproximação metodológica entre as práticas médicas tradicional (oriental) e convencional (ocidental) se faz necessária para os avanços na área da saúde em todo o mundo (ABE et al., 2011), requerendo-se o aperfeiçoamento da habilidade diagnóstica dos profissionais de saúde para elaborar com mais acurácia os programas de tratamento, engrandecendo o conhecimento científico humano na caracterização dos fenômenos clínicos.

O diagnóstico e o tratamento no contexto da MTC são processos de constante combinação do pensamento empírico e do pensamento teórico, sendo uma evolução do pensamento e de tomada de decisões (WANG et al., 2010). A diferenciação de padrão para o tratamento é o método que sintetiza as características do padrão primário para o tratamento dos pacientes, cujo valor científico depende de um processo de diferenciação acurado. No entanto, a diferenciação do padrão tende a ser personalizada, o que contradiz a necessidade de prescrições mais padronizadas. Portanto, a investigação da relação diagnóstico-prescrição se justifica pelo estabelecimento maior qualidade nos critérios diagnósticos para obter máxima de acurácia e confiabilidade na assistência e melhor avaliação da relação custo/benefício. Em contraste com a medicina ocidental, que possui protocolos de tratamento para diversas doenças, na literatura científica atual não existem protocolos definidos de seleção de acupontos para os padrões dada a possibilidade de seleção de acupontos locais, distais específicos e sintomáticos, dentre outros fatores (O'CONNOR & BENSKY, 1987).

Diante da impossibilidade de realização de utilização de casos reais – nos quais o diagnóstico real da MTC não é conhecido – a simulação de casos clínicos surge como alternativa razoável para fornecer subsídios para uma análise mais criteriosa dos processos diagnósticos, desde que aplicada de uma maneira sistemática. Portanto, o uso de métodos computacionais se justifica pela necessidade de reduzir a imprecisão inerente ao processo de diagnose, uma vez que a semiologia é extremamente vasta e se confunde aos vários padrões da MTC principalmente pela subjetividade no momento da avaliação.

1.5 Problemas

- a) Qual é o desempenho diagnóstico de examinadores para diferenciação de padrões em pacientes com diagnósticos corretos e incorretos?
- b) Há diferença de concordância interexaminadores para diferenciação de padrões em grupos de pacientes com diagnósticos corretos e incorretos?
- c) Há diferença de concordância interexaminadores para prescrição de acupontos em grupos de pacientes com diagnósticos corretos e incorretos?
- d) Como se comportam tais possíveis diferenças em condições específicas e genéricas de diagnóstico e prescrição terapêutica?
- e) Qual o efeito das características pessoais e profissionais na acurácia para diferenciação de padrões?

1.6 Hipóteses

Hipotetiza-se que examinadores treinados em MTC terão pior desempenho estatístico para diferenciação de padrões em relação ao método computacional em estudo (PDA), devido ao contraste dos critérios subjetivos e objetivos utilizados no procedimento, respectivamente. Em contraste, hipotetiza-se que o desempenho diagnóstico será melhor nos casos com erros diagnósticos do PDA. Em relação à concordância para diferenciação de padrões e prescrição de acupontos, hipotetiza-se que esta não seja diferente em função dos desfechos diagnósticos do PDA (diagnóstico correto e incorreto). Adicionalmente, hipotetiza-se que em condições mais genéricas de diagnóstico e prescrição – isto é, somente o órgão interno é especificado – a concordância tanto para diagnóstico quanto para prescrição seja maior. Finalmente, hipotetiza-se que características relativas à atuação profissional possuam influência sobre a acurácia diagnóstica dos examinadores.

1.7 Objetivos

1.7.1 Geral

Investigar a concordância interexaminadores e desempenho diagnóstico para diferenciação de padrões e prescrição de acupontos sob condições de diagnóstico correto e incorreto.

1.7.2 Específicos

- a) Descrever o desempenho diagnóstico de um grupo de examinadores especialistas em MTC para diferenciação de padrões;
- b) Comparar a concordância interexaminadores para diferenciação de padrões em grupos simulados com diagnósticos corretos e incorretos;
- c) Comparar a concordância interexaminadores para prescrição de acupontos para grupos simulados com diagnósticos corretos e incorretos.
- d) Analisar as comparações de concordância em condições mais genéricas e específicas de diagnóstico e prescrição;
- e) Correlacionar características pessoais e profissionais com a acurácia diagnóstica.

Capítulo 2 Métodos

2.1 Delineamento do estudo

Este estudo apresenta delineamento observacional transversal. Este estudo acompanhou o *Guidelines for Reporting Reliability and Agreement Studies* – GRRAS (KOTTNER et al., 2011). Outro documento específico, o *Standards for Reporting Interventions in Clinical Trials of Acupuncture* – STRICTA (MACPHERSON et al., 2001; MACPHERSON et al., 2010), foi utilizado como referência para determinação das características dos profissionais envolvidos na pesquisa.

2.2 Amostra

2.2.1 Local de realização do estudo

Os participantes foram recrutados junto ao corpo docente de cursos de pós-graduação *lato sensu* em Acupuntura na cidade do Rio de Janeiro (RJ, Brasil). Cada docente foi abordado individualmente para ser esclarecido sobre os objetivos da pesquisa, seus métodos e análise crítica dos riscos e benefícios. O estudo (preenchimento do questionário e análise dos dados) foi realizado no Laboratório de Simulação Computacional e Modelagem em Reabilitação (LSCMR) do Centro Universitário Augusto Motta (UNISUAM).

2.2.2 Cálculo e justificativa do tamanho amostral

O principal fator que influencia o poder de análise estatística é o número de casos que serão simulados. De fato, considerando-se a classificação de $\kappa > 0,40$ como confiabilidade moderada ou superior, observa-se que para detectar valores de confiabilidade $\kappa > 0,40$ para variáveis dicotômicas não é vantajoso utilizar mais do que três examinadores por caso, uma vez que os efeitos no poder dos teste de hipótese e na amplitude dos intervalos de confiança são pequenos. Adicionalmente, recomenda-se uma maior quantidade de casos como uma estratégia mais efetiva para maximizar o poder de análise (SIM & WRIGHT, 2005) Mais especificamente, foi demonstrado que conforme o número de examinadores aumenta, o número

necessário de sujeitos diminui, mas a economia no tamanho amostral obtido pelo aumento do número de examinadores rapidamente diminui após cinco examinadores (ALTAYE et al., 2011). Considerando-se o exposto acima, duas categorias (diagnóstico correto e incorreto) e utilizando-se o questionário com 30 casos simulados/categoria, necessita-se de pelo menos 5 participantes.

2.3 Critérios de inclusão e exclusão

Foram incluídos no estudo os participantes que atenderam simultaneamente aos seguintes critérios:

1. Profissionais com graduação na área de saúde em curso reconhecido pelo Ministério da Educação: Biomedicina, Enfermagem, Educação Física, Fisioterapia, Fonoaudiologia, Medicina, Nutrição, Odontologia, Psicologia, Terapia Ocupacional;
2. Profissionais com habilitação em acupuntura em nível *lato sensu* devidamente registrados no respectivo conselho de classe;
3. Profissionais em prática clínica há pelo menos um ano;
4. Assinatura do termo de consentimento livre e esclarecido (Apêndice 1), após leitura do mesmo sobre os objetivos, riscos e potenciais benefícios da participação na pesquisa.

Foram excluídos do estudo os participantes que solicitaram a sua exclusão do estudo para o responsável desta pesquisa ou por motivos que impedissem a coleta de dados (ex.: ausência ou desistência).

2.4 Procedimentos

1ª etapa: Geração e análise do banco de dados. Foi utilizado o banco de dados de padrões de trabalhos anteriores (FERREIRA, 2011), consistindo em 73 padrões de órgãos internos descritos por um total de 509 manifestações. A consistência e a qualidade do banco de dados de padrão foram computacionalmente testadas antes da simulação, de modo a garantir a exaustividade e exclusividade mútua dos padrões no banco de dados, bem como a unicidade de manifestações em cada padrão.

2ª etapa: Simulação de pacientes. Trezentos pares de perfis de manifestações verdadeiros e falsos foram simulados dentre os padrões pelo algoritmo *SimTCM* (FERREIRA & PACHECO, 2014), um pacote do projeto *SuiteTCM* (DE SÁ FERREIRA, 2013). Os

pacientes foram simulados por meio da seleção de uma quantidade variável e aleatória de manifestações considerando-se os quatro exames ($I_p + A_O + I_q + P_a$). Cada manifestação sorteada foi excluída do conjunto de possíveis manifestações para evitar recorrência da mesma manifestação no respectivo caso simulado (método de amostragem aleatória sem reposição). Da mesma forma, cada padrão foi removido do conjunto de possíveis padrões para simulação do seu respectivo perfil de manifestação falso, sendo repostos para possível sorteio na próxima simulação. Foi assumido que a probabilidade de cada manifestação, assim como de cada padrão, na população em geral era dada e segue uma distribuição uniforme. Adicionalmente, a proporção de informação ruidosa (*proportion of noisy information, PNI*) foi ajustada como nula. A simulação dos perfis pelo *SimTCM* exporta diversos dados, dos quais foram utilizados para análise: (1) o rótulo da condição-alvo sob simulação; (2) o rótulo do padrão sorteado para simulação, que pode ser diferente do rótulo anterior no caso de um padrão com perfil falso; (3) o perfil de manifestações separado por vírgulas para caracterização dos casos clínicos; e (4) o código binário representando o paciente como “1” ou “0” caso o perfil simulado seja verdadeiro (padrão-alvo = simulado) ou falso (padrão-alvo $\neq$ simulado).

3ª etapa: Análise dos desfechos diagnósticos. A análise dos desfechos diagnósticos dos pacientes simulados com perfis verdadeiros e falsos simulados foi realizada pelo PDA utilizando seus dois critérios $F\%$ e $N\%_{-28.5}$. Os pacientes simulados foram separados em categorias de acordo com o resultado da diferenciação de padrões obtida pelo PDA em comparação ao padrão simulado pelo *SimTCM*, nominalmente ‘diagnóstico correto’, ‘diagnóstico incorreto’ e ‘diagnóstico ausente’. Os pacientes simulados com perfis verdadeiros e falsos, mas com diagnóstico ausente foram excluídos das etapas subsequentes.

4ª etapa: Elaboração do questionário para caracterização da amostra. Um questionário autoaplicável foi elaborado com sete questões sobre informações sobre o examinador conforme recomendado pelo o STRICTA (MACPHERSON et al., 2001; MACPHERSON et al., 2010): idade, sexo, área de graduação em saúde, atividade profissional, período de conclusão do curso de especialização, tempo de formação em acupuntura e duração do curso de especialização (Apêndice 1).

5ª etapa: Elaboração do caderno de casos clínicos. Foram sorteados aleatoriamente 30 pacientes simulados com perfis verdadeiros de manifestações para cada um dos dois desfechos de diagnóstico: correto e incorreto. Outros 30 pacientes simulados foram sorteados aleatoriamente dentre os perfis falsos de manifestação, totalizando 90 questões para avaliação do diagnóstico chinês e prescrição de acupontos (Apêndice 2). As questões foram elaboradas e apresentadas de forma direta e clara, evitando-se duplicidade na interpretação. As 90 questões

foram aleatoriamente distribuídas em 30 grupos de 3 casos utilizando um gerador online de números aleatórios (www.random.org) para evitar concentração sequencial de desfechos e assim uma melhor distribuição dos casos pelo questionário.

6ª etapa: *Aplicação dos questionários*. Os questionários foram entregues aos examinadores por via impressa, de modo independente entre os examinadores, contendo uma exposição prévia da pesquisa e sobre as vantagens que essa poderá trazer para a sociedade e em particular para o respondente. Adicionalmente, foi disponibilizado um formulário de resposta com espaços delimitados para o preenchimento do código da resposta referente ao padrão e aos acupontos (Apêndice 3). As respostas dos examinadores foram pareadas de acordo com a numeração da questão. Os questionários foram auto-aplicados e pesquisador não interferiu no seu preenchimento, podendo o examinador utilizar o tempo que julgasse necessário. Foram disponibilizadas ao respondente para consulta: uma tabela impressa com 74 opções de resposta, incluindo os 73 padrões do banco de dados e 1 opção “Não é possível identificar o padrão” (Apêndice 4); uma tabela impressa com os 361 acupontos (WHO, 1993), 17 pontos extra comumente utilizados na prática clínica (O’CONNOR & BENSKY, 1987) e a opção “Nenhum acuponto” (Apêndice 5); e a referência utilizada para construção do banco de dados de padrões e suas respectivas manifestações (MACIOCIA, 1996). Aos participantes foi dado um tempo de 10 minutos para familiarização com as tabelas contendo as opções de resposta. De modo a reproduzir as condições clínicas, foi permitida, mas não estimulada, ao respondente a consulta à obra na íntegra. Para as respostas sobre o diagnóstico, foi requisitada a indicação de resposta única; para as respostas sobre os acupontos, foi requisitada a sugestão de 1 a 8 respostas.

2.5 Análise dos dados

Os algoritmos *SimTCM* e PDA estão implementados em rotinas independentes em LabVIEW 2014 (National Instruments, EUA) e executados em um computador com 2.26 GHz Intel ® Core 2 Duo microprocessador com 2 GB RAM rodando em compilação em Mac OS X (Apple Inc., EUA).

Os parâmetros relacionados ao desempenho estatístico diagnóstico de cada examinador foram obtidos a partir das tabelas de contingência 2×2 (Tabela 2), feita a partir da simulação pelo *SimTCM* e identificação pelo examinador conforme a seguir:

- Condição de simulação pelo *SimTCM*: padrão verdadeiro (condição-alvo presente; =1) ou padrão falso (condição-alvo ausente; =0)

- Resultado da identificação pelo examinador: simulado = identificado (=1) ou simulado $\neq$ identificado (=0).

Os dados coletados dos questionários referentes ao diagnóstico e acupontos de cada paciente simulado foram tabulados conforme a Tabela 3 para análise estatística.

2.5.1 Variáveis de controle

De acordo com o STRICTA (MACPHERSON et al., 2001; MACPHERSON et al., 2010), o conhecimento dos praticantes de MTC influencia o modo como o praticante seleciona os acupontos e, portanto, se torna uma informação valiosa que pode ter efeito sobre os desfechos de um estudo e sobre a generalização dos resultados. Assim, as seguintes variáveis de controle serão coletadas: idade (anos), sexo, afiliação institucional (professores de pós-graduação; profissional autônomo), área de graduação em saúde, tempo de treinamento em acupuntura (horas), e tempo de formação em acupuntura (anos).

As variáveis de controle foram utilizadas para caracterização da amostra de examinadores e para avaliação da sua associação com as variáveis de desfecho.

2.5.2 Variáveis de desfecho

Primariamente, foram calculados os parâmetros *ACC*, *SEN*, *ESP*, *VPP* e *VPN* de cada examinador em comparação ao resultado da simulação. O coeficiente κ foi calculado separadamente para identificação de padrões, agrupado pelo grupo de pacientes simulados e identificados com pelo PDA com diagnóstico correto e incorreto. O coeficiente ι foi calculado para seleção de acupontos, agrupado por resultado de diferenciação pelo PDA (diagnóstico correto, e incorreto). Secundariamente, foram investigadas as correlações entre as variáveis de desfecho e de controle.

2.5.3 Análise estatística

A análise estatística foi realizada no programa R 3.1.1 (R CORE TEAM, 2014) utilizando-se os pacotes ‘boot’ (DAVISON & HINKLEY, 1997; CANTY & RIPLEY, 2014), ‘caret’ (KUHN, 2014), ‘irr’ (GAMER et al., 2012), ‘psy’ (FALISSARD, 2012) e ‘xlsx’ (DRAGULESCU, 2014) para Mac OS X (Apple Inc., EUA). O valor de significância estatística adotado será $p < 0,05$ (testes unicaudais).

Análise descritiva: Valores no texto e nas tabelas foram exibidos como mediana[mínimo; máximo] ou frequência absoluta e relativa (%) para variáveis contínuas ou categóricas, respectivamente. Histogramas foram gerados para mostrar a distribuição empírica das variáveis de concordância obtidas pelo método de reamostragem (*bootstrap*).

Análise de desempenho diagnóstico: Foram calculadas a *ACC*, *SEN*, *ESP*, *VPP* e *VPN* conforme as equações 1-5, por examinador e também agrupadas por resultado de diferenciação pelo PDA (diagnóstico correto e incorreto).

Análise de concordância interexaminador: Foi utilizado o coeficiente κ de Light (LIGHT, 1971) para avaliar a concordância para diferenciação de padrões, recomendado para estudos com cruzamento completo (*fully-crossed design*) em que todos os casos são classificados por múltiplos examinadores (HALLGREN, 2012). O coeficiente ι (JANSON & OLSSON, 2001) foi calculado para avaliar a concordância multivariada entre múltiplos examinadores para prescrição de acupontos. Os intervalos de confiança de 95% (IC95%) foram calculados com o procedimento *bootstrap* e o método *bias-corrected accelerated interval* (BCa) com 1000 replicações (EFRON & TIBSHIRANI, 1998). O *p* valor empírico foi calculado com o método de $(r+1)/(B+1)$, onde *r* representa o número de replicações que produzem uma estatística maior ou igual àquela calculada com os dados (NORTH et al., 2002). Estes parâmetros foram interpretados qualitativamente como concordância pobre (<0,00), leve (0,00 a 0,20), razoável (0,21 a 0,40), moderada (0,41 a 0,60), substancial (0,61 a 0,80) ou quase perfeita (0,81 a 1,00) (LANDIS & KOCH, 1977). Além da análise por rótulo do padrão, análises mais detalhadas de concordância para diferenciação de padrões e prescrição de acupontos foram realizadas considerando-se os sistemas internos como um dos principais atributos básicos que compõem o rótulo do padrão (Tabela 1).

Análise associativa: O coeficiente de correlação de *Spearman* foi utilizado para análise da correlação entre as variáveis quantitativas de controle e desfecho (KANJI, 2006). Os valores de correlação e seus respectivos níveis qualitativos de associação foram descritos como sem associação (0,00), associação desprezível ($\pm 0,01$ a $\pm 0,20$), fraca ($\pm 0,21$ a $\pm 0,40$), moderada ($\pm 0,41$ a $\pm 0,70$), forte ($\pm 0,71$ a $\pm 0,99$) ou perfeita ($\pm 1,00$) (WHO, 2001).

2.6 Aspectos éticos

O protocolo experimental foi submetido ao Comitê de Ética antes da execução do estudo, em consonância com a resolução 466/2012 e foi aprovado para realização (CAAE: 35723214.6.0000.5235) (Anexo 1). O presente estudo encontra-se registrado na Organização

Mundial de Saúde sob o número (*universal trial number*) U1111-1165-0313 e submetido ao Registro Brasileiro de Ensaios Clínicos (ReBEC), aguardando revisão.

Todos os indivíduos participantes assinaram um termo de consentimento livre e esclarecido (Apêndice 6) após serem informados da natureza do estudo e do protocolo a ser realizado.

- A) Análise crítica dos riscos:** Os questionários serão distribuídos de modo individual e anônimo para os participantes, assegurando-se a sua confidencialidade, privacidade, proteção da imagem e não-estigmatização. Entretanto, os participantes podem se sentir desconfortáveis ao por ter dificuldades em responder às questões técnicas do questionário. Para minimizar esse efeito, os resultados não serão divulgados individualmente e somente o respondente terá acesso ao seu resultado individual.
- B) Análise crítica dos benefícios:** Os benefícios incluem a informação ao participante sobre seu desempenho no diagnóstico dos casos simulados. Os benefícios gerais para a população podem incluir a diminuição de erros de diagnóstico na MTC e a prevenção de possíveis erros na prescrição dos acupontos, beneficiando a todos que utilizam a prática da MTC como intervenção terapêutica, prevenindo e/ou diminuindo os riscos e danos ao paciente.
- C) Responsabilidade do pesquisador e da instituição:** O pesquisador e a Instituição se responsabilizarão por qualquer dano pessoal ou moral referente à integridade física e ética que a pesquisa possa comportar.
- D) Critérios para suspender ou encerrar a pesquisa:** O estudo será imediatamente suspenso na ocorrência de qualquer falha metodológica ou técnica observada pelo pesquisador, cabendo ao mesmo a responsabilidade de informar a todos os participantes o motivo da suspensão. Também será suspenso caso seja percebido qualquer risco ou dano à saúde dos sujeitos participantes, consequentes à pesquisa, que não foram previstos no termo de consentimento. Quando atingir a coleta de dados necessária a pesquisa será encerrada.
- E) Local da pesquisa:** Esta pesquisa será realizada no Laboratório de Simulação Computacional e Modelagem em Reabilitação (LSCMR) do Programa de Pós-graduação em Ciências da Reabilitação, no Centro Universitário Augusto Motta, *campus* Bonsucesso (RJ).
- F) Demonstrativo de infraestrutura:** A Instituição onde será feito o estudo possui a infraestrutura necessária para o desenvolvimento da pesquisa com ambiente adequado. O LSCMR possui os softwares e computadores para processamento e análise dos dados.

- G) Propriedade das informações geradas:** Deve-se salientar que não haverá nenhuma cláusula restritiva para a divulgação dos resultados da pesquisa, e que os dados coletados serão utilizados única e exclusivamente para comprovação do experimento. Os resultados serão submetidos à publicação, sendo favoráveis ou não às hipóteses do estudo.
- H) Acesso aos dados do pesquisador/Instituição:** Em qualquer etapa do estudo, o participante poderá acessar o profissional responsável, Dra. Ingrid Jardim de Azeredo Souza Oliveira (CREFITO 49658-F) no telefone (21) 98318-0221. Se tiver alguma consideração ou dúvida sobre a ética da pesquisa, entre em contato com o Comitê de Ética em Pesquisa (CEP): Praça das Nações, nº 34 - Bonsucesso, Rio de Janeiro – RJ, Tel.: (21) 3882-9797 (ramal 1015), e-mail: comitedeetica@unisuam.edu.br.
- I) Indenização em caso de dano:** Em caso de dano pessoal diretamente causado pelos procedimentos propostos neste estudo, você terá direito a tratamento médico, bem como às indenizações legalmente estabelecidas, caso se justifiquem.
- J) Ressarcimento:** Não haverá despesas pessoais para você em qualquer fase do estudo, nem compensação financeira relacionada à sua participação.
- K) Sigilo:** O projeto será mantido em sigilo na íntegra por um prazo de 1 ano.

Capítulo 3 – Resultados

O Quadro 1 apresenta os manuscritos elaborados e submetidos para os periódicos científicos.

Quadro 1: Manuscritos submetidos.

Autores	Título	Periódico	Classificação WebQualis 2014
Ferreira AS, Oliveira IJAS	Are interrater agreements for pattern differentiation and acupuncture prescription influenced by diagnostic errors? A single-blinded study protocol	Journal of Integrative Medicine	B1 (IF: 0.916)
Oliveira IJAS, Ferreira AS	Effects of diagnostic errors in pattern differentiation and acupuncture prescription: A single-blinded, interrater agreement study	Evidence-based Complementary and Alternative Medicine	Sem classificação na Área 21 (IF: 2.175)

Manuscript 1 – *Journal of Integrative Medicine*

Are interrater agreements for pattern differentiation and acupuncture prescription different under diagnostic errors? A single-blinded study protocol

Abstract

Background: The influence of diagnostic errors on the agreement of traditional Chinese medicine (TCM) experts for pattern differentiation and acupuncture prescription remains unknown. This study protocol is designed to compare the interrater agreement of TCM experts on pattern differentiation and acupoints prescription between two groups with different diagnostic outcomes.

Methods & Design: This is a single-blinded, observational cross-sectional study. A dataset with 509 manifestations about internal organ's patterns will be organized using the four examinations. Three hundred human patients will be simulated, whereas 90 patients will be separated into three groups (n=30 each) according to the diagnostic outcome determined by a computational model: correct, incorrect, and missed diagnosis. A hardcopy questionnaire with 90 questions – one for each simulated patient – will be delivered face-to-face to a minimal of 5 TCM experts for the indication of a single diagnosis (among 73 patterns) and up to 8 acupoints (among 378 acupoints) for each simulated patient. Contingency tables 2×2 will be used for assessment of diagnostic performance (accuracy, sensitivity, specificity, positive and negative predictive values). Light's κ and Janson and Olsson's ι coefficients will be calculated for assessment of interrater agreement for pattern differentiation and acupuncture prescription, respectively.

Discussion: Planning studies with an integrative approach is challenging but a necessary step

towards the development of an evidence-based practice of TCM. It is expected that the results of this study provide new insights about the effects of diagnostic outcomes in interrater agreement for diagnosis and acupuncture prescription.

Funding resources: This study is supported by the grant (nº E-26/103.066/2012) from the Fundação Carlos Chagas Filho de Amparo à Pesquisa no Estado do Rio de Janeiro (FAPERJ).

Keywords: reliability; diagnosis; study protocol; traditional Chinese medicine; rehabilitation.

Trial Registration: UTM U1111-1165-0313, ReBEC (under review – pending).

1 Background

Traditional medicines experienced an increasingly interest in theoretical, experimental, and clinical research since its recognition in the Alma-Ata Declaration^[1]. Particularly, traditional Chinese medicine (TCM) was developed by a society geographic, social, and culturally distant from the Western community during the last 3,000 years^[2,3]. In Brazil, the free-access to Eastern medicine techniques using public health institutions for disease prevention, treatment or rehabilitation of the population also increased in recent years^[4], mainly due to the implementation in 2006 of the National Policy on Integrative and Complementary Practices^[5].

The diagnostic process of TCM has a unique feature: patterns, the TCM's counterpart of Western diseases, are identified through a process named pattern differentiation. The collection of clinical manifestations of an individual – the so-called manifestation profile – is obtained using the *four examinations*, namely inspection (IP), auscultation-olfaction (AO), inquiry (IQ) and palpation (PA)^[6]. As a corollary, the diagnosis is considered as subjective because only the five senses are used to gather meaningful clinical data and must be interpreted by a TCM expert; no equipment or diagnostic exam was developed for collecting

data for pattern differentiation until the last decades. Though TCM practitioners follow systematic-philosophic relationships between the humans and nature to identify the diagnosis^[7], TCM's pattern differentiation comprises a procedure subjected to errors as any other diagnostic system.

In Western medical practice, errors are difficult to recognize but are not rare^[8]. Many cases of diagnostic errors lie in the difficulty for differentiating between two closely related or similar diagnoses. Nonetheless, they might have very different prognosis or therapeutic options^[9,10]. Likewise, different TCM patterns might lead to distinct treatment choices such as herbs combinations and selection of acupoints. In contrast with Western medicine that has treatment protocols for various diseases, there are no defined protocols of acupoints for patterns because of the personalized aspect of TCM's diagnostic process and the possibility of selecting acupoints using a variety of criteria^[6]. Therefore, it is important to assess both the amount of agreement for TCM diagnosis and the diagnostic accuracy itself to determine its validity in both clinical and research scenarios, mainly among different raters. In this sense, high interrater agreement – the degree to which raters achieve identical results under similar assessment conditions – and diagnostic accuracy are important characteristics of any model used for classification^[11].

2 Significance and Previous Results

Previous studies investigated the agreement for pattern differentiation and/or for acupuncture prescription^[12-20], though they present important limitations either from the TCM or scientific perspective. Among these limitations we found a lack of calculating and reporting either statistical measures of agreement^[13-15,17] or a lack of investigating the relationship between diagnosis and therapeutic prescription^[14,16-20]. Finally, the above-cited studies used real human patients, in which the *true pattern* was unknown and therefore it is not possible to

assess accuracy with a gold-standard model. Most importantly, no study investigated the influence of diagnostic errors on both the agreement of TCM experts for pattern differentiation and for acupuncture prescription.

Our group developed automated systems to study the pattern differentiation process, namely the manifestation profile simulation algorithm (MPSA) and pattern differentiation algorithm (PDA)^[21] as a part of a large project on computerization of TCM research and practice^[22]. The MPSA was proposed^[23] for simulation of manifestation profiles with known diagnosis, with manifestation profiles representing either a case or a control individual. The same study^[23] introduced the PDA and its objective criterion (explained information, $F_{\%}$) for selection of candidate patterns and ranking them as diagnostic hypothesis. In a sequential study, the PDA incorporated^[24] a second objective criterion (available information, $N_{\%}$) for candidate pattern selection and diagnostic hypotheses ranking. Using the same dataset of 69 *zangfu* patterns in both studies, a statistically significant increase in diagnostic accuracy was found using both PDA's criteria $F_{\%}$ and $N_{\%}$ as compared to $F_{\%}$ alone (accuracy: 94.7% vs. 93.2%; sensitivity = 89.8% vs. 86.5%; specificity = 99.5% vs. 99.9%, respectively; $P < 0.001$). As related to the second criterion, a specific proportion of available information ($N_{\%} = 28.5\%$) was required to achieve the optimal diagnostic accuracy, which was recommended as a cutoff value. Recently, the MPSA was improved and incorporated in a more advanced model for simulation of human patients, namely the *SimTCM* model^[25].

The PDA and MPSA were used to investigate diagnostic errors in pattern differentiation; it was possible because the *true target-pattern* of the human patients were known from the simulation procedure^[26]. In that study, diagnostic outcomes were separated into correct diagnosis, misdiagnosis or no diagnosis and the database was expanded to 73 *zangfu* patterns. Using the four examinations the lowest misdiagnosis and undiagnosis rates were observed (6.0% and 1.4%, respectively) and shared manifestations among dual patterns

was identified as an important source of error in pattern differentiation.

3 Study Protocol Aims

This study protocol aims to investigate the interrater agreement for pattern differentiation and acupoints prescription under correct and incorrect diagnostic outcomes. More specifically, this protocol is designed (*i*) to compare the diagnostic accuracy and interrater agreement of TCM experts for pattern differentiation between two groups of human patients simulated with different diagnostic outcomes, (*ii*) to compare interrater agreement of TCM experts for acupoints prescription between the two groups of humans patients simulated with different diagnostic outcomes, and (*iii*) to describe the diagnostic performance of a sample of TCM experts using a questionnaire with human patients simulated with known diagnosis.

4 Methods/Study Design

This is an observational, cross-sectional study that follows the Guidelines for Reporting Reliability and Agreement Studies (GRRAS)^[11]. This study protocol also follows the Standards for Reporting Interventions in Clinical Trials of Acupuncture (STRICTA)^[27,28] where applicable for determining the characteristics of the TCM experts enrolled in the study. Figure 1 exhibits the study flowchart.

PLEASE INSERT FIGURE 1 HERE

4.1 Ethics

This protocol was elaborated in accordance with the resolution 466/2012 from the National Health Council^[29] and approved by the institutional ethics Committee prior to its execution (CAAE: 35723214.6.0000.5235). All participants must sign an informed consent form to

participate in this study after an explanation about the study aims, potential risks and benefits associated with their participation. A certified professional (IJASO) with a two-year training period (Instituto Sohaku-In, RJ, Brazil) and nine years of TCM clinical experience will perform the data collection. Computational routines were developed and implemented by the same researcher (ASF) who is a certified professional (Instituto Sohaku-In, RJ, Brazil) with a two-year training period and has 13 years of TCM clinical experience.

4.2 Sample size justification

The major factor that influences the power of statistical analysis is the number of cases; it is not advantageous to increase the number of raters because the effects on the statistical power and amplitude of confidence intervals are small. It was demonstrated that as the number of raters increases the required number of cases diminishes, although the saving in sample size rapidly decreases after five raters^[30]. Therefore, a minimum sample size of five raters was required.

Because the available formulas for determining the sample size consider up to 5 outcome categories (and our data contained 73 categories for diagnosis), the following procedure was adopted. A separated routine was written to derive samples sizes for 2, 3, 4, and 5 outcome categories considering the confidence interval perspective and parameters set to: $\kappa_0 = 0.75$, $95CI\%=[0.61; 0.80]$, and five raters. The obtained sample sizes for these outcome categories were then fitted to an exponential model, adjusted as

$n = 400.1789 \times category^{-0.622923}$, ($R^2 = 0.942$), where n is the required sample size. Using the fitted model for extrapolation to the number of patterns (=73), the minimum sample of 28 simulated patients per group was required.

4.3 Inclusion and exclusion criteria

Raters who simultaneously meet the following criteria will be included in the study: (i) bachelor degree in any health-related course recognized by the National Ministry of Education, including biomedicine, nursing, physical therapy, speech therapy, physician, nutrition, dentistry, psychology, and occupational therapy; (ii) postgraduate training in acupuncture registered at the respective professional council; (iii) clinical practice for at least one year; (iv) signature of the informed consent form after reading about the objectives, potential risks and benefits for participating in this research. Raters will be excluded from the study upon request to the principal investigator.

4.4 Participants screening and admission

Raters will be assessed for eligibility from postgraduate courses in acupuncture at local institutions nearby the Rio de Janeiro city (RJ, Brazil). The principal investigator will contact each rater in an independent manner for clarification of the study aims, methods and presentation of a critical analysis of potential risks and benefits. Raters will be blinded regarding the results of the other raters enrolled in this study and to the diagnostic outcomes of the questionnaire.

A self-administered questionnaire for sample characterization as well as for checking the adherence for inclusion criteria was developed with seven questions about personal and professional characteristics of the rater^[27,28]: age, sex, area of graduate course, professional occupation, period of completion of the course of expertise, time and duration of acupuncture training in the specialization course. Questionnaires will be filled *in loco* immediately after the participant's admission. Data analysis will be conducted at the Laboratory of Computational Modeling and Simulation in Rehabilitation (RJ, Brazil) after data collection from all raters.

4.5 Dataset of patterns

The database of patterns consists of 73 *zangfu* patterns from a previous work^[26], described by a total of 509 unique manifestations separated from examination method: Ip ($n = 103$, 20.2%), AO ($n = 31$, 6.1%), Iq ($n = 349$, 68.6%), Pa ($n = 26$, 5.1%). The consistency and quality of the database will be computationally tested before simulation of human patients to ensure that patterns are mutually exhaustive and have no duplicated manifestations in each pattern.

4.6 Simulation of human patients

Patients will be computer-simulated by selecting a variable and random number of manifestations considering all four examinations using the *SimTCM* model^[22,25]. For this simulation, it was assumed that the probability of each manifestation and pattern in the general population follows a uniform probability mass function. No user intervention is required other than this set-up of the simulation. The following outputs from *SimTCM* will be exported to a TXT file to be used in subsequent stages: (i) the label of the target-pattern under simulation; (ii) the label of the pattern randomly selected for simulation; (iii) the manifestation profile as comma-separated values; and (iv) the binary code representing the patient as "1" or "0" if the human patient simulated corresponds to either a true profile (*SimTCM* target-pattern = *SimTCM* simulated pattern) or a false profile (*SimTCM* target-pattern $\neq$ *SimTCM* simulated pattern). Three hundred pairs of true and false manifestations profiles will be simulated from all 73 patterns.

4.7 Assessment of pattern differentiation and diagnostic outcomes of the simulated data

The pattern differentiation of the simulated human patients will be performed by PDA using

its both criteria $F\%$ and $N\%$ -28.5%. PDA uses as input the manifestation profiles from the above-cited TXT file and outputs the identified pattern. No user intervention is required other than this set-up of the data analysis.

The analysis of the diagnostic outcomes will be performed in the following manner. The identified diagnosis by PDA will be compared to the label of the simulated pattern by *SimTCM* to generate three possible outcomes: ‘correct diagnosis’ (simulated pattern = identified pattern), ‘misdiagnosis’ (simulated pattern $\neq$ identified pattern) or ‘undiagnosis’ (no pattern was identified by PDA). Simulated patients with true and false profiles will be separated into categories according to the diagnostic outcome. False profiles with misdiagnosis will be excluded from the subsequent stages, as well as all simulated patients with the undiagnosis outcome.

4.8 Elaboration of the questionnaire with clinical cases of simulated human patients

A total of 60 patients will be randomly selected among the *true profiles* for each of the two diagnostic outcomes, being 30 with *correct diagnosis* and other 30 with *misdiagnosis*. Another 30 patients will be randomly selected among the *false profiles* with *correct diagnosis*, summing up 90 patients to compose the questionnaire. All questions will be prepared and presented clearly, avoiding dubious interpretation. The 90 patients will be randomly distributed into 30 groups of 3 cases each to avoid concentration of outcomes and thus a better distribution of diagnostic outcomes within the questionnaire. Randomizations will be performed separately for each one of the three groups of patients using an online pseudo-random generator of numbers and sequences (www.random.org).

4.9 Face-to-face interview of TCM experts using the questionnaire

The hardcopy of the questionnaire will be delivered face-to-face, containing a brief survey of

possible risks and benefits of this research to society and in particular to the respondent. The questionnaire is self-administered and the investigator will not interfere in its completion. The following material will be available to the rater: a hardcopy table with 74 response options (73 patterns, 1 option "It is not possible to identify the pattern"); a hardcopy table with 379 response options (361 channel acupoints^[31], 17 miscellaneous acupoints^[6], 1 option "No acupoint"); and the reference book used for the construction of database standards with their manifestations^[32]. To the raters will be allowed 10 minutes to familiarize themselves with the tables containing the response options, though no time limit will be posed to complete the form. To reproduce clinical conditions, the rater is allowed to consult the reference book, though shall not be encouraged to. For questions about the diagnosis, it will be required only one answer; for questions about the acupoints, it will be required between 1 and 8 answers. An individual, separated response form will be provided for filling the respective code regarding the identified pattern and the prescribed acupoints for each patient.

4.10 Data analysis

The rater's response to each patient will be paired to the numbering of the questions. The coding for responses about the pattern in range 1-74 and 1-379 for patterns and acupoints will be typed in an electronic worksheet also paired to the numbering of the questions.

Parameters related to the diagnostic performance of each rater will be obtained from 2×2 contingency tables (Table 1)^[21] made from the comparison between the results of the simulation by *SimTCM* (i.e. gold-standard) and the pattern differentiation by each rater (i.e. diagnostic test result): true-positives (*TP*; manifestation profiles simulated with the target-pattern that were correctly identified by the rater as present), true-negatives (*TN*; manifestation profiles not simulated with the target-pattern that were correctly identified by the rater as absent), false-negatives (*FN*; manifestation profiles simulated with the target-

pattern that were erroneously identified by the rater as absent), and false-positives (*FP*; manifestation profiles not simulated with the target-pattern that were erroneously identified by the rater as present).

PLEASE INSERT TABLE 1 HERE

4.11 Statistical analysis plan

Descriptive analysis. Data analysis will be reported based on single measurements of each rater; group values will be presented as median [minimum; maximum] for continuous variables, and absolute and relative frequencies (%) for categorical variables.

Interrater agreement of TCM experts. The Light's κ (kappa) coefficient^[33] recommended for studies with fully crossed design in which all cases are classified by multiple raters^[34] will be calculated for testing the agreement regarding pattern differentiation (1 variable: the identified pattern), grouped by diagnostic outcome. The Janson and Olsson's ι (iota) coefficient^[35] recommended for studies with multivariate analysis by multiple raters on the same participants will be calculated for testing the agreement regarding acupuncture prescription (8 variables: acupoint #1, acupoint #2, ..., acupoint #8). The 95% confidence interval [95% CI] will be calculated using the bootstrap procedure and bias-corrected accelerated method (BCa) with $B=1,000$ replications^[36]. The empirical p value will be calculated using $(r + 1)/(B + 1)$, where r is the number of replications that produce greater than or equal to that calculated with statistical data^[37]. Coefficients of agreement will be qualitatively interpreted as poor (< 0.00), slight (0.00 to 0.20), fair (0.21 to 0.40), moderate (0.41 to 0.60), substantial (0.61 to 0.80), or almost perfect (0.81 to 1.00)^[38].

Diagnostic performance of TCM experts. Accuracy, sensitivity, specificity, positive predictive value, and negative predictive values will be calculated (Table 1)^[39,40] for both

single rating and whole sample rating using the output from *SimTCM* as gold-standard and the output from TCM raters. These data will be grouped by diagnostic outcome, considering the true profiles with correct and incorrect pattern differentiation as performed by PDA.

Associative analysis. The Spearman correlation coefficient will be used to analyze the correlation between control and outcome variables^[41]. The correlation values and their qualitative levels of association will be described as no association (0.00), negligible (± 0.01 to ± 0.20), weak (± 0.21 to ± 0.40), moderate (± 0.41 to ± 0.70), strong (± 0.71 to ± 0.99) or perfect association (± 1.00)^[42].

The value of statistical significance is $p < 0.05$ (one-tailed tests).

4.12 Computational resources

A computer with 2.26 GHz Intel ® Core 2 Duo microprocessor with 2 GB RAM running Mac OS X 10.10 (Apple Inc., USA) will be used for data tabulation and analysis. The *SimTCM* and PDA algorithms are implemented as independent computational routines in LabVIEW 2014 (National Instruments, USA) running on Windows Vista compilation (Microsoft Corp., USA). Data from raters will be tabulated in an electronic worksheet in Excel for Mac 2011 (Microsoft Corp., USA) using automatic data validation. Statistical analysis will be performed in R 3.1.1^[43] using 'boot'^[44], 'caret'^[45], 'irr'^[46], 'kappaSize'^[47], 'psy'^[48], and 'xlsx'^[49] packages.

5 Discussion

This is a study protocol aiming to investigate the interrater agreement for pattern differentiation and acupoint prescription under correct and incorrect diagnostic outcomes. Developing a study protocol for unveiling interrater agreement and diagnostic accuracy for pattern differentiation is particularly challenging: to combine traditional theories with

contemporary instruments and procedures without mutual interference is virtually impossible.

Particularly, the major challenge comprised the need of a gold-standard method for assessing the accuracy of pattern differentiation and the presentation of these cases to the TCM experts in a systematic fashion. We chose to use patients simulation instead of real cases, and consequently to present them as a written list of manifestations. Although this methodological choice provides a gold-standard method for measuring accuracy, it might compromise the observation of clinically meaningful information for pattern differentiation. As a corollary, diagnostic accuracy itself may be underestimated as compared to real human patients from whom additional data can be obtained during the interview. Nonetheless, the use of written clinical cases has been the method of knowledge transmission regarding TCM for at least two millennia^[2,3,6] and thus it is consistent with the TCM's mode of thinking. Because face-to-face interviews have been considered as the approach of choice for data collection in epidemiological studies^[50], our questionnaire becomes an important tool in the study of diagnostic accuracy when combined with robust software for data simulation and analyzing the responses of TCM experts.

From the methodological point-of-view, this study protocol also highlights new methodologies to the TCM field, which are already available in epidemiology, statistics and related areas. The usage of separated coefficients of agreement for the univariate (pattern differentiation $\rightarrow \kappa$) and multivariate problems (acupuncture prescription $\rightarrow \iota$) comprises a novelty of our study in comparison to the previous ones on this subject^[12-20]. The majority of the previous studies used simple percent agreement for assessing agreement, which is not recommended mainly because it is not chance-corrected^[11,36]. Although both κ and ι have their own limitations, they possess desirable characteristics for coefficients of agreement and are applicable to TCM problems as well^[35-37].

The Chinese art of pattern differentiation for diagnosis is essentially a subjective

analysis. Therefore, the quantification and standardization of diagnostic procedures for TCM practitioners is an urgent need. The last two decades have witnessed the growth of importance of research related to computerization of pattern differentiation^[7,21]. Important advances in biomedical instrumentation and software development opened new possibilities to research on this important aspect regarding TCM examination. Therefore, planning research studies with an integrative design is challenging but a necessary step towards the development of an evidence-based practice of TCM. It is expected that the results of this study might provide new insights about the effects of diagnostic outcomes in interrater agreement for acupuncture prescription.

Competing interests

None declared.

Author's contribution

IJASO designed the study protocol and drafted the manuscript. ASF designed the study protocol, drafted the manuscript, and developed the computational methods. All authors revised and approved the final version of the manuscript.

Funding

This study is supported by the grant (nº E-26/103.066/2012) from the Fundação Carlos Chagas Filho de Amparo à Pesquisa no Estado do Rio de Janeiro (FAPERJ).

References

1. World Health Organization. Primary Health Care Report of the International Conference on Primary Health Care, Alma-Ata, USSR, 6–12 September 1978. Geneva, 1978.

[\[External link\]](#)

2. T'ao L. Chinese medicine during the Chin (1127-1234) and Yuan (1234-1368) eras. *Chin Med J.* 1955; 73(3):241–256. [[PubMed](#)]
3. Lee T, Cheng CF, Chang CS. Some early records of nervous and mental diseases in traditional Chinese medicine. *Chin Med J.* 1962; 81:55–59. [[PubMed](#)]
4. Fragoso APS. Ensaio clínicos com acupuntura: Questões bioéticas e metodológicas. *Perspectiva Online.* 2010; 4(15):151–158. [[External link](#)]
5. Brasil, Ministério da Saúde. Secretaria de Atenção à Saúde. Departamento de Atenção Básica. Política Nacional de Práticas Integrativas e Complementares no SUS - PNPIC-SUS. Brasília: Ministério da Saúde, 2006. 92 p. [[External link](#)]
6. O'Connor J, Bensky D. *Acupuncture a comprehensive text.* Seattle: Eastland Press; 1987.
7. Ferreira AS, Lopes AJ. Chinese Medicine Pattern Differentiation and Its Implications for Clinical Practice. *Chin J Integr Med.* 2011; 17(11):818–823. [[PubMed](#)]
8. Weingart SN, Wilson RM, Gibberd RW, Harrison B. Epidemiology of medical error. *Br Med J* 2000; 320(18):774-777. [[PubMed](#)]
9. Kuhn GJ. Diagnostic errors. *Acad Emerg Med.* 2002; 9(7):740–750. [[PubMed](#)]
10. Schiff GD, Hasan O, Kim S, Abrams R, Cosby R, Lambert BL, Elstein AS, Hasler S, Kabongo ML, Krosnjar N, Odwazny R, Wisniewski MF, McNutt RA. Diagnostic errors in medicine: analysis of 583 physician-reported errors. *Arch Int Med.* 2009; 169(20):1881–1887. [[PubMed](#)]
11. Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, Roberts C, Shoukri M, Streiner DL. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol.* 2011; 64(1):96–106. [[PubMed](#)]
12. Sung JJY, Leung WK, Ching JYL, Lao L, Zhang G, Wu JCY, Liang SM, Xie H, Ho YP, Chan LS, Bernam B, Chan FKL. Agreements among traditional Chinese medicine practitioners in the diagnosis and treatment of irritable bowel syndrome. *Aliment*

- Pharmacol Ther. 2004; 20(10):1205–1210. [[PubMed](#)]
13. Zhang GG, Lee W, Bausell B, Lao L, Handwerger B, Berman B. Variability in the traditional Chinese medicine (TCM) diagnoses and herbal prescriptions provided by three TCM practitioners for 40 patients with rheumatoid arthritis. *J Altern Complement Med.* 2005; 11(3):415–421. [[PubMed](#)]
 14. Zhang GG, Singh B, Lee W, Handwerger B, Lao L, Berman B. Improvement of agreement in TCM diagnosis among TCM practitioners for persons with the conventional diagnosis of rheumatoid arthritis: effect of training. *J Altern Complement Med.* 2008; 14(4):381–386. [[PubMed](#)]
 15. Coyetaux RR, Chen W, Lindemuth CE, Tan Y, Reilly AC. Variability in the diagnosis and point selection for persons with frequent headache by traditional Chinese medicine acupuncturists. *J Altern Complement Med.* 2006; 12(9):863–872. [[PubMed](#)]
 16. Mist S, Ritenbaugh C, Aickin M. Effects of questionnaire-based diagnosis and training on inter-rater reliability among practitioners of traditional Chinese medicine. *J Altern Complement Med.* 2009; 15(7):703–709. [[PubMed](#)]
 17. O'Brien KA, Abbas E, Zhang J, Guo Z, Luo R, Bensoussan A, Komesaroff PA. An investigation into the reliability of Chinese medicine diagnosis according to the eight guiding principles and zang-fu theory in Australians with hypercholesterolemia. *J Altern Complement Med.* 2009; 15(3):259–266. [[PubMed](#)]
 18. Grant SJ, Schnyer RN, Chang DH, Fahey P, Bensoussan A. Interrater Reliability of Chinese Medicine Diagnosis in People with Prediabetes. *Evid-Based Complem Altern Med.* 2013; 2013: Article ID 710892. [[PubMed](#)]
 19. Birkeflet O, Laake P, Vollestad NK. Poor multi-rater reliability in TCM pattern diagnoses and variation in the use of symptoms to obtain a diagnosis. *Acupunct Med.* 2014; 32(4):325–332. [[PubMed](#)]

20. Xu ZX, Xu J, Yan JJ, Wang YQ, Guo R, Liu GP, Yan HX, Qian P, Hong YJ. Analysis of the diagnostic consistency of Chinese medicine specialists in cardiovascular disease cases and syndrome identification based on the relevant feature for each label learning method. *Chin J Integr Med*. 2014 Jul 30. [Epub ahead of print] [[PubMed](#)]
21. Ferreira AS. Advances in Chinese medicine diagnosis: From traditional methods to computational models. In: Kuang H. Recent advances in Chinese medicine. Croatia: InTech. 2012. [[External link](#)]
22. Ferreira AS. Promoting the Integrative Medicine by the computerization of traditional Chinese medicine for scientific research and clinical practice: The SuiteTCM Project. *J Integr Med*. 2013; 11(2):135–139 [[PubMed](#)]
23. de Sá Ferreira A. Statistical validation of strategies for Zang-fu single pattern differentiation. *J Chin Integr Med*. 2008; 6(11):1109–1116. [[PubMed](#)]
24. Ferreira AS. Diagnostic accuracy of pattern differentiation algorithm based on traditional Chinese medicine theory: a stochastic simulation study. *Chin Med*. 2009; 4:24. [[PubMed](#)]
25. Ferreira AS, Pacheco AG. SimTCM: A human patient simulator with application to diagnostic accuracy studies of Chinese medicine. *J Integr Med*. 2014; December; Epub ahead of print. [[External link](#)]
26. Sá Ferreira A. Misdiagnosis and undiagnosis due to pattern similarity in Chinese medicine: a stochastic simulation study using pattern differentiation algorithm. *Chin Med*. 2011; 6:13. [[PubMed](#)]
27. MacPherson H, Altman DG, Hammerschlag R, Youping L, Taixing W, White A, Moher D, & STRICTA Revision Group. Revised STAndards for Reporting Interventions in Clinical Trials of Acupuncture (STRICTA): extending the CONSORT statement. *PLoS Medicine*. 2010; 7(6):e1000261. [[PubMed](#)]
28. MacPherson H, White A, Cummings M, Jobst K, Rose K, Niemtzw R. Standards for

- reporting interventions in controlled trials of acupuncture: The STRICTA recommendations. *STAndards for Reporting Interventions in Controlled Trails of Acupuncture*. *Acupunct Med*. 2002; 20(1):22–25. [[PubMed](#)]
29. Brasil. Conselho Nacional de Saúde. Resolução nº 466, 2012. Diretrizes e Normas regulamentadoras de pesquisa envolvendo seres humanos. Brasília, 13 jun. 2013. [[External link](#)]
30. Altaye M, Donner A, Eliasziw M. A general goodness-of-fit approach for inference procedures concerning the kappa statistic. *Stat Med*. 2011; 20(16):2479–2488. [[PubMed](#)]
31. World Health Organization. Standard acupuncture nomenclature. World Health Organization, Geneva: 1993. [[External link](#)]
32. Maciocia G. Os Fundamentos da Medicina Tradicional Chinesa: Um Texto Abrangente para Acupunturistas e Fitoterapeutas. Rio de Janeiro: Roca; 1996.
33. Light RJ. Measures of response agreement for qualitative data: Some generalizations and alternatives. *Psychol Bull*. 1971; 76(5):365–377. [[External link](#)]
34. Hallgren KA. Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutor Quant Methods Psychol*. 2012; 8(1):23–34. [[PubMed](#)]
35. Janson H, Olsson U. A measure of agreement for interval or nominal multivariate observations. *Educ Psychol Meas* 2001; 6(21):277–289. [[External link](#)]
36. Efron B, Tibshirani RJ. An introduction to the bootstrap. Florida, CRC Press, 1998.
37. North BV, Curtis D, Sham PC. A note on the calculation of empirical P-values from Monte Carlo procedures. *Am J Hum Genet*. 2002; 71(2); 439–441. [[PubMed](#)]
38. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977; 33(1):159–174. [[External link](#)]
39. Altman DG, Bland JM. Diagnostic tests. 1: Sensitivity and specificity. *BMJ*. 1994;308(6943):1552. [[PubMed](#)]

40. Altman DG, Bland JM. Diagnostic tests 2: Predictive values. *BMJ*. 1994; 309(6947):102.
[[PubMed](#)]
41. Kanji G. 100 statistical tests. 3a edição. Sage Publications, London, 2006.
42. World Health Organization. Health research methodology: A guide for training in research methods. World Health Organization, Geneva: 2001. [[External Link](#)]
43. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. [[External link](#)]
44. Canty A, Ripley B. boot: Bootstrap R (S-Plus) Functions. 2014. R package version 1.3-11. [[External link](#)]
45. Kuhn M. Contributions from Jed Wing, Steve Weston, Andre Williams, Chris Keefer, Allan Engelhardt, Tony Cooper, Zachary Mayer, Brenton Kenkel, the R Core Team and Michael Benesty. caret: Classification and Regression Training. 2014. R package version 6.0-37. [[External link](#)]
46. Gamer M, Lemon J, Singh IFP. irr: Various Coefficients of Interrater Reliability and Agreement. 2012. R package version 0.84. [[External link](#)]
47. Rotondi MA. kappaSize: Sample Size Estimation Functions for Studies of Interobserver Agreement. 2013. R package version 1.1. [[External link](#)].
48. Dragulescu AA. xlsx: Read, write, format Excel 2007 and Excel 97/2000/XP/2003 files. 2014. R package version 0.5.7. [[External link](#)]
49. Falissard B. psy: Various procedures used in psychometry. 2012. R package version 1.1. [[External link](#)]
50. Siemiatycki J, Campbell S. Nonresponse bias and early versus all responders in mail and telephone surveys. *Am J Epidemiol*. 1984, 120(2):291–301. [[PubMed](#)]

Figure legends

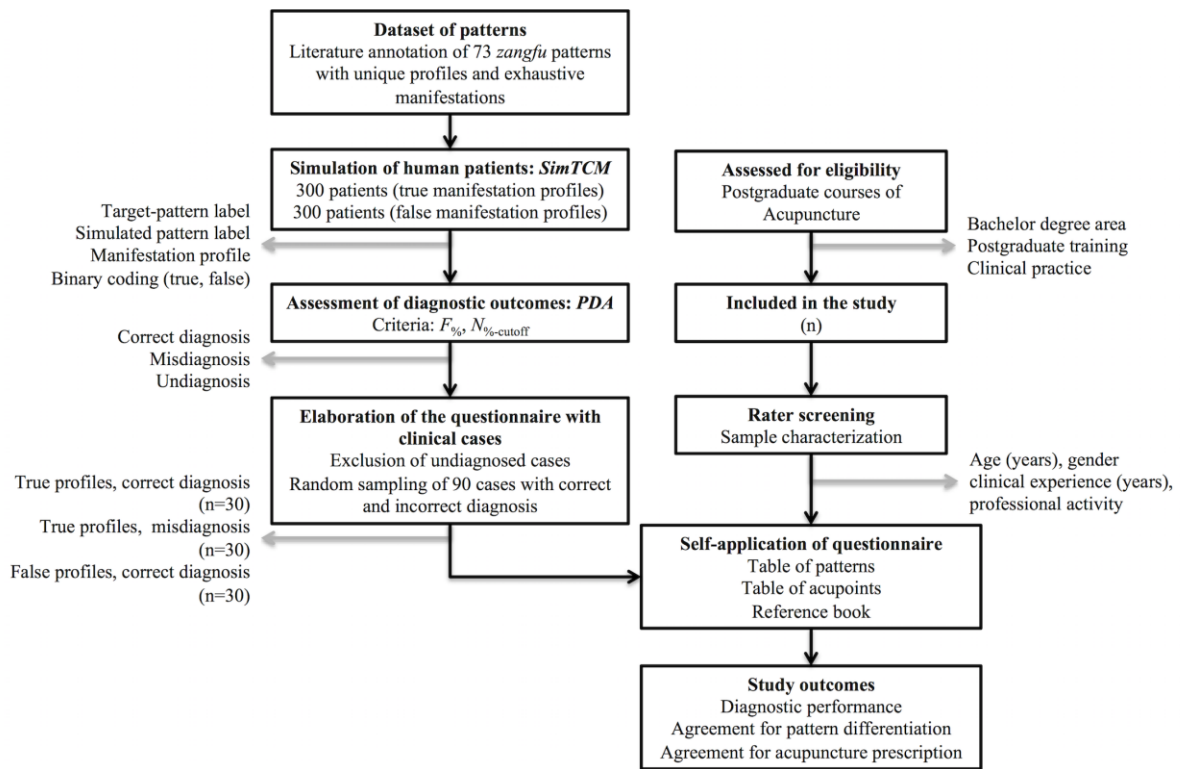


Figure 1: Study flowchart.

Table 1: Contingency table for assessment of diagnostic performance of traditional Chinese medicine experts using human patients simulation.

		<i>SimTCM</i> output (Reference test)		Parameters
		True manifestation profile ^A (Simulated = target)	False manifestation profile ^B (Simulated ≠ target)	
Result from TCM expert (New test)	Target-pattern + (Identified = target)	True-positive (<i>TP</i>)	False-positive (<i>FP</i>)	$PPV = \frac{TP}{TP + FP}$
	Target-pattern - (Identified ≠ target)	False-negative (<i>FN</i>)	True-negative (<i>TN</i>)	$NPV = \frac{TN}{TN + FN}$
	Parameters	$SEN = \frac{TP}{TP + FN}$	$SPE = \frac{TN}{TN + FP}$	$ACC = \frac{TP + TN}{TP + FP + TN + FN}$

^A Target-pattern present; ^B Target-pattern absent. *SEN*: sensitivity; *SPE*: specificity; *PPV*: positive predictive value; *NPV*: negative predictive value; *ACC*: accuracy.

**Manuscripto 2 – Evidence-based Complementary Alternative
Medicine**

**Effects of diagnostic errors in pattern differentiation and acupuncture
prescription: A single-blinded, interrater agreement study**

Authors: Ingrid Jardim de Azeredo Souza Oliveira¹, Arthur de Sá Ferreira¹

Citation (Pubmed format): Oliveira IJAS, Ferreira AS

Affiliations: ¹Laboratory of Computational Simulation and Modeling in Rehabilitation,
Postgraduate Program of Rehabilitation Science, Centro Universitário Augusto
Motta/UNISUAM, Rio de Janeiro, Brazil

Corresponding address: Arthur de Sá Ferreira, D.Sc., Postgraduate Program of
Rehabilitation Science, Centro Universitário Augusto Motta/UNISUAM, Praça das Nações
34, Bonsucesso, Rio de Janeiro, RJ, CEP 21041-010, Tel: +5521 3882-9797 (extension 1015)

E-mails:

Ingrid Jardim de Azeredo Souza Oliveira (Oliveira IJAS): guigarden@yahoo.com.br

Arthur de Sá Ferreira (Ferreira AS): arthur_sf@ig.com.br; arthurde@unisuamdoc.com.br

Tables: 2

Figures: 3

References: 45

Word count (abstract): 199

Keywords: 4

Word count (main text): 5460

Abstract

This single-blinded, observational cross sectional study compared the interrater agreement of TCM experts for pattern differentiation and acupoints prescription between two groups of human patients simulated with different diagnostic outcomes. A dataset with 509 manifestations about *zangfu* patterns was organized. Patients were simulated and separated into groups (n=30 each) according to the diagnostic outcome determined by a computational model: correct and incorrect. A hardcopy questionnaire with 90 simulated patient was delivered face-to-face to 6 TCM experts (4 males, 4-year minimal of clinic experience) who were asked to indicate a single pattern (among 73) and 8 acupoints (among 378). Interrater agreement was higher for pattern differentiation than for acupuncture prescription. Interrater agreement on pattern differentiation was classified as slight for both groups with correct diagnosis (Light's kappa=0.167, 95%CI=[0.108; 0.254]) and incorrect diagnosis outcome (Light's kappa=0.190, 95%CI=[0.120; 0.286]). Interrater agreement on acupuncture prescription was also slight for both groups of correct (iota=0.029, 95%CI=[0.015; 0.057]) and incorrect diagnosis (iota=0.040, , 95%CI=[0.023; 0.058], p=0.075). Analysis of diagnostic performance of raters yielded accuracy=60.9%, sensitivity=21.7%, and specificity=100%. An overall improvement in the interrater agreement and diagnostic accuracy was observed when the data were analyzed using codes for the internal systems instead of the pattern's labels.

Keywords: reliability; diagnosis; traditional Chinese medicine; rehabilitation.

1. Introduction

Diagnostic errors are difficult to recognize but are not rare in the Western practice [1]. The difficulty for differentiating between two closely related or similar diagnoses, with possibly very different prognosis or therapeutic options, is acknowledged as a source of error [2,3]. Traditional medicines, in particular traditional Chinese medicine (TCM), are no exceptions. The systematic-philosophic relationships between humans and nature applied by TCM experts [4] do not guarantee an error-free diagnostic process [5]; different patterns that require distinct therapeutic choices regarding acupuncture prescription might also be confused. In contrast with the available treatment regimens for various diseases in the Western medicine, there are no defined protocols of acupoints for patterns mainly because of both the personalized approach of TCM's diagnostic process and the large variety of criteria for selecting acupoints [6]. In this sense, high interrater agreements – i.e. the degree to which raters achieve identical results under similar assessment conditions rating the same items [7] – alongside an accurate diagnosis are two important characteristics of diagnostics models.

Previous studies reported a variable degree of interrater agreement on pattern differentiation and/or therapeutic prescription [8-16], though they present important limitations either from the TCM or scientific perspectives. For instance, there were a lack of calculating and/or reporting statistics of agreement [9-11,13], or a lack of investigating the pattern differentiation and therapeutic prescription for the same cases [10,12-16]. Most importantly, all the above-cited studies used real human patients with a narrow range of diseases and corresponding TCM patterns, in which the *true patterns* were unknown – and therefore it was not possible to assess the diagnostic accuracy of TCM experts with a gold-standard model. However, no study investigated the magnitude of the influence of diagnostic errors on the agreement of TCM experts for pattern differentiation and therapeutic prescription.

Our group developed automated systems to study the pattern differentiation process. The manifestation profile simulation algorithm (MPSA) was proposed [17] for simulation of cases and controls with known diagnosis; the current version of MPSA is implemented in the *SimTCM* model [18,19]. The pattern differentiation algorithm (PDA) was introduced in the same study [17] and its current version [20] applies two objective criteria – explained ($F\%$) and available information ($N\%$), respectively – for selection of candidate patterns and ranking them as diagnostic hypothesis. Both PDA and MPSA were used to investigate diagnostic errors in pattern differentiation; because the true target-pattern of human patients was known from the *SimTCM* and were paired to the diagnosis obtained from the PDA [5], the diagnostic outcomes were separated into correct diagnosis, misdiagnosis or no diagnosis. Using the four examinations, the lowest misdiagnosis and undiagnosis rates among 73 *zangfu* patterns were observed and shared manifestations among dual patterns was identified as an important source of error in pattern differentiation [5].

Assessing the interrater agreement of TCM experts for pattern differentiation and acupuncture prescription under different diagnostic outcomes might provide new insights about the causes and/or consequences of diagnostic errors in this traditional medical practice. Therefore, this study investigated the effects of diagnostic outcomes on the interrater agreement for pattern differentiation and acupoint prescription. More specifically, this study compared the interrater agreement of TCM experts for pattern differentiation and acupoints prescription between two groups of simulated patients with different diagnostic outcomes.

2. Material and Methods

2.1 Study design

This was an observational, cross-sectional study that followed the Guidelines for Reporting Reliability and Agreement Studies (GRRAS) [7]. This study also followed the STandards for

Reporting Interventions in Clinical Trials of Acupuncture (STRICTA) [21,22] where applicable for determining the characteristics of the TCM experts enrolled in the study. Figure 1 exhibits the study flowchart. Raters were blinded regarding the results of the other raters enrolled in this study.

<PLEASE INSERT FIGURE 1 HERE>

2.2 Ethics

The institutional ethics committee approved this study prior to its execution (CAAE: 35723214.6.0000.5235). All participants signed an informed consent form to participate in this study after being informed about the study aims, potential risks and benefits related to their participation.

2.3 Raters screening and admission

Raters were independently assessed for eligibility from postgraduate courses in acupuncture at local institutions nearby the Rio de Janeiro city (RJ, Brazil). Raters who simultaneously met the following criteria were included in the study: (1) bachelor degree in any health-related course recognized by the National Ministry of Education; (2) postgraduate training in acupuncture registered at the respective professional council; (3) clinical practice for at least one year; (4) signature of the informed consent form after reading about the objectives, potential risks and benefits for participating in this research. All raters enrolled in this study answered *in loco* a self-administered questionnaire about their personal and professional characteristics for sample characterization.

2.4 Dataset of patterns

The database of patterns consisted of 73 *zangfu* patterns developed for a previous work, presenting 509 unique manifestations separated from examination method: inspection ($n=103$, 20.2%), auscultation-olfaction ($n=31$, 6.1%), inquiry ($n=349$, 68.6%), and palpation ($n=26$, 5.1%). The consistency and quality of the database was computationally tested before simulation of human patients to ensure that patterns were mutually exhaustive and had no duplicated manifestations describing the same pattern [5].

2.5 Simulation of human patients

Patients were simulated by selecting a variable and random number of manifestations considering all four examinations using the *SimTCM* model [18,19]. For this simulation, it was assumed that the probability of each manifestation and pattern in the general population follows a uniform probability mass function. No user intervention was required other than the initial set-up. The set of manifestations in a given pattern have mutual relationships due to the same or similar pathogenesis. Likewise, different patterns might share manifestations due to the same or similar etiology. This mutual relationship between the randomly selected manifestations for a given pattern was ensured by selecting the manifestations for a given pattern – either k or q – from the same *pattern* as described in the dataset. It is worth noticing that the dataset was developed to simultaneously maintain the internal relationship among manifestations of a given pattern and the co-occurrence of the same manifestation among different patterns.

The *SimTCM* model output a TXT file used in the subsequent stages of this research with the following data: (1) the label of the target-pattern under simulation; (2) the label of the pattern randomly selected for simulation; (3) the manifestation profile as comma-separated values; and (4) the binary code representing the patient as ‘1’ or ‘0’ if the simulated patient corresponds to either a true profile (target-pattern = simulated pattern) or a false profile

(target-pattern $\neq$ simulated pattern), respectively. A total of 300 pairs of true and false manifestations profiles were simulated among all 73 patterns.

2.6 Assessment of diagnostic outcomes of the simulated sample of patients

To objectively test whether the simulated patient could have its pattern differentiated (either correctly or not) among several candidates in a dataset, an automated model for pattern differentiation must be used. PDA was chosen among other automated methods because it is the most accurate automated method for pattern differentiation considering a large set of patterns regardless of an underlying medical disease [23].

Pattern differentiation of the sample of simulated patients was performed by the PDA using its both criteria $F\%$ and $N\%-28.5\%$ for separation into diagnostic outcomes. PDA's diagnostic performance is high for simulated patients (accuracy = 94.7%; sensitivity = 89.8%; specificity = 99.5%) [20]. PDA used as input the manifestation profiles from the above-cited TXT file and output the identified pattern. The diagnostic outcomes were obtained by comparing the label of the simulated pattern by *SimTCM* with the label of the identified pattern by PDA, yielding one of these three outcomes: 'correct diagnosis' (simulated pattern = identified pattern), 'incorrect diagnosis' (simulated pattern $\neq$ identified pattern), or 'undiagnosis' (no pattern was identified as a diagnosis by PDA) [5]. Simulated patients with true and false profiles were separated into categories of correct and incorrect diagnostic outcome. Simulated patients with the undiagnosis outcome were excluded from subsequent stages because they lack a pattern label identified from PDA for comparison with TCM experts. Likewise, false profiles with incorrect diagnosis were excluded from the subsequent stages because they do not characterize a true-negative condition (i.e. false profiles with a correct diagnosis). Therefore, false profiles with incorrect diagnosis do not contribute to the analysis of diagnostic accuracy using the 2x2 contingency table.

172

173 *2.7 Elaboration of the questionnaire with clinical cases of simulated human patients*

174 Sixty simulated patients were randomly selected among the *true profiles* for each of the two
 175 diagnostic outcomes, being 30 with correct diagnosis and other 30 with incorrect diagnosis.
 176 Other 30 patients were randomly selected among the *false profiles* with correct diagnosis,
 177 summing up 90 patients to compose the questionnaire. More explicitly, a true profile with
 178 incorrect diagnosis represents a patient with a given target-pattern who was (mis)diagnosed
 179 with another pattern. Such a kind of manifestation profile is required to test the ability of the
 180 rater to correctly identify the target-pattern under conditions in which the automated method
 181 PDA failed. On the contrary, a false profile with correct diagnosis comprises a patient without
 182 the target-pattern who was (mis)diagnosed with the target-pattern. This kind of data is
 183 necessary to test the ability of the rater to correctly exclude the target-pattern under conditions
 184 in which the automated method PDA failed again.

185 All questions were prepared and presented clearly, avoiding dubious interpretation.

186 The 90 patients were then randomly distributed into 30 groups of 3 cases each to avoid
 187 concentration of outcomes and thus providing a better distribution of diagnostic outcomes
 188 within the questionnaire.

189

190 *2.8 Face-to-face interview of TCM experts using the questionnaire*

191 The hardcopy of the questionnaire were delivered face-to-face by one author who is a
 192 certified professional (IJASO) with a two-year training period and nine years of TCM clinical
 193 experience.

194 The questionnaire was self-administered and the investigator did not interfere in its
 195 filling. The following material was available to the rater for usage and consulting: a hardcopy
 196 table with 74 response options (73 *zangfu* patterns [24] + 1 option "It is not possible to

identify the pattern"); a hardcopy table with 379 response options (361 channel acupoints [25] + 17 miscellaneous acupoints [6] + 1 option "No acupoint"); and the reference book used for the construction of database standards with their manifestations [24]. To the raters was allowed 10 minutes to familiarize themselves with the tables containing the response options, though no time limit will be posed to complete the form. To reproduce clinical conditions, the rater was allowed to consult the reference book, though was not encouraged to. Only one answer was required for questions about the diagnosis, whereas between 1 and 8 answers for questions were required about the acupoints. An independent response form was provided for each rater for filling the respective code regarding the identified pattern and the prescribed acupoints for each simulated patient in the questionnaire.

2.9 Raw data tabulation and synthesis

Data analysis was conducted at the Laboratory of Computational Modeling and Simulation in Rehabilitation (RJ, Brazil) after data collection from all raters. The rater's responses to each patient were paired to the numbering of the questions. The coding for responses about the specific 74 options for diagnosis and 379 options for prescription were typed into an electronic worksheet by one researcher (IJASO) and checked by the other one (ASF), also paired to the numbering of the questions.

Because a specific *zangfu* pattern's label is composed by several pieces of information regarding the morbid condition itself (i.e. the nature, the location, and the affected internal organ or channel), a more general set of codes was provided regarding the affected internal system. By doing this, it was possible to investigate this study's outcome regarding specific and general aspects of TCM pattern differentiation. Therefore, *zangfu* patterns were coded according to the affected *zangfu* itself: heart (*xin*=1), liver (*gan*=2), spleen (*pi*=3), lung (*fei*=4), kidneys (*shen*=5), pericardium (*xinbao*=6), small intestine (*xiaochang*=7), gallbladder

(*dan*=8), stomach (*wei*=9), large intestine (*dachang*=10), bladder (*pangguang*=11), and triple energizer (*sanjiao*=12). Because more than one *zangfu* system was affected in the same pattern (e.g. *fei-pi-shen* to generate *tanyin*), they were combined as separate codes: liver-gallbladder (*gan-dan*=13), stomach-spleen (*wei*=14), kidney-lung (*shen-fei*=15), kidney-heart (*shen-xin*=16), lung-spleen-kidney (*fei-pi-shen*=17), and lung-spleen-kidney-xin (*fei-pi-shen-xin*=18). Acupoints were coded with the same sequence as the internal organs for correspondence with the specific channels (codes 1 to 12 only); additional codes were provided for the governing vessel (*dumai*=19), conception vessel (*renmai*=20), extra-channel acupoints (=21), and no acupoints (=22).

Parameters related to the diagnostic performance of each rater were obtained from 2×2 contingency tables [23] made from the comparison between the results of the simulation by *SimTCM* (gold-standard method) and the pattern differentiation by each rater:

- *True-positives (TP)*: manifestation profiles simulated with the target-pattern that were correctly identified by the rater as present;
- *False-negatives (FN)*: manifestation profiles simulated with the target-pattern that were erroneously identified by the rater as absent;
- *False-positives (FP)*: manifestation profiles not simulated with the target-pattern that were erroneously identified by the rater as present; and
- *True-negatives (TN)*: manifestation profiles not simulated with the target-pattern that were correctly identified by the rater as absent.

2.10 Statistical analysis

Data analysis was reported based on single measurements of each rater. Values were presented as median [minimum; maximum] for continuous variables, and absolute and relative frequencies (%) for categorical variables. Histograms of the empirical values obtained

for the coefficients of agreement under the bootstrap procedure were generated. All statistics were grouped by diagnostic outcome, considering the true profiles with correct and incorrect pattern differentiation as performed by PDA. The value of statistical significance is $p < 0.05$ (one-tailed tests).

The Light's κ (kappa) coefficient [26] recommended for studies with fully crossed design in which all cases are classified by multiple raters [27] was calculated, grouped by diagnostic outcome. The Janson and Olsson's ι (iota) coefficient [28] recommended for studies with multivariate analysis by multiple raters on the same participants was calculated for testing the agreement regarding acupuncture prescription. The 95% confidence interval (95%CI) was calculated using the bootstrap procedure and bias-corrected accelerated method (BCa) with $B=1,000$ replications [29]. Empirical p values were calculated using $(r + 1)/(B + 1)$, where r is the number of replications that produce greater than or equal to that calculated with statistical data [30]. Null hypotheses for group with correct diagnosis comprised $\kappa=0$ and $\iota=0$ (agreement not better than chance), whereas the null hypotheses for the group with incorrect diagnosis comprised $\kappa_{\text{incorrect}}=\kappa_{\text{correct}}$ and $\iota_{\text{incorrect}}=\iota_{\text{correct}}$ (no difference between correct and incorrect diagnoses) for pattern differentiation and acupuncture prescription, respectively. Both coefficients of agreement were qualitatively interpreted as poor (< 0.00), slight (0.00 to 0.20), fair (0.21 to 0.40), moderate (0.41 to 0.60), substantial (0.61 to 0.80), or almost perfect (0.81 to 1.00) [31].

Accuracy, sensitivity, specificity, positive and negative predictive values were calculated for both single rating and grouped data [32,33]. Accuracy was tested using a binomial procedure to assess if this rate was better than chance (i.e. no information rate, $H_0=50\%$).

The Spearman's ρ correlation coefficient [34] was used to analyze the association between control and outcome variables. The correlation values and their qualitative levels

were described as: no association (0.00), negligible association (± 0.01 to ± 0.20), weak association (± 0.21 to ± 0.40), moderate association (± 0.41 to ± 0.70), strong association (± 0.71 to ± 0.99), and perfect association (± 1.00) [35].

The major factor that influences the power of statistical analysis is the number of cases; it is not advantageous to increase the number of raters because the effects on the statistical power and amplitude of confidence intervals are small [36]. It was demonstrated that as the number of raters increases the required number of cases diminishes, although the saving in sample size rapidly decreases after five raters [37]. Therefore, a sample size of five raters was required to investigate 30 patients per category. Because the available formulas for determining the sample size consider up to 5 outcome categories (and our data contained 73 categories for diagnosis), the following procedure was adopted. A separated routine was written to derive samples sizes for 2, 3, 4, and 5 outcome categories considering the confidence interval perspective and parameters set to: $\kappa_0 = 0.75$, $95\text{CI}\% = [0.61; 0.80]$, and five raters. The obtained sample sizes for these outcome categories were then fitted to an exponential model, adjusted as $n = 400.1789 \times \text{category}^{0.622923}$, ($R^2 = 0.942$), where n is the required sample size. Using the fitted model for extrapolation to the number of patterns (=73), the minimum sample of 28 simulated patients per group was required

2.11 Computational resources

A computer with 2.26 GHz Intel ® Core 2 Duo microprocessor with 2 GB RAM running Mac OS X 10.10 (Apple Inc., USA) was used for data tabulation and analysis. The *SimTCM* and PDA algorithms were implemented as independent computational routines in LabVIEW 2014 (National Instruments, USA) running on Windows Vista compilation (Microsoft Corp., USA). Data from raters were tabulated in an electronic worksheet in Excel for Mac 2011 (Microsoft Corp., USA) using automatic data validation. Statistical analysis was performed in

R 3.1.1 [38] using 'boot' [39], 'caret' [40], 'irr' [41], 'kappaSize' [42], 'psy' [43], and 'xlsx' [44] packages using customized routines for data reading and analyses. Randomizations were performed using an online pseudo-random generator of numbers and sequences (www.random.org). Computational routines were developed and implemented by the same researcher (ASF) who is a certified professional with a two-year training period and 13 years of TCM clinical experience. All tables with raw data and computational routines for data analysis in R language are freely available from the authors upon request by e-mail.

3. Results

Seventeen TCM raters were independently assessed for eligibility between November 2014 and December 2014. Eleven raters were excluded because they were unavailable during the period for data collection (n=9), not yet registered in the professional council (n=1), or missed the appointment after two scheduling (n=1). Six raters (all physiotherapists, one also a medical doctor) were included in the study and filled in both questionnaires (Table 1). All raters stated that they perform pattern differentiation before acupuncture treatment selection in their daily professional activity.

<PLEASE INSERT TABLE 1 HERE>

Figure 2 exhibits the empirical histograms for the bootstrap resampling of κ and ι considering the *zangfu* patterns coded by their specific labels. In general, interrater agreement was higher for pattern differentiation than for acupuncture prescription. A significant, slight interrater agreement on pattern differentiation was observed for simulated patients with the correct diagnosis outcome ($\kappa=0.167$, 95%CI=[0.108; 0.254], $p<0.001$). No significant difference in interrater agreement was observed for simulated patients with incorrect

diagnosis outcome ($\kappa=0.190$, 95%CI=[0.120; 0.286], $p=0.330$). Likewise, a significant slight significant interrater agreement for acupuncture prescription was observed for the group of simulated patients with correct ($\kappa=0.029$, 95%CI=[0.015; 0.057], $p<0.001$) diagnostic outcome, although no significant statistical change was noticed for the group with the incorrect diagnosis outcome ($\kappa=0.040$, 95%CI=[0.023; 0.058], $p=0.075$).

<PLEASE INSERT FIGURE 2 HERE>

An overall improvement in the interrater agreement was observed when the simulated patients were analyzed using the codes for the internal systems (Figure 3); again, interrater agreement was higher for pattern differentiation than for acupuncture prescription. A significant, fair interrater agreement on pattern differentiation was observed for simulated patients with the correct diagnosis outcome ($\kappa=0.216$, 95%CI=[0.156; 0.309], $p<0.001$). No significant change in interrater agreement was observed for simulated patients with incorrect diagnosis outcome ($\kappa=0.248$, 95%CI=[0.170; 0.339], $p=0.256$). Nonetheless, a significant yet slight interrater agreement for acupuncture prescription was observed for either group of simulated patients with correct ($\kappa=0.046$, 95%CI=[0.024; 0.091]) and incorrect ($\kappa=0.062$, 95%CI=[0.037; 0.093], $p=0.102$) diagnostic outcome, again without statistical significance between diagnostic outcomes.

<PLEASE INSERT FIGURE 3 HERE>

Table 2 summarizes the results of the diagnostic performance of each rater, as well as the group summary. The diagnostic performance of the raters for identifying *zangfu* patterns with correct diagnosis showed that only 2 raters (33%) performed pattern differentiation

better than chance (accuracy= 60.9% [56.7; 70.0]). Sensitivity was low (21.7%), whereas specificity was high (100%); positive predictive values were also high (100%), with low negative predictive values (56.1%). Similar results were observed for identifying *zangfu* patterns with incorrect diagnosis according to PDA. Two raters (33%) performed pattern differentiation better than chance (accuracy= 58.3% [51.7, 63.3]), and presented low sensitivity (16.7%), high specificity and positive predictive values (100%), and low negative predictive value (54.5%).

<PLEASE INSERT TABLE 2 HERE>

An overall improvement in the diagnostic performance was also observed for the pattern differentiation using the coding by the affected internal system. Five raters (83%) performed pattern differentiation better than chance in cases with correct diagnosis (accuracy= 66.7% [55.0; 73.3]). Again, sensitivity was low (36.7%) and specificity conversely high (93.4%). Positive predictive values were also high (88.0%), with low negative predictive values (60.5%). Regarding patients with incorrect diagnosis, the same five raters (83%) performed pattern differentiation better than chance (accuracy= 67.5% [58.3, 75.0]), and presented low sensitivity (38.3%), high specificity and positive predictive values (93.4% and 88.3%, respectively), and low negative predictive value (61.1%).

The association analysis regarding the correct diagnostic outcome showed no significant positive correlations of diagnostic accuracy with being a postgraduate course professor ($\rho=0.399$, $p=0.217$), time since postgraduate ($\rho=0.277$, $p=0.297$), or age ($\rho=0.087$, $p=0.435$). No significant negative correlations were also observed between diagnostic accuracy and being supervisor of a clinic-school ($\rho=-0.396$, $p=0.218$) or sex ($\rho=-0.105$, $p=0.422$). No significant positive correlations were also observed for diagnostic accuracy and

being a postgraduate course professor ($\rho=0.399$, $p=0.217$), time since postgraduate ($\rho=0.277$, $p=0.297$), or age ($\rho=0.232$, $p=0.329$). No significant negative correlations were observed between diagnostic accuracy and sex ($\rho=-0.315$, $p=0.272$) or being supervisor of a clinic-school ($\rho=-0.198$, $p=0.353$).

4. Discussion

This study investigated the effects of diagnostic outcomes on the TCM interrater agreement for pattern differentiation and acupoint prescription using both specific and general characteristics of patterns. The main findings of our study comprised: (1) interrater agreements on pattern differentiation and acupoint prescription were slight regardless of whether the diagnosis was accurate or not, (2) interrater agreement was fair for differentiating combinations of affected internal organs, although it remained slight for acupuncture prescription, (3) diagnostic accuracy of TCM raters for differentiating *zangfu* patterns was not better than chance for most raters, although it was better for differentiating the affected internal organs, and (4) no significant effects of personal and professional variables were detected. To the best of our knowledge, this is the first study to simultaneously assess the interrater agreement for pattern differentiation and acupuncture prescription in specific and general TCM theories, along with the possible effects of the diagnostic errors in these outcomes.

The observed low agreement for TCM diagnosis may be due to several factors, particularly related to TCM theories and methodological aspects of this study, which are discussed in a separated section. As related to TCM theories, the systematic-philosophic relationship used for pattern differentiation tries to provide clear distinctions among all *zangfu* patterns [4,6]. However, it does not guarantee the occurrence of pathognomonic manifestations; several manifestations are shared among patterns and therefore may confuse

the TCM expert when diagnosing a patient [5]. The lack of international standards for describing each pattern's manifestation and no training before filling in questionnaires are another two potential source of variability in diagnosis, albeit the latter condition is a close representation of the clinical scenario. Nonetheless, our results are consonant with previous ones investigating TCM agreement on diagnosis of specific *zangfu* patterns: slight interrater agreement (unspecified $\kappa=0.11$; four raters) [8]; interrater agreement below chance expectations (κ values not available; three raters) [9]; 'little' agreement (κ values not available; three raters) [11]; slight interrater agreement (Fleiss' $\kappa=0.112$; four pairs of raters) [12]; and slight interrater agreement (κ in range 0.014 to 0.179, eight raters) [15]. In contrast, our results were lower than those observed by Xu et al. [16], who reported slight to almost perfect interrater agreement (Cohen's κ in range 0.005 to 0.801; two raters). However, the raters applied very specific diagnostic criteria that might have helped achieve so high agreements. Our observations that interrater agreement was better for more generic aspects of TCM diagnosis and prescription are also similar to others studies: interrater agreements varying between slight ($\kappa=0.15$) and almost perfect ($\kappa=0.87$) among three raters [13]; and moderate interrater agreement (Cohen's $\kappa=0.56$; two raters) [14].

It is worth noticing that the above-cited studies focused on the paradigm of 'pattern|disease' [45], in which few TCM patterns are studied within the context of a given disease at any stage of its natural history: irritable bowel syndrome [8], rheumatoid arthritis [9], frequent headaches [11], mixed sample of healthy participants and patients with chronic diseases [12], hypercholesterolemia [13], prediabetes [14], fertile and infertile women [15], and cardiovascular diseases [16]. This approach limits their external validity to a more general interpretation of the TCM practice. Therefore, since our study did not specify an underlying disease and patterns were randomly selected to compose the simulated sample of patients our

findings are considered as representative regarding both specific *zangfu* patterns and affected internal organs.

The observed low agreement for acupuncture prescription may be explained mainly by the likewise low agreement in pattern differentiation itself, along again with other unique aspects of acupuncture practice. There are plenty of criteria for selecting and combining acupoints: local, distant, specific, painful, *yin-yang* combinations, balance and imbalanced combinations are among the most commonly used [6,24]. In addition, each acupoint has several therapeutic indications that might be partially shared between either adjacent or distant acupoints, on the same channel or not [6,24]. Our results are also consonant with two above-cited studies that also investigated TCM agreement on prescription and reported slight interrater agreement on herbal therapy ($\kappa=0.16$; four raters) [8] and ‘little’ agreement on acupuncture prescription (κ values not available; three raters) [11]. Using data provided from Coeytaux et al. [11] – in which 37 patients with frequent cephalaea were interviewed by three TCM raters – and our methods described in this study, we found a slight agreement on acupuncture prescription considering either the acupoint’s label ($\tau=0.028$ [0.016; 0.042]) and the acupoint’s channel ($\tau=0.026$ [-0.001; 0.047]).

It is worth discussing our findings in light of two famous principles related to the personalized approach of TCM for diagnosis and treatment. On the one hand, the traditional statement ‘different treatments for the same pattern’ [4] acknowledges that it is not an error to treat the same pattern with a variety of acupoint sets. This statement holds because the implicit information herein is that the same pattern also manifests as different collections of signs and/or symptoms in each person due to personal and environmental factors [6]. Hence, it is possible that different treatments are selected for the same pattern depending on the patient’s manifestation profile and it should not be considered as a source of error. On the other hand, other traditional statement ‘the same treatment for different patterns’ [4]

acknowledges that to apply the same acupoints' set to treat different patterns is not an error. This statement also holds because of another implicit information: each acupoint presents different collections of therapeutic actions [6]. It is indeed possible that the same acupoints' set is indicated for different patterns depending on their expected therapeutic action and thus it should not be considered as a source of error for prescription. Because our sample of TCM raters was not aware about the true diagnosis of each patient and the automated 'simulation-identification' procedure guaranteed that each manifestation profile corresponded to a unique diagnosis, the observed slight agreements for pattern differentiation and acupoint prescription are not related to these statements.

The lower diagnostic accuracy of TCM raters regarding the simulated sample with the correct diagnosis identified by PDA was expected under both specific and general aspects of diagnosis; while raters relied on subjective data analysis, the PDA applied objective criteria for constructing a list of diagnostic hypothesis and selecting a diagnosis. Such a low accuracy of the TCM raters was accompanied by a low sensitivity and a high specificity, which indicates that raters are less capable of correctly identify the true pattern as present but are more capable of correctly identify the true pattern as absent. Therefore, the subjective criteria adopted by TCM raters to perform pattern differentiation seem to act mainly to exclude unlikely diagnostic hypotheses rather than to provide the true one – which in turn also helps explain the low interrater agreement for pattern differentiation. Because literature showed that the interrater agreement on pattern differentiation might be significantly improved after supervised practice [8,10,12], efforts should be made directed to test whether training might also improve the diagnostic accuracy of TCM raters.

An unanticipated, interesting finding was the diagnostic accuracy under the incorrect diagnosis (i.e. false profiles). It must be emphasized that this group comprised simulated patients with a unique diagnosis (known from the *SimTCM* output) that were misdiagnosed by

PDA using its quantitative criteria. PDA's misdiagnosis rate was low – nearly 6.0% – for these 73 *zangfu* patterns [5]; nonetheless, this result may be interpreted as the superiority of TCM raters to perform pattern differentiation in cases that PDA failed to report a correct diagnosis. This result is encouraging but raises a new challenge: to improve the detection of misdiagnosed patterns using PDA or other automated methods for pattern differentiation.

We found no personal (age, sex) or professional (clinic or scholar activity, time since postgraduate) variable to be associated with the diagnostic accuracy. Although our study was not designed to identify association between variables, these results may be used to plan sample sizes for larger studies aiming to determine if these factors are indeed determinants of an accurate pattern differentiation, if any.

Collectively, our findings raise a novel explanation for the low interrater agreement for pattern differentiation and even lower agreement for acupuncture prescription: the nonlinear, multivariate nature of patterns and acupoints. *Zangfu* patterns can be defined by at least four vectorial dimensions (e.g. *fei-qi* deficiency: vital substances, *qi*; internal organs, *fei*; pathophysiologic mechanisms, deficiency; and clinical manifestations) [6,24], whilst acupoints can be defined by at least three dimensions, two scalar plus a vectorial one (e.g. LI-4 *hegu*: channel number, 4; internal organ, large intestine-*dachang*; and therapeutic indications) [6,24]. If the sharing of clinical manifestations and therapeutic indications among vectorial variables of patterns and acupoints are also considered [5], the whole model of 'patient → TCM diagnosis → TCM prescription → treatment' is also a nonlinear one with two sequential stages. The input of clinical manifestations into the first stage of such a multivariate nonlinear model partially explains the variability in the outputted diagnosis because a variety of subsets of clinical manifestations may be present in the same pattern in different subjects. In sequence, inputting the diagnosis into the second stage of this multivariate nonlinear model for acupoint selection adds more variability to the outputted

acupoints. Adding some ‘noisy information’ in either or both stages – for instance, originated from the variability in rater’s own knowledge, clinical experience and training, ability to recognize manifestations or to differentiate between two similar conditions – turns into a scenario in which different raters might provide a variety of diagnostics and acupoints for the same subject. We thus encourage other researchers to investigate whether this model is of clinical value for improving TCM interrater agreement and possibly diagnostic accuracy as well.

There are limitations that need to be discussed for a proper interpretation of our results. Firstly, the presentation of cases in hardcopy questionnaires instead of real persons may be acknowledged as a potential source of variability in diagnosis; the simulation of cases being another interaction factor. However, the presentation of clinical cases is the core of TCM transmission since ancient times and therefore raters were familiarized with this type of presentation of clinical cases. In addition, the simulation of cases used manifestations as described in current literature and the usage of simulation procedures has been increasingly the method of choice for diagnostic analysis in TCM [18,19]. Although the simulation procedure did not cover all possible combinations of manifestations for a *zangfu* pattern, the random sampling method provided a variety of combinations that may occur in daily clinical practice. Most importantly, the usage of advanced statistical methods for data simulation (*SimTCM*) and analysis (κ , τ , bootstrapping, confusion matrices) is a major strength of our study in comparison to the previous ones and are strongly recommended for future studies on this subject.

5. Conclusions

Interrater agreement is slight for differentiating *zangfu* patterns and prescribing acupoints, regardless of whether the diagnosis is accurate or not. Interrater agreement is better for pattern

differentiation but yet slight for acupuncture prescription based on the selected internal organs. Diagnostic accuracy of TCM raters for differentiating *zangfu* patterns is not better than chance for most raters, although it is better for differentiating the affected internal organs, in particular when an automated method failed to provide the correct diagnosis.

Acknowledgements

We are in debit with Dr. Sohaku Raimundo César Bastos from the Instituto Sohaku-in Foundation for providing the conditions for data collection.

Funding

This study is supported by the grant (nº E-26/103.066/2012) from the Fundação Carlos Chagas Filho de Amparo à Pesquisa no Estado do Rio de Janeiro (FAPERJ).

References

- [1] Weingart SN, Wilson RM, Gibberd RW, Harrison B. Epidemiology of medical error. *British Medical Journal*, vol. 320, no. 18, pp. 774–777, 2000. [[PubMed](#)]
- [2] Kuhn GJ. Diagnostic errors. *Academic Emergency Medicine*, vol. 9, no. 7, pp. 740–750, 2002. [[PubMed](#)]
- [3] Schiff GD, Hasan O, Kim S, Abrams R, Cosby R, Lambert BL, Elstein AS, Hasler S, Kabongo ML, Krosnjar N, Odwazny R, Wisniewski MF, McNutt RA. Diagnostic errors in medicine: analysis of 583 physician-reported errors. *Archives of Internal Medicine*, vol. 169, no. 20, pp. 1881–1887, 2009. [[PubMed](#)]
- [4] Ferreira AS, Lopes AJ. Chinese Medicine Pattern Differentiation and Its Implications for Clinical Practice. *Chinese Journal of Integrative Medicine*, vol. 17, no. 11, pp.

- 546 818–823, 2011. [[PubMed](#)]
- 547 [5] Sá Ferreira A. Misdiagnosis and undiagnosis due to pattern similarity in Chinese
548 medicine: a stochastic simulation study using pattern differentiation algorithm. *Chinese*
549 *Medicine*, vol. 2011, no. 6, pp. 13, 2011. [[PubMed](#)]
- 550 [6] O'Connor J, Bensky D. Acupuncture a comprehensive text. Seattle: Eastland Press;
551 1987.
- 552 [7] Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, Roberts C,
553 Shoukri M, Streiner DL. Guidelines for Reporting Reliability and Agreement Studies
554 (GRRAS) were proposed. *Journal of Clinical Epidemiology*, vol. 64, no. 1, pp. 96–106,
555 2011. [[PubMed](#)]
- 556 [8] Sung JY, Leung WK, Ching JYL, Lao L, Zhang G, Wu JCY, Liang SM, Xie H, Ho
557 YP, Chan LS, Bernam B, Chan FKL. Agreements among traditional Chinese medicine
558 practitioners in the diagnosis and treatment of irritable bowel syndrome. *Alimentary*
559 *Pharmacology and Therapeutics*, vol. 20, no. 10, pp. 1205–1210, 2004. [[PubMed](#)]
- 560 [9] Zhang GG, Lee W, Bausell B, Lao L, Handwerger B, Berman B. Variability in the
561 traditional Chinese medicine (TCM) diagnoses and herbal prescriptions provided by
562 three TCM practitioners for 40 patients with rheumatoid arthritis. *The Journal of*
563 *Alternative and Complementary Medicine*, vol. 11, no. 3, pp. 415–421, 2005. [[PubMed](#)]
- 564 [10] Zhang GG, Singh B, Lee W, Handwerger B, Lao L, Berman B. Improvement of
565 agreement in TCM diagnosis among TCM practitioners for persons with the
566 conventional diagnosis of rheumatoid arthritis: effect of training. *The Journal of*
567 *Alternative and Complementary Medicine*, vol. 14, no. 14, pp. 381–386, 2008.
568 [[PubMed](#)]
- 569 [11] Coyetaux RR, Chen W, Lindemuth CE, Tan Y, Reilly AC. Variability in the diagnosis
570 and point selection for persons with frequent headache by traditional Chinese medicine

- 571 acupuncturists. *The Journal of Alternative and Complementary Medicine*, vol. 12, no. 9,
572 pp. 863–872, 2006. [[PubMed](#)]
- 573 [12] Mist S, Ritenbaugh C, Aickin M. Effects of questionnaire-based diagnosis and training
574 on inter-rater reliability among practitioners of traditional Chinese medicine. *The*
575 *Journal of Alternative and Complementary Medicine*, vol. 15, no. 7, pp. 703–709, 2009.
576 [[PubMed](#)]
- 577 [13] O’Brien KA, Abbas E, Zhang J, Guo Z, Luo R, Bensoussan A, Komesaroff PA. An
578 investigation into the reliability of Chinese medicine diagnosis according to the eight
579 guiding principles and zang-fu theory in Australians with hypercholesterolemia. *The*
580 *Journal of Alternative and Complementary Medicine*, vol. 15, n. 3, pp. 259–266, 2009.
581 [[PubMed](#)]
- 582 [14] Grant SJ, Schnyer RN, Chang DH, Fahey P, Bensoussan A. Interrater Reliability of
583 Chinese Medicine Diagnosis in People with Prediabetes. *Evidence-Based*
584 *Complementary and Alternative Medicine*, vol. 2013, Article ID 710892, 2013.
585 [[PubMed](#)]
- 586 [15] Birkeflet O, Laake P, Vollestad NK. Poor multi-rater reliability in TCM pattern
587 diagnoses and variation in the use of symptoms to obtain a diagnosis. *Acupuncture in*
588 *Medicine*, vol. 32, no. 4, pp. 332–352, 2014. [[PubMed](#)]
- 589 [16] Xu ZX, Xu J, Yan JJ, Wang YQ, Guo R, Liu GP, Yan HX, Qian P, Hong YJ. Analysis
590 of the diagnostic consistency of Chinese medicine specialists in cardiovascular disease
591 cases and syndrome identification based on the relevant feature for each label learning
592 method. *Chinese Journal of Integrative Medicine*, 2014 Jul 30. [Epub ahead of print]
593 [[PubMed](#)]
- 594 [17] de Sá Ferreira A. Statistical validation of strategies for Zang-fu single pattern
595 differentiation. *Journal of Chinese Integrative Medicine*, vol. 6, no. 11, pp. 1109–1116,

- 596 2008. [[PubMed](#)]
- 597 [18] Ferreira AS, Pacheco AG. SimTCM: A human patient simulator with application to
598 diagnostic accuracy studies of Chinese medicine. *Journal of Integrative Medicine*,
599 2014; December; Epub ahead of print. [[External link](#)]
- 600 [19] Ferreira AS. Promoting the Integrative Medicine by the computerization of traditional
601 Chinese medicine for scientific research and clinical practice: The SuiteTCM Project.
602 *Journal of Integrative Medicine*, vol. 11, no. 2, pp. 135–139, 2013. [[PubMed](#)]
- 603 [20] Ferreira AS. Diagnostic accuracy of pattern differentiation algorithm based on
604 traditional Chinese medicine theory: a stochastic simulation study. *Chinese Medicine*,
605 vol. 4, pp. 24, 2009. [[PubMed](#)]
- 606 [21] MacPherson H, Altman DG, Hammerschlag R, Youping L, Taixing W, White A, Moher
607 D, & STRICTA Revision Group. Revised STAndards for Reporting Interventions in
608 Clinical Trials of Acupuncture (STRICTA): extending the CONSORT statement. *PLoS*
609 *Medicine*, vol. 7, no. 6, pp. e1000261, 2010. [[PubMed](#)]
- 610 [22] MacPherson H, White A, Cummings M, Jobst K, Rose K, Niemtzow R. Standards for
611 reporting interventions in controlled trials of acupuncture: The STRICTA
612 recommendations. STAndards for Reporting Interventions in Controlled Trails of
613 Acupuncture. *Acupuncture in Medicine*, vol. 20, no. 1, pp. 22–25, 2002. [[PubMed](#)]
- 614 [23] Ferreira AS. Advances in Chinese medicine diagnosis: From traditional methods to
615 computational models. In: Kuang H. Recent advances in Chinese medicine. Croatia:
616 InTech. 2012. [[External link](#)]
- 617 [24] Maciocia G. Os Fundamentos da Medicina Tradicional Chinesa: Um Texto Abrangente
618 para Acupunturistas e Fitoterapeutas. Rio de Janeiro: Roca; 1996.
- 619 [25] World Health Organization. Standard acupuncture nomenclature. World Health
620 Organization, Geneva: 1993. [[External link](#)]

- 621 [26] Light RJ. Measures of response agreement for qualitative data: Some generalizations
622 and alternatives. *Psychology Bulletin*, vol. 76, no. 5, pp. 365–377, 1971. [[External link](#)]
- 623 [27] Hallgren KA. Computing Inter-Rater Reliability for Observational Data: An Overview
624 and Tutorial. *Tutorials in Quantitative Methods for Psychology*, vol. 8, no. 1, pp.
625 23–34, 2012. [[PubMed](#)]
- 626 [28] Janson H, Olsson U. A measure of agreement for interval or nominal multivariate
627 observations. *Educational and Psychological Measurement*, vol. 6, no. 21, pp. 277–289,
628 2001. [[External link](#)]
- 629 [29] Efron B, Tibshirani RJ. An introduction to the bootstrap. Florida, CRC Press, 1998.
- 630 [30] North BV, Curtis D, Sham PC. A note on the calculation of empirical P-values from
631 Monte Carlo procedures. *American Journal of Human Genetics*, vol. 71, no. 2, pp.
632 439–441, 2002. [[PubMed](#)]
- 633 [31] Landis JR, Koch GG. The measurement of observer agreement for categorical data.
634 *Biometrics*, vol. 33, no. 1, pp. 159–174, 1977. [[External link](#)]
- 635 [32] Altman DG, Bland JM. Diagnostic tests. 1: Sensitivity and specificity. *British Medical*
636 *Journal*, vol. 308, no. 6943, pp. 1552, 1994. [[PubMed](#)]
- 637 [33] Altman DG, Bland JM. Diagnostic tests 2: Predictive values. *British Medical Journal*,
638 vol. 309, no. 6947, pp. 102, 1994. [[PubMed](#)]
- 639 [34] Kanji G. 100 statistical tests. 3a edição. Sage Publications, London, 2006.
- 640 [35] World Health Organization. Health research methodology: A guide for training in
641 research methods. World Health Organization, Geneva: 2001. [[External Link](#)]
- 642 [36] Sim J, Wright CC. The kappa statistics in reliability studies: Use, interpretation, and
643 sample size requirements. *Physical Therapy*, vol. 85, no. 3, pp. 257-268, 2005.
644 [[PubMed](#)]
- 645 [37] Altaye M, Donner A, Eliasziw M. A general goodness-of-fit approach for inference

procedures concerning the kappa statistic. *Statistics in Medicine*, vol. 20, no. 16, pp. 2479–2488, 2011. [[PubMed](#)]

[38] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. [[External link](#)]

[39] Canty A, Ripley B. boot: Bootstrap R (S-Plus) Functions. 2014. R package version 1.3-11. [[External link](#)]

[40] Kuhn M. Contributions from Jed Wing, Steve Weston, Andre Williams, Chris Keefer, Allan Engelhardt, Tony Cooper, Zachary Mayer, Brenton Kenkel, the R Core Team and Michael Benesty. caret: Classification and Regression Training. 2014. R package version 6.0-37. [[External link](#)]

[41] Gamer M, Lemon J, Singh IFP. irr: Various Coefficients of Interrater Reliability and Agreement. 2012. R package version 0.84. [[External link](#)]

[42] Rotondi MA. kappaSize: Sample Size Estimation Functions for Studies of Interobserver Agreement. 2013. R package version 1.1. <http://CRAN.R-project.org/package=kappaSize>.

[43] Falissard B. psy: Various procedures used in psychometry. 2012. R package version 1.1. [[External link](#)]

[44] Dragulescu AA. xlsx: Read, write, format Excel 2007 and Excel 97/2000/XP/2003 files. 2014. R package version 0.5.7. [[External link](#)]

[45] Shin BC, Kim S, Cho YH. Syndrome pattern and its application in parallel randomized controlled trials. *Chinese Journal of Integrative Medicine*, vol. 19, no. 3, pp. 163–171, 2013. [[Pubmed](#)]

Table 1: Descriptive data of the studied sample.

Variable	Value*
Sample size, n	6
Male	4 (66.7%)
Female	2 (33.3%)
Professional activity	
Clinical consultant	6 (100%)
Postgraduate professor	5 (83.3%)
Supervisor of clinic-school	3 (50.0%)
Age, years	43 [37; 64]
Formal training and practicing	
Duration of postgraduate course, years	2 [2; 2.5]
Acupuncture and TCM theory, hours	800
Acupuncture training, hours	400
Time since postgraduate, years	12 [4; 33]

*Median [minimum; maximum] for continuous variables; frequency (%) for categorical variables.

Table 2: Diagnostic performance of six raters for pattern differentiation grouped by diagnostic outcome.

Outcome Descriptor	Correct diagnosis (true profiles)						Incorrect diagnosis (false profiles)					
	ACC [95%CI]	P-value	SEN	SPE	+PV	−PV	ACC [95%CI]	P-value	SEN	SPE	+PV	−PV
Patterns												
Rater 1	61.7 [48.2; 73.9]	0.046	23.3	100	100	56.6	58.3 [44.9; 70.9]	0.123	16.7	100	100	54.5
Rater 2	58.3 [44.9; 70.9]	0.123	16.7	100	100	55.6	56.7 [43.2; 69.4]	0.183	13.3	100	100	53.6
Rater 3	60.0 [46.5; 72.4]	0.078	20.0	100	100	55.6	58.3 [44.9; 70.9]	0.123	16.7	100	100	54.5
Rater 4	56.7 [43.2; 69.4]	0.183	13.3	100	100	53.6	51.7 [38.4; 64.8]	0.449	3.3	100	100	50.8
Rater 5	70.0 [56.8; 81.2]	0.001	40.0	100	100	62.5	63.3 [49.9; 75.4]	0.026	26.7	100	100	57.7
Rater 6	61.7 [48.2; 73.9]	0.046	23.3	100	100	56.6	61.7 [48.2; 73.9]	0.046	23.3	100	100	56.6
Group-median	60.9 [56.7; 70.0]	NT	21.7	100	100	56.1	58.3 [51.7; 63.3]	NT	16.7	100	100	54.5
Internal organs												
Rater 1	65.0 [51.6; 76.9]	0.014	33.3	96.7	90.9	59.2	65.0 [51.6; 76.9]	0.014	33.3	96.7	90.9	59.2
Rater 2	61.7 [48.2; 73.9]	0.046	23.3	100	100	56.6	63.3 [49.9; 75.4]	0.026	26.7	100	100	57.7
Rater 3	73.3 [60.3; 83.9]	<0.001	56.7	90.0	85.0	67.5	70.0 [56.8; 81.2]	0.001	50.0	90.0	83.3	64.3
Rater 4	55.0 [41.6; 67.9]	0.259	23.3	86.7	63.6	53.1	58.3 [44.9; 70.9]	0.123	30.0	86.7	69.2	55.3
Rater 5	68.3 [55.0; 79.7]	0.003	46.7	90.0	82.4	62.8	75.0 [61.2; 85.3]	<0.001	60.0	90.0	85.7	69.2
Rater 6	68.3 [55.0; 79.7]	0.003	40.0	96.7	92.3	61.7	70.0 [56.8; 81.2]	0.001	43.3	96.7	92.9	63.0
Group-median	66.7 [55.0; 73.3]	NT	36.7	93.4	88.0	60.5	67.5 [58.3; 75.0]	NT	38.3	93.4	88.3	61.1

ACC: accuracy; 95%CI: 95% confidence interval; SEN: sensitivity; SPE: specificity; +PV: positive predictive value; −PV: negative predictive

value; NT: not tested.

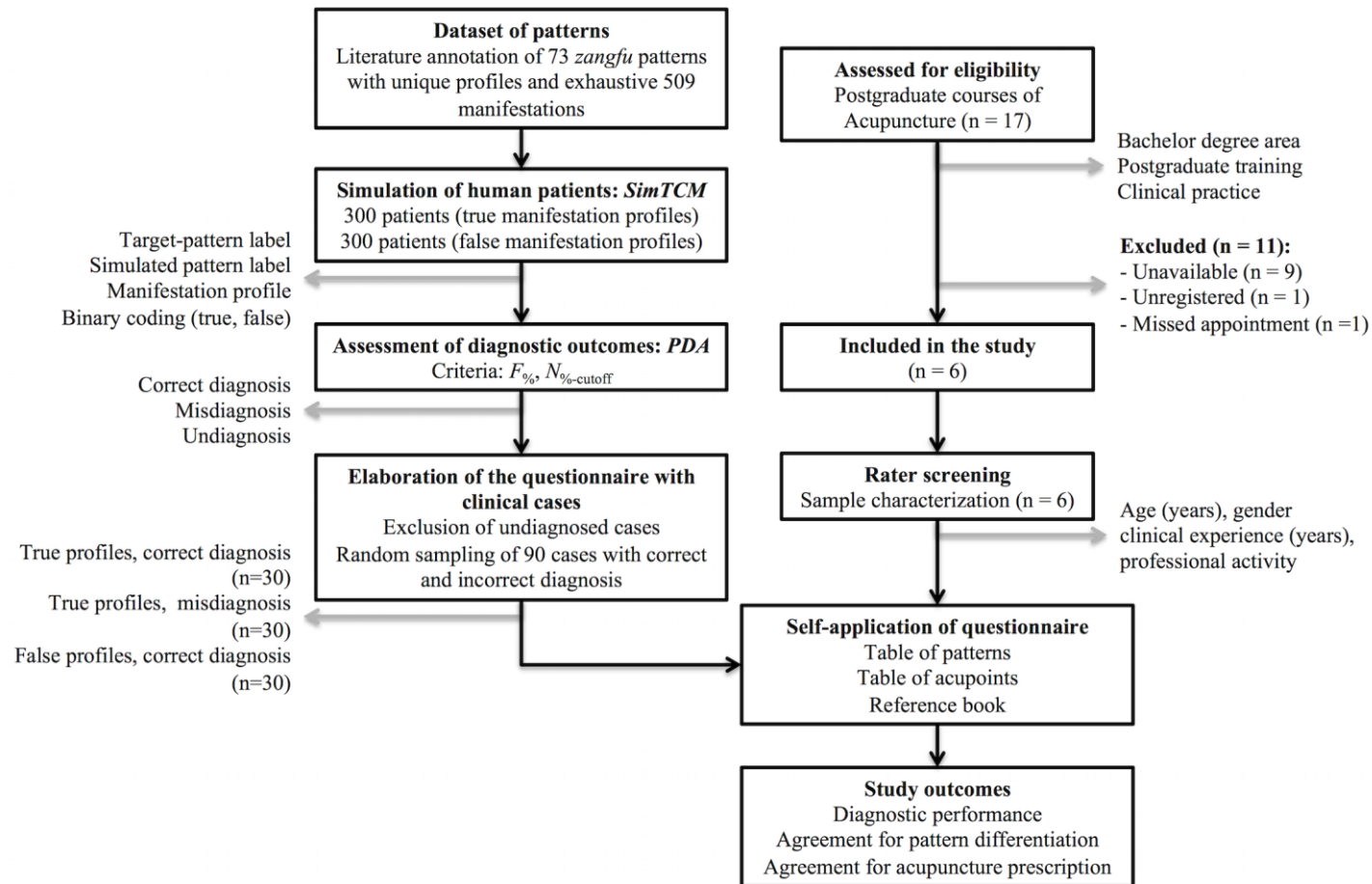


Figure 1: Study flowchart.

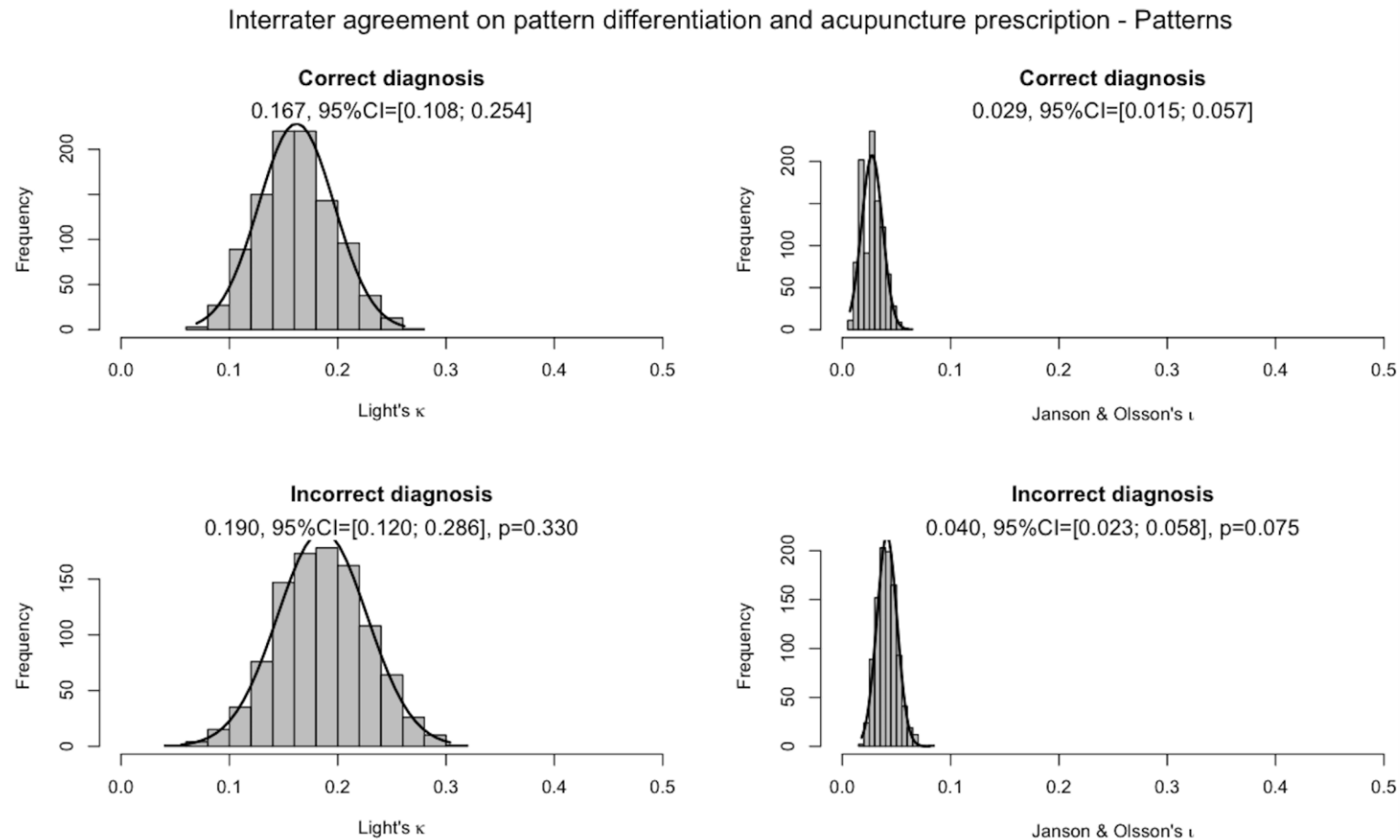


Figure 2: Bootstrap analysis of interrater agreement on pattern differentiation (left column) and acupuncture prescription (right column) estimated from the specific labels of 73 *zangfu* patterns of simulated human patients grouped by correct (upper row, n=30) or incorrect (lower row, n=30) diagnosis.

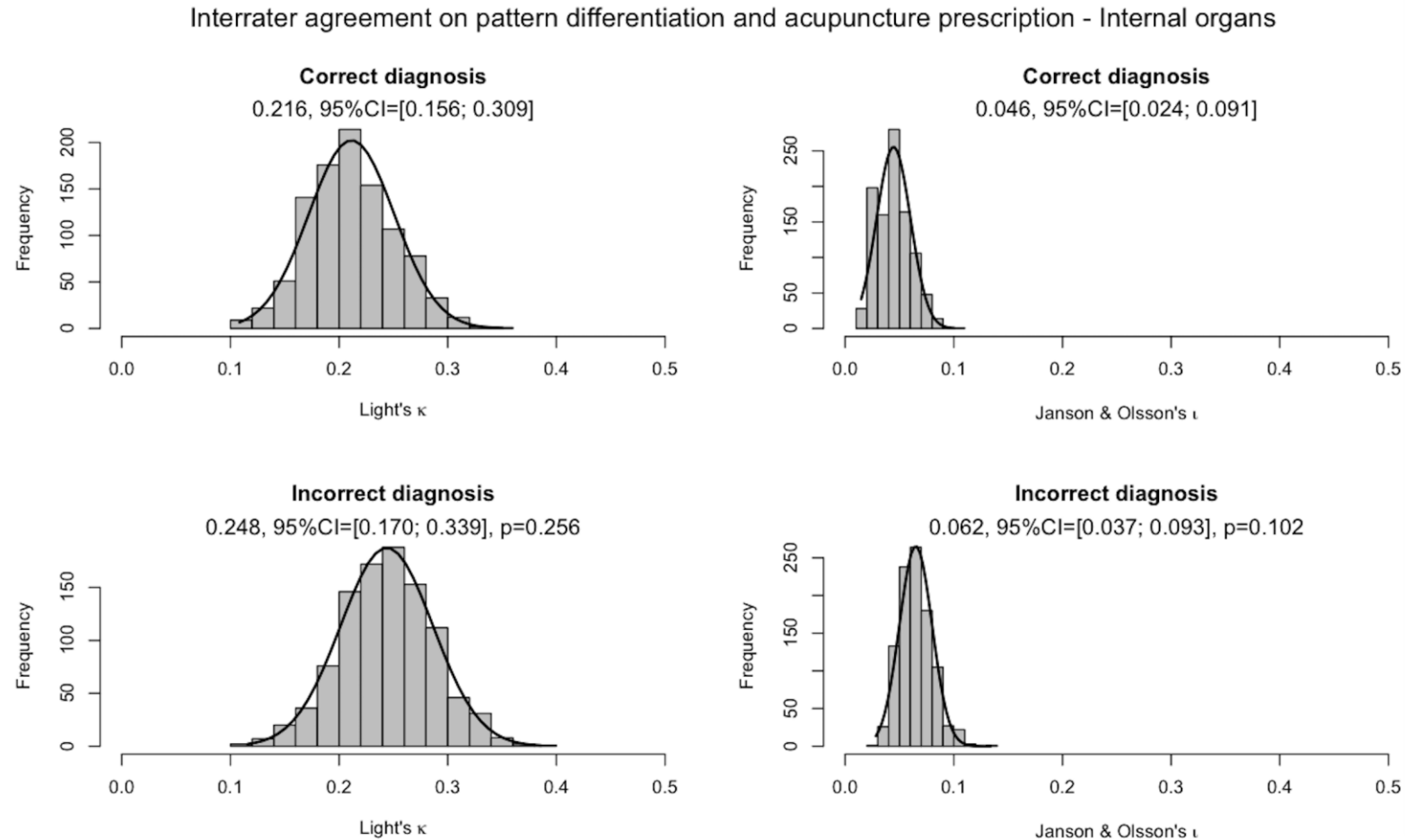


Figure 3: Bootstrap analysis of interrater agreement on pattern differentiation (left column) and acupuncture prescription (right column) estimated from the affected *zangfu* systems of simulated human patients grouped by correct (upper row, $n=30$) or incorrect (lower row, $n=30$) diagnosis.

Capítulo 4 – Considerações finais

Registra-se que não se tem informação de outro estudo na simulação computacional em diagnóstico de padrão em MTC e devido ao fato, houve algumas dificuldades quanto à execução desde a busca bibliográfica, pois existe uma escassez sobre o tema em questão, até a entrega dos questionários, pois muitas foram as tentativas em vão em se conseguir docentes de pós-graduação e com disponibilidade em resolver os casos simulados, uma vez que são 90 questões de raciocínio para o diagnóstico e prescrição dos acupontos e, a média de tempo percorreu cerca de uma hora e meia.

A pesquisa mostrou que existe uma diferença entre avaliar um paciente por meio do uso de padrão de *zangfu* e avaliar baseado apenas na identificação dos órgãos internos, o que pareceu este, ser melhor identificado pelos profissionais na prática clínica. A acurácia diagnóstica dos examinadores em MTC para diferenciação dos padrões *zangfu*, não foi a melhor possibilidade para a maioria, embora fosse melhor para diferenciar os órgãos internos afetados, em especial, quando um método automatizado não conseguiu apresentar o diagnóstico correto.

Devido ao fato de a maioria dos entrevistados serem professores de pós-graduação e tempo de treinamento entre 4 e 33 anos, além de supervisão de estágio na prática da acupuntura, observou-se uma acurácia acima dos 60% nos casos de diagnóstico por padrão e identificação dos órgãos internos. Aventa-se a possibilidade de que outros profissionais com menos tempo de formação e experiência clínica possam apresentar uma acurácia, sensibilidade e especificidade bem diferentes da realidade dos examinadores deste estudo, o que sugere-se que até mesmo com profissionais experientes, uma ferramenta de diagnóstico pode ajudar a aumentar a acurácia na prática clínica e prescrição dos acupontos.

Existem ainda outras perspectivas na visão da pesquisa a serem estudadas futuramente, como estudar outros perfis de profissionais, recém-formados e alunos em formação da MTC e, se no caso da simulação do diagnóstico incorreto, quais seriam os efeitos que uma prescrição incorreta dos acupontos poderia ocasionar no que diz respeito a funcionalidade da prescrição, se os pontos mesmo incorretamente prescritos funcionam ou não, ou se ainda, são eficientes porém não eficazes na prática clínica.

O estudo da simulação no ramo do diagnóstico chinês e prescrição de acupontos assegura uma potencial fonte de treinamento e aperfeiçoamento para os praticantes de MTC, contribuindo para o desenvolvimento das capacidades necessárias ao domínio da competência

nas áreas de saúde, e na sistematização da assistência, garantindo a qualidade do tratamento, melhorando a acurácia diagnóstica, bem como a relação custo/benefício no momento da avaliação, minimizando possíveis riscos e intercorrências.

Devido ao aumento do acesso gratuito da população aos serviços derivados das medicinas orientais, em especial na fisioterapia, tanto no âmbito preventivo quanto da reabilitação, recomenda-se o uso da simulação para capacitar os especialistas em acupuntura, agindo como coadjuvante na melhora global dos pacientes atendidos, nas diversas disfunções, com baixo risco de eventos colaterais.

REFERÊNCIAS

- ABE GC, RAMOS PE, FONTES SV, YONAMINE B, MELLO CAS, QUADROS AJ, CHUN WP, JIA JE, SCHMIDT B, OLIVEIRA ASB. Poliomielite e síndrome pós-polio pela medicina tradicional chinesa: da fisiopatologia ao diagnóstico. **Rev Neurocienc.** 2011;19(2):365-81.
- ALTAYE M, DONNER A, ELIASZIW M. A general goodness-of-fit approach for inference procedures concerning the kappa statistic. **Stat Med.** 2011; 20(16): 2479-88.
- ALTMAN DG, BLAND JM. Diagnostic tests. 1: Sensitivity and specificity. **BMJ.** 1994;308(6943):1552.
- ALTMAN DG, BLAND JM. Diagnostic tests 2: Predictive values. **BMJ.** 1994;309(6947):102.
- BIRKEFLET O, LAAKE P, VØLLESTAD NK. Poor multi-rater reliability in TCM pattern diagnoses and variation in the use of symptoms to obtain a diagnosis. **Acupunct Med.** 2014;32(4):325-32.
- BRASIL, Ministério da Saúde. Secretaria de Atenção à Saúde. Departamento de Atenção Básica. Política Nacional de Práticas Integrativas e Complementares no SUS - PNPIC-SUS. Brasília: Ministério da Saúde, 2006. 92 p.
- BRASIL. Conselho Nacional de Saúde. Resolução nº 466, 2012. Diretrizes e Normas regulamentadoras de pesquisa envolvendo seres humanos. Brasília, 13 jun. 2013.
- CANTY A, RIPLEY B. boot: Bootstrap R (S-Plus) Functions. 2014. R package version 1.3-11.
- CAO H, BOURCHIER S, LIU J. Does Syndrome Differentiation Matter? A Meta-Analysis of Randomized Controlled Trials in Cochrane Reviews of Acupuncture. **Med Acupunct.** 2012;24(2):68-76.
- CHENG XN. **Chinese Acupuncture and Moxibustion.** Beijing: Foreign Language Press, 1987.
- CINTRA MER, PEREIRA PPG. Percepções de Corpo Identificadas entre Pacientes e Profissionais de Medicina Tradicional Chinesa do Centro de Saúde Escola do Butantã. **Saúde Soc.** 2012; 21(1):193-205.
- COEYTAUX RR, CHEN W, LINDEMUTH CE, TAN Y, REILLY AC. Variability in the diagnosis and point selection for persons with frequent headache by traditional Chinese medicine acupuncturists. **J Altern Complement Med.**;12(9):863-72.
- COHEN J. A coefficient of agreement for nominal scales. **Educational and Psychological Measurement**, vol20, p.37–46, 1960.

- COHEN J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. **Psychological Bulletin**, vol.70, p.213–220, 1968.
- DAVISON AC, HINKLEY DV. (1997) *Bootstrap Methods and Their Applications*. Cambridge University Press, Cambridge. ISBN 0-521-57391-2
- DE SÁ FERREIRA A. Evidence-based practice of Chinese medicine in physical rehabilitation science. **Chin J Integr Med**. 2013; 19(10):723-9.
- DE SÁ FERREIRA A. Promoting integrative medicine by computerization of traditional Chinese medicine for scientific research and clinical practice: The SuiteTCM Project. **J Integr Med**. 2013; 11(2):135-9.
- DEVEREAUX PJ, BHANDARI M, CLARKE M, MONTORI VM, COOK DJ, et al. (2005) Need for expertise based randomised controlled trials. **BMJ** 330:88.
- DOYLE DJ, ARELLANO R. The virtual Anesthesiology training simulation system. **Can J Anaesth**, 1995;42:267-273.
- DRAGULESCU AA. xlsx: Read, write, format Excel 2007 and Excel 97/2000/XP/2003 files. R package version 0.5.7. 2014. <http://CRAN.R-project.org/package=xlsx>
- EFRON B, TIBSHIRANI RJ. *An introduction to the bootstrap*. Florida, CRC Press, 1998.
- FALISSARD B. psy: Various procedures used in psychometry. 2012. R package version 1.1 <http://CRAN.R-project.org/package=psy>
- FERREIRA A. Statistical validation of strategies for Zang-fu single pattern differentiation. **Zhong Xi Yi Jie He Xue Bao**. 2008 Nov;6(11):1109-16.
- FERREIRA AS, LOPES AJ. Chinese medicine pattern differentiation and its implications for clinical practice. **Chin J Integr Med**. 2011;17(11):818-23.
- FERREIRA AS, PACHECO AGF. SimTCM: A human patient simulation with application to Chinese medicine diagnostic accuracy studies. **J Integr Med** 2014; In Press.
- FERREIRA AS. Diagnostic accuracy of pattern differentiation algorithm based on Chinese medicine theory: a stochastic simulation study. **Chin Med**. 2009; 4:24.
- FIREMAN Z, SEGAL A, KOPELMAN Y, STERNBERG A, CARASSO R. Acupuncture treatment for irritable bowel syndrome. A double-blind controlled study. **Digestion**. 2001;64(2):100-3.
- FLAWS B. **The Classic of Difficulties**. 4 edition. Boulder: Blue Poppy Press; 2004.
- FLEISS JL. Measuring nominal scale agreement among many raters. **Psychological Bulletin**, vol76, p.378–382, 1971.

- FRAGOSO AP, FERREIRA AD. Statistical distribution of acupoint prescriptions for sensory-motor impairments in post-stroke subjects. **Chin J Integr Med.** 2012; Dec 3. [Epub ahead of print]
- FRAGOSO APS, FERREIRA AS. A Fisiopatologia do acidente vascular cerebral e suas sequelas segundo a Medicina Chinesa. **Corpus et Scientia.** 2012; 8(3):167-75.
- FRAGOSO APS. Ensaio clínico com acupuntura: Questões bioéticas e metodológicas. **Perspectiva Online.** 2010; 4(15):151-8.
- FROOM P, BENBASSAT J. Inconsistencies in the classification of preventive interventions. **Prev Med.** 2000; 31(2 Pt 1):153-8.
- GRANT SJ, SCHNYER RN, CHANG DH, FAHEY P, BENSOUSSAN A. Interrater Reliability of Chinese Medicine Diagnosis in People with Prediabetes. **Evid-Based Complem Altern Med.** 2013; 2013: Article ID 710892, 8 pages.
- GUANG JY. The mode of thinking in Chinese clinical medicine: characteristics, steps and forms. **Clin Acupunct Oriental Med.** 2001; 2(1):23-8.
- GUYATT GH, TUGWELL PX, FEENY DH, HAYNES RB, DRUMMOND M. A framework for clinical evaluation of diagnostic technologies. **CMAJ.** 1986;134(6):587-94.
- HALLGREN KA. Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. **Tutor Quant Methods Psychol.** 2012; 8(1): 23-34.
- HSIEH J, HONDA AF, SUÁREZ-FARINAS M, SANSON M, KEDHAR S, MAURO J, FRANCIS J, BADAMO J, DIAZ VA, KEMPEN JH, LATKANY PA. Fundus image diagnostic agreement among multiple raters in uveitis utilizing free and open source software. **Can J Ophthalmol.** 2013; 48(4):227-234.
- HUANG MJ, CHEN MY. Integrated design of the intelligent web based Chinese Medical Diagnostic System (CMDS) systematic development for digestive health. **Expert Syst Appl.** 2007;32(2):658-73.
- IOM (Institute of Medicine). Clinical Practice Guidelines We Can Trust. Washington, DC: The National Academies Press, 2011.
- JANSON H, OLSSON U. A measure of agreement for interval or nominal multivariate observations. **Educ Psychol Meas.** 2001; 6(21):277-289.
- KANJI G. 100 statistical tests. 3a edição. Sage Publications, London, 2006.
- KOTTNER AJ, AUDIGÉ L, BRORSON S, DONNER GAJEWSKIE BJ, HROBJARTSSON A, ROBERTS C, SHOUKRIH M, STREINER DL. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. **J Clin Epidemiol.** 2011; 64(1):96-106.

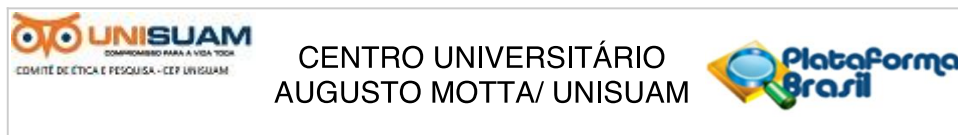
- KRAEMER, H.C. Extension of the kappa coefficient. **Biometrics**, vol.36:207–216, 1980.
- KUHN M. Contributions from Jed Wing, Steve Weston, Andre Williams, Chris Keefer, Allan Engelhardt, Tony Cooper, Zachary Mayer, Brenton Kenkel, the R Core Team and Michael Benesty. (2014). caret: Classification and Regression Training. R package version 6.0-37. <http://CRAN.R-project.org/package=caret>
- LANDIS JR, KOCH GG. The measurement of observer agreement for categorical data. **Biometrics**. 1977; 33(1):159-74.
- LEE T, CH'ENG CF, CHANG CS. Some early records of nervous and mental diseases in traditional Chinese medicine. **Chin Med J**. 1962; 81:55-9.
- LIGHT RJ. Measures of response agreement for qualitative data: Some generalizations and alternatives. **Psychol Bull**. 1971; 76(5):365-77.
- LIU GP, LI GZ, WANG YL, WANG YQ. Modelling of inquiry diagnosis for coronary heart disease in traditional Chinese medicine by using multi-label learning. **BMC Complement Altern Med**. 2010; 10:37.
- LU AP, CHEN KJ. Integrative medicine in clinical practice: From pattern differentiation in traditional Chinese medicine to disease treatment. **Chin J Integr Med**. 2009; 15(2):152.
- LUIZ AB, CORDOVIL I, BARBOSA FILHO J, FERREIRA AS. Zangfu zheng (patterns) are associated with clinical manifestations of zang shang (target-organ damage) in arterial hypertension. **Chin Med.**, 2011; 6:23.
- MACIOCIA G. **Os Fundamentos da Medicina Tradicional Chinesa: Um Texto Abrangente para Acupunturistas e Fitoterapeutas**. Rio de Janeiro: Roca; 1996.
- MACLURE M, WILLETT WC. Misinterpretation and misuse of the kappa statistic. **Am J Epidemiol**, 1987; 126:161–169.
- MACPHERSON H, WHITE A, CUMMINGS M, JOBST K, ROSE K, NIEMTZOW R; STandards for Reporting Interventions in Controlled Trails of Acupuncture. Standards for reporting interventions in controlled trials of acupuncture: The STRICTA recommendations. STandards for Reporting Interventions in Controlled Trails of Acupuncture. **Acupunct Med**. 2002;20(1):22-5.
- MACPHERSON H, ALTMAN DG, HAMMERSCHLAG R, YOUPING L, TAIXIANG W, WHITE A, MOHER D, & STRICTA Revision Group. Revised STandards for Reporting Interventions in Clinical Trials of Acupuncture (STRICTA): extending the CONSORT statement. **PLoS Medicine**. 2010; 7(6):e1000261.

- MIST S, RITENBAUGH C, AICKIN M. Effects of questionnaire-based diagnosis and training on inter-rater reliability among practitioners of traditional Chinese medicine. **J Altern Complement Med.** 2009; 15(7):703-709.
- NORTH BV, CURTIS D, SHAM PC. A note on the calculation of empirical P-values from Monte Carlo procedures. **Am. J. Hum. Genet.** 2002; 71(2); 439-441.
- O'BRIEN KA, ABBAS E, ZHANG J, GUO Z, LUO R, BENSOUSSAN A, KOMESAROFF PA. An investigation into the reliability of Chinese medicine diagnosis according to the eight guiding principles and *zang-fu* theory in Australians with hypercholesterolemia. **J Altern Complement Med.** 2009, 15(3):259-266.
- O'CONNOR J, BENSKY D. **Acupuncture a comprehensive text.** Seattle: Eastland Press; 1987.
- PRADY SL, MACPHERSON H (2007) Assessing the utility of the standards for reporting trials of acupuncture (STRICTA): a survey of authors. **J Altern Complement Med** 13:939-43.
- QUEIROZ MS. O sentido do conceito de medicina alternativa e movimento vitalista: uma perspectiva teórica introdutória. In: NASCIMENTO, M.C. (Org.). **As duas faces da montanha: estudos sobre medicina chinesa e acupuntura.** São Paulo: Hucitec, 2006. p. 19-39.
- R CORE TEAM (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: <http://www.R-project.org/>.
- RASHID A, JAKOBIEC FA. Avoiding the major complication of ophthalmic pathology: misdiagnosis. A review of three common diagnostic challenges. **Semin Ophthalmol.** 2014 Nov;29(5-6):468-74.
- SÁ FERREIRA A. Advances in Chinese Medicine Diagnosis: From Traditional Methods to Computational Models. In: **Recent Advances in Theories and Practice of Chinese Medicine.** Prof. Haixue Kuang (Ed.), 2012.
- SÁ FERREIRA A. Misdiagnosis and undiagnosis due to pattern similarity in Chinese medicine: a stochastic simulation study using pattern differentiation algorithm. **Chin Med.** 2011; 6:13.
- SCHAECHTER J, NEUSTEIN SM. P6 acupuncture point stimulation for prevention of postoperative nausea and vomiting. **Anesthesiology** 2008; 109:155-6;
- SHÙ-DÉ, J. **Case studies on pattern identification from the personal experience of Jião Shù-Dé.** Taos, New Mexico: 2006.

- SIEMIATYCKI J, CAMPBELL S. Nonresponse bias and early versus all responders in mail and telephone surveys. **Am J Epidemiol** 1984; 120(2): 291-301.
- SIM, J., WRIGHT, C.C. The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requirements. **Phys Ther.** 2005; 85(3):257-68.
- SUNG JJY, LEUNG WK, CHING JYL, LAO L, ZHANG G, WU JCY, LIANG SM, XIE H, HO YP, CHAN LS, BERMAN B, CHAN FKL. Agreements among traditional Chinese medicine practitioners in the diagnosis and treatment of irritable bowel syndrome. **Aliment Pharmacol Ther** 2004; 20:1205–1210.
- T'AO L. Chinese medicine during the Chin (1127-1234) and Yuan (1234-1368) eras. **Chin Med J.** 1955; 73(3):241-56.
- THOMPSON WD, WALTER SD. A reappraisal of the kappa coefficient. **J Clin Epidemiol** 1988; 41(10): 949-58.
- THOMPSON WG, LONGSTRETH GF, DROSSMAN DA, HEATON KW, IRVINE EJ, MULLER-LISSNER SA. Functional bowel disorders and functional abdominal pain. **Gut** 1999; 45 (Suppl. 2): II43–7.
- WANG LY, ZHONG CJ, CHEN W. Study of thinking ability in clinical practice enhanced by adopting PBL of logic to medical students. **Zhong Guo Gao Deng Yi Xue JiaoYu.** 2010; 1:118-20.
- WANG X, QU H, LIU P, CHENG Y. A self-learning expert system for diagnosis in traditional Chinese medicine. **Expert Syst Appl.** 2004; 26(4):557-66.
- WORLD HEALTH ORGANIZATION. **Acupuncture: a review and analysis of report on controlled clinical trials.** World Health Organization, Geneva: 2003.
- WORLD HEALTH ORGANIZATION. **Estrategia de la OMS sobre medicina tradicional 2014-2023.** Geneva; 2013.
- WORLD HEALTH ORGANIZATION. **General Guidelines for Methodologies on Research and Evaluation of Traditional Medicine.** Geneva; 2000
- WORLD HEALTH ORGANIZATION. **Health research methodology: A guide for training in research methods.** World Health Organization, Geneva: 2001.
- WORLD HEALTH ORGANIZATION. **Primary Health Care Report of the International Conference on Primary Health Care,** Alma-Ata, USSR, 6–12 September 1978. Geneva, 1978.
- WORLD HEALTH ORGANIZATION. **Standard acupuncture nomenclature.** World Health Organization, Geneva: 1993.

- XUE Z, XU J, YAN J, WANG Y, GUO R, LIU G, YAN H, QIAN P, HONG Y. Analysis of the Diagnostic Consistency of Chinese Medicine Specialists in Cardiovascular Disease Cases and Syndrome Identification Based on the Relevant Feature for Each Label Learning Method. **Chin J Integr Med.** 2014; Epub ahead of print.
- YAMAMURA Y. **Acupuntura tradicional: a arte de inserir.** 2. ed. São Paulo: Roca, 2001.
- YAMAMURA Y. **Entendendo medicina chinesa e acupuntura.** São Paulo: Centre AO, 2006.
- YANG SZ, CHACE C. Translators. **The systematic classic of acupuncture and moxibustion.** Boulder: Blue Poppy Press; 2004.
- YANG SZ, LI JY. Translators. **Pi Wei Lun (Treatise on the Spleen & Stomach)** Boulder: Blue Poppy; 1993.
- YANG W, MENG F, JIANG Y. **Diagnostics of Traditional Chinese Medicine.** Academy Press: Beijing, 1998.
- YEUNG WF, CHUNG KF, POON MM, HO FY, ZHANG SP, ZHANG ZJ, ZIEA ET, WONG TAAM V. Prescription of chinese herbal medicine and selection of acupoints in pattern-based traditional chinese medicine treatment for insomnia: a systematic review. **Evid Based Complement Alternat Med.** 2012; 2012:902578.
- ZASLAWSKI C. Clinical reasoning in traditional Chinese medicine: implications for clinical research. **Clin Acupunct Oriental Med.** 2003; 4(2-3):94-101.
- ZHANG GG, LEE W, BAUSELL B, LAO L, HANDWERGER B, BERMAN B. Variability in the traditional Chinese medicine (TCM) diagnoses and herbal prescriptions provided by three TCM practitioners for 40 patients with rheumatoid arthritis. **J Altern Complement Med.** 2005; 11(3):415-421.
- ZHANG GG, SINGH B, LEE W, HANDWERGER B, LAO L, BERMAN B. Improvement of agreement in TCM diagnosis among TCM practitioners for persons with the conventional diagnosis of rheumatoid arthritis: effect of training. **J Altern Complement Med.** 2008; 14(4):381-386.
- ZHENG N, WU Z. TCM-SIRD: an integrated aided system for traditional Chinese medicine *Sizheng*. **Proceedings of 2004 IEEE International Conference on Systems, Man and Cybernetics**, pp. 3864-3868, The Hague, The Netherlands, October 10-13, 2004.
- ZHOU J, ZHU J, CHEN M, JIANG M, ZHANG Z, ZHAN Z, ZHANG X. Logical thinking in pattern differentiation of traditional Chinese medicine. **J Tradit Chin Med.** 2013; 33(1):137-40.

Anexo 1 – Carta de aprovação do comitê de ética em pesquisa



PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: Efeitos do erro diagnóstico na confiabilidade interexaminadores para diferenciação de padrões e prescrição de acupontos

Pesquisador: Ingrid Jardim de Azeredo Souza

Área Temática:

Versão: 2

CAAE: 35723214.6.0000.5235

Instituição Proponente: SOCIEDADE UNIFICADA DE ENSINO AUGUSTO MOTTA

Patrocinador Principal: FUN CARLOS CHAGAS F. DE AMPARO A PESQUISA DO ESTADO DO RIO DE JANEIRO - FAPERJ
Financiamento Próprio

DADOS DO PARECER

Número do Parecer: 817.427

Data da Relatoria: 08/10/2014

Apresentação do Projeto:

O projeto versa sobre questões da medicina tradicional chinesa, que tem sido cada vez mais empregada no Brasil e objeto de pesquisa científica em todo o mundo. Em especial, o projeto traz objetivos relacionados à acupuntura, buscando quantificar a confiabilidade no diagnóstico e prescrição da técnica. Está bem apresentado no ambiente da Plataforma Brasil, cumprindo os requisitos básicos do sistema CEP-CONEP.

Objetivo da Pesquisa:

Investigar os efeitos do erro diagnóstico na confiabilidade interexaminadores para diferenciação de padrões e prescrição de acupontos.

Avaliação dos Riscos e Benefícios:

Os riscos e benefícios foram apresentados no projeto e no termo de consentimento.

Comentários e Considerações sobre a Pesquisa:

A folha de rosto está devidamente preenchida; há embasamento científico que justifique a pesquisa; os objetivos estão bem definidos; há explicação dos métodos que serão empregados; o cálculo do tamanho da mostra está presente; há critérios de inclusão e exclusão bem definidos; a avaliação crítica de riscos e benefícios foi feita; há explicitação da responsabilidade do

Endereço: Praça das Nações nº 34 TEL: (21)3882-9797 (Ramal : 1015)

Bairro: Bonsucesso

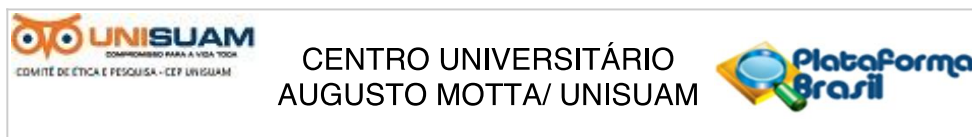
CEP: 21.041-010

UF: RJ

Município: RIO DE JANEIRO

Telefone: (21)3882-9797

E-mail: comitedeetica@unisuum.edu.br



Continuação do Parecer: 817.427

pesquisador; os critérios para suspender a pesquisa estão definidos; há apresentação de orçamento financeiro; o compromisso de divulgação foi descrito; garantia de acesso ao pesquisador foi dada. As pendências apontadas no parecer anterior sobre forma de recrutamento, local de preenchimento do questionário e esclarecimentos sobre ressarcimento foram todas cumpridas.

Considerações sobre os Termos de apresentação obrigatória:

O termo de consentimento está conciso, como deve ser, e apresenta: o título do projeto abaixo do título da folha, os riscos, os benefícios, garantia de esclarecimento, ausência de prejuízo caso recuse, garantia de sigilo, compromisso de divulgação dos resultados, ausência de ressarcimento pela participação, contatos do pesquisador e do CEP, e não apresenta espaços em branco. As pendências apontadas no parecer anterior sobre linguagem do objetivo e ajustes gramaticais foram todas cumpridas.

Recomendações:

Não há recomendações a fazer.

Conclusões ou Pendências e Lista de Inadequações:

O projeto está aprovado.

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

Considerações Finais a critério do CEP:

O projeto está aprovado.

Cabe ressaltar que o pesquisador se compromete em anexar na Plataforma Brasil um relatório ao final da realização da pesquisa. Pedimos a gentileza de utilizar o modelo de relatório final que se encontra na página eletrônica do CEP-UNISUAM (<http://www.unisuam.edu.br/index.php/introducao-comite-etica-em-pesquisa>). Além disso, em caso de evento adverso, cabe ao pesquisador relatar, também através da Plataforma Brasil.

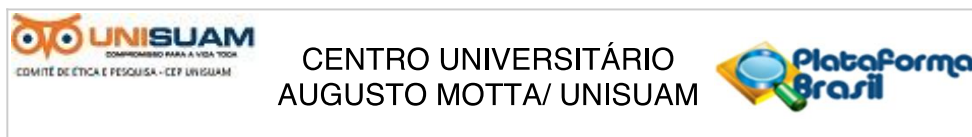
Endereço: Praça das Nações nº 34 TEL: (21)3882-9797 (Ramal : 1015)

Bairro: Bonsucesso **CEP:** 21.041-010

UF: RJ **Município:** RIO DE JANEIRO

Telefone: (21)3882-9797

E-mail: comitedeetica@unuam.edu.br



Continuação do Parecer: 817.427

RIO DE JANEIRO, 02 de Outubro de 2014

Assinado por:
Míriam Raquel Meira Mainenti
(Coordenador)

Endereço: Praça das Nações nº 34 TEL: (21)3882-9797 (Ramal : 1015)
Bairro: Bonsucesso **CEP:** 21.041-010
UF: RJ **Município:** RIO DE JANEIRO
Telefone: (21)3882-9797 **E-mail:** comitedeetica@unisuum.edu.br

Apêndice 1 – Questionário para caracterização da amostra

Examinador ID: __ __	
1) Data de nascimento: __ __ / __ __ / __ __ __ __	Idade: __ __ anos
2) Sexo: () Masculino () Feminino	
3) Área de graduação em saúde: () Biomedicina () Enfermagem () Educação Física () Fisioterapia () Fonoaudiologia () Medicina () Nutrição () Odontologia () Psicologia () Terapia Ocupacional () Outros. Especifique: _____	
4) Atividade profissional: () Profissional com atuação clínica em acupuntura (consultório, hospital, domiciliar, entre outros) () Professor de curso de especialização em acupuntura () Supervisor de estágio em acupuntura em clínica-escola	
5) Mês/ano de conclusão do curso de especialização: __ __ / __ __ __ __	
6) Tempo de formação em acupuntura: __ __ anos __ __ meses	
7) Duração do curso de especialização em acupuntura: __ __ anos __ __ meses	

Apêndice 2 – Caderno de casos clínicos



CENTRO UNIVERSITÁRIO AUGUSTO MOTTA
Vice-reitoria Acadêmica
Programa de Pós-graduação em Ciências da Reabilitação – PPGCR
Mestrado em Ciências da Reabilitação

Efeitos do erro diagnóstico na confiabilidade interexaminadores para diferenciação de padrões e prescrição de acupontos

CADERNO DE CASOS CLÍNICOS

Instruções para aplicação dos questionários. Os questionários são autoaplicáveis e o pesquisador não interferirá no seu preenchimento. Estão disponíveis para consulta: 1) tabela com **74 opções** de resposta de diagnóstico; 2) tabela com **379 opções** de resposta de acupontos; e 3) uma **referência bibliográfica**. Será dado um tempo de 10 minutos para familiarização com as tabelas contendo as opções de resposta.

DIAGNÓSTICO: indicar **1 item** da ficha de padrões.

ACUPONTOS: sugerir **de 1 a 8 itens** da ficha de acupontos.

Você pode fazer pausas durante sua participação.
Utilize o tempo que julgar necessário para preenchimento.

Caso 1

Cobertura espessa, sensação de frio, pulso profundo.

Caso 2

Cobertura espessa, constipação.

Caso 3

Prolapso uterino, ejaculação precoce, micção frequente, pulso debilitado, respingos após micção.

Caso 4

Língua vermelho-púrpura, desconforto torácico.

Caso 5

Calafrios, pulso superficial, obstrução nasal, cobertura branca, espirro.

Caso 6

Cansaço matinal, anorexia.

Caso 7

Ausência de sede, diarreia, anorexia.

Caso 8

Queimação no ânus, náusea, cobertura pegajosa, cobertura amarela.

Caso 9

Pontos púrpuros, menstruação com coágulos escuros.

Caso 10

Fome constante, dor abdominal, desejo de comer arroz cru.

Caso 11

Febre, urina escura, inabilidade para digerir gordura, cobertura pegajosa.

Caso 12

Memória debilitada, língua vermelha, calor dos cinco palmos.

Caso 13

Cobertura espessa, dispneia, náusea, cobertura pegajosa.

Caso 14

Cobertura branca, distensão abdominal.

Caso 15

Crises de dispneia, língua vermelha.

Caso 16

Edema em membros inferiores, palpitação, frio nas costas.

Caso 17

Dor abdominal, língua pálida, borborismo, pulso profundo.

Caso 18

Pulso lento, secreção vaginal branca, plenitude no epigástrico.

Caso 19

Incontinência urinária, cefaleia occipital crônica.

Caso 20

Língua pálida, língua edemaciada.

Caso 21

Dispneia, lábios cianóticos, língua púrpuro-azulada, respiração debilitada.

Caso 22

Língua pálida, dor abdominal, pulso profundo, diarreia.

Caso 23

Surdez, boca seca à noite, insônia (acorda com frequência), cobertura ausente, depressão, pulso vazio, emissão noturna, dor nos ossos, tontura persistente, sudorese noturna, constipação.

Caso 24

Língua vermelha, sensação de sufocamento.

Caso 25

Cobertura pegajosa, cobertura espessa, calafrios.

Caso 26

Cobertura branca, língua pálida.

Caso 27

Micção dolorida, desconforto torácico, calor no tórax.

Caso 28

Perda do equilíbrio, coma, língua rígida, febre, opistótono.

Caso 29

Dor abdominal, diarreia.

Caso 30

Tremor dos membros, língua vermelho-escura.

Caso 31

Gosto ruim na boca, fome constante.

Caso 32

Constipação, cobertura marrom.

Caso 33

Tontura, face pálida, palpitação noturna leve e ocasional, face opaca.

Caso 34

Edema no ânus, cobertura amarela, urina escura, cobertura seca.

Caso 35

Prurido no nariz, dor de garganta, tosse seca.

Caso 36

Pulso debilitado, frequência urinária, língua pálida.

Caso 37

Sensação de peso na cabeça, cobertura pegajosa.

Caso 38

Língua pálida, face pálida, língua edemaciada, pulso em nó, apatia, língua úmida, sudorese, palpitação diurna leve e ocasional, cansaço, pulso debilitado.

Caso 39

Tontura, língua levemente seca, insônia, lábios pálidos, laterais alaranjadas, língua seca, cefaleia pior à tarde/noite, língua pálida, laterais mais pálidas, flutuação dos olhos.

Caso 40

Perda do equilíbrio, língua rígida, coma.

Caso 41

Flatulência, cobertura espessa.

Caso 42

Cefaleia surda orbicular, insônia, tremores musculares, menstruação escassa, fezes ressecadas, face pálida, cefaleia pior à tarde/noite, amenorreia, língua seca, tontura, câibras.

Caso 43

Face pálida, metrorragia.

Caso 44

Distensão abdominal, pulso profundo, flatulência, borborismo.

Caso 45

Cobertura pegajosa, dor no hipocôndrio.

Caso 46

Insônia, cefaleia pior à tarde/noite, debilidade muscular, laterais mais pálidas.

Caso 47

Pulso profundo, lombalgia, frio nas costas, expectoração espumante, frio nas pernas.

Caso 48

Língua vermelha, constipação.

Caso 49

Diarreia, lassitude.

Caso 50

Dispneia ao esforço físico, palpitação diurna leve e ocasional, mãos frias, sensação de frio, desconforto precordial.

Caso 51

Língua pálida, dor escrotal, plenitude no hipogástrio, vômito, pulso em corda.

Caso 52

Cobertura amarela, língua vermelha.

Caso 53

Cefaleia frontal surda, membros debilitados, língua pálida.

Caso 54

Flutuação dos olhos, laterais mais pálidas, unhas secas, convulsões, laterais alaranjadas, fezes ressecadas, tremores musculares, insônia, língua pálida, cefaleia surda orbicular, cefaleia surda (topo da cabeça), tetania, pulso agitado, cegueira para cores, câibras, língua levemente seca, memória debilitada, cefaleia pior à tarde/noite.

Caso 55

Pulso vazio, cabelo sem brilho, cabelo quebradiço, perda de dentes.

Caso 56

Diarreia, anorexia.

Caso 57

Vômito após ingestão, sangramento gengival, constipação, cobertura amarela, sede (líquidos frios).

Caso 58

Parestesia nos membros, tontura, memória debilitada, convulsões, lábios pálidos, laterais alaranjadas, cefaleia surda orbicular, oligomenorréia, insônia, cabelo seco, constipação, face opaca, unhas secas, laterais mais pálidas, face pálida, cansaço, pulso agitado, fezes ressecadas, unhas quebradiças, tetania, cefaleia surda (topo da cabeça), unhas escuras, unhas denteadas.

Caso 59

Coma, lábios cianóticos, pulso em nó, membros frios.

Caso 60

Hematêmese (sangue escuro), petéquias púrpuras, pulso agitado.

Caso 61

Boca seca, sede.

Caso 62

Cansaço, língua pálida.

Caso 63

Hemorroidas, pulso fino.

Caso 64

Sudorese profusa, respiração debilitada, língua curta.

Caso 65

Boca seca, pulso vazio, língua seca.

Caso 66

Tendência a bater ou xingar, ponta edemaciada, pulso escorregadio, pápulas sobre a fissura central.

Caso 67

Língua pálida, pulso profundo.

Caso 68

Corpo magro, evacuação difícil, língua pálida, boca seca, língua vermelha.

Caso 69

Urgência urinária, cobertura amarela.

Caso 70

Vômito, náusea, língua rosada.

Caso 71

Dispneia ao esforço físico,, agitação mental, edema facial.

Caso 72

Pulso profundo, corpo magro, pulso debilitado.

Caso 73

Incontinência urinária, espermatorréia.

Caso 74

Língua edemaciada, joelhos frios, falta de força de vontade.

Caso 75

Opressão torácica, bochechas vermelhas, expectoração pegajosa, secreção nasal amarela, sede.

Caso 76

Cefaleia surda orbicular, tetania, insônia, cegueira para cores, face pálida, cefaleia pior à tarde/noite, amenorreia, câibras, cabelo seco, cefaleia pior com atividade.

Caso 77

Apatia, sudorese, plenitude precordial, mãos frias.

Caso 78

Cefaleia pior com relaxamento, cefaleia aguda orbicular, cefaleia lateral, sensação de exaltação, tontura severa, laterais mais vermelhas, gritos de fúria, língua vermelha, cefaleia piora com calor, zumbido.

Caso 79

Secreção nasal amarela, sabor oleoso, lábios secos, opressão torácica.

Caso 80

Desconforto torácico, face cianótica.

Caso 81

Pulso profundo, borborismo, flatulência.

Caso 82

Náusea, língua rosada, vômito.

Caso 83

Garganta seca, evacuação difícil, língua seca.

Caso 84

Tremor dos membros, pulso agitado.

Caso 85

Pulso profundo, pulso lento, língua pálida.

Caso 86

Anorexia, prolapso estomacal.

Caso 87

Diarreia, incontinência urinária, espermatorreia.

Caso 88

Face pálida, memória debilitada, parestesia nos membros, menstruação escassa, cefaleia surda (topo da cabeça), laterais mais pálidas, convulsões, pulso fino, debilidade muscular, unhas secas.

Caso 89

Sede (pequenos goles), vômito.

Caso 90

Gosto ruim na boca, desejo de comer terra.

OBRIGADO PELA SUA PARTICIPAÇÃO

[illegible]

Apêndice 4 – Tabela de padrões



CENTRO UNIVERSITÁRIO AUGUSTO MOTTA
Vice-reitoria Acadêmica
Programa de Pós-graduação em Ciências da Reabilitação – PPGCR
Mestrado em Ciências da Reabilitação

Efeitos do erro diagnóstico na confiabilidade interexaminadores para diferenciação de padrões e prescrição de acupontos

01 Agitação do Fogo do Xin	38 Calor-Cheio no Xiaochang
02 Colapso do Yang do Xin	39 Deficiência do Xiaochang e Frio
03 Deficiência do Qi do Xin	40 Dor do Qi do Xiaochang
04 Deficiência do Xue do Xin	41 Infestação de Vermes no Xiaochang
05 Deficiência do Yang do Xin	42 Qi Preso no Xiaochang
06 Deficiência do Yin do Xin	43 Calor-Umidade na Dan
07 Estase do Xue do Xin	44 Deficiência da Dan
08 Tanyin-Fogo afetando o Xin	45 Deficiência da Pangguang e Frio
09 Ascendência do Yang do Gan	46 Umidade-Calor na Pangguang
10 Aumento do Yang do Gan Causando Vento	47 Umidade-Frio na Pangguang
11 Calor Extremo Gerando Vento no Gan	48 Deficiência do Qi do Fei
12 Calor-Umidade no Gan e na Dan	49 Deficiência do Yin do Fei
13 Chama do Fogo do Gan em Ascendência	50 Invasão do Fei pelo Vento-Calor
14 Deficiência do Xue do Gan	51 Invasão do Fei pelo Vento-Frio
15 Deficiência do Xue do Gan Causando Vento	52 Invasão do Fei pelo Vento-Secura
16 Deficiência do Yin do Gan	53 Invasão do Fei pelo Vento-Água
17 Estagnação do Frio no Gan Jing	54 Obstrução do Qi do Fei
18 Estagnação do Qi do Gan	55 Secura do Fei
19 Estase do Xue do Gan	56 Tanyin-Calor Obstruindo o Fei
20 Fogo no Gan e Dan	57 Tanyin-Fluidos Obstruindo o Fei
21 Ausência de Firmeza do Qi do Shen	58 Tanyin-Umidade Obstruindo o Fei
22 Deficiência da Jing do Shen	59 Deficiência do Qi do Pi
23 Deficiência do Qi do Shen	60 Deficiência do Yang do Pi
24 Deficiência do Yang do Shen	61 Pi Não Controlando o Xue
25 Deficiência do Yang do Shen: Água Fluindo em Abundância para o Fei	62 Submersão do Qi do Pi
26 Deficiência do Yang do Shen: Água Fluindo em Abundância para o Xin	63 Umidade-Calor Invadindo o Pi
27 Deficiência do Yin do Shen	64 Umidade-Frio Invadindo o Pi
28 Deficiência do Yin do Shen: Fogo-Vazio Resplandecendo	65 Deficiência do Qi do Wei
29 Shen Falhando ao Receber o Qi	66 Deficiência do Wei e Frio
30 Tanyin Obscurecendo a Shen	67 Deficiência do Yin do Wei
31 Calor Obstruindo o Dachang	68 Estase do Xue no Wei
32 Calor no Dachang	69 Fogo no Wei
33 Colapso do Dachang	70 Frio Invadindo o Wei
34 Frio Invadindo o Dachang	71 Rebelião do Qi do Wei em Ascendência
35 Frio no Dachang	72 Retenção dos Alimentos no Wei
36 Secura no Dachang	73 Umidade-Calor Invadindo o Wei e o Pi
37 Umidade-Calor no Dachang	74 Não é possível identificar o padrão

Apêndice 5 – Tabela de acupontos

1 HT-1 Jiquan	66 KI-11 Henggu	131 GB-10 Fubai
2 HT-2 Qingling	67 KI-12 Dahe	132 GB-11 Touqiaoyin
3 HT-3 Shaohai	68 KI-13 Qixue	133 GB-12 Wangu
4 HT-4 Lingdao	69 KI-14 Siiman	134 GB-13 Benshen
5 HT-5 Tongli	70 KI-15 Zhongzhu	135 GB-14 Yangbai
6 HT-6 Yinxi	71 KI-16 Huangshu	136 GB-15 Toulinqi
7 HT-7 Shenmen	72 KI-17 Shangqu	137 GB-16 Muchuang
8 HT-8 Shaofu	73 KI-18 Shiguan	138 GB-17 Zhengying
9 HT-9 Shaochong	74 KI-19 Yindu	139 GB-18 Chengling
10 LU-1 Zhongfu	75 KI-20 Futonggu	140 GB-19 Naokong
11 LU-2 Yunmen	76 KI-21 Youmen	141 GB-20 Fengchi
12 LU-3 Tianfu	77 KI-22 Bulang	142 GB-21 Jianjing
13 LU-4 Xiabai	78 KI-23 Shenfeng	143 GB-22 Yuanze
14 LU-5 Chize	79 KI-24 Lingxu	144 GB-23 Zhejin
15 LU-6 Kongzui	80 KI-25 Shencang	145 GB-24 Riyue
16 LU-7 Lieque	81 KI-26 Yuzhong	146 GB-25 Jingmen
17 LU-8 Jingcu	82 KI-27 Shufu	147 GB-26 Daimai
18 LU-9 Taiyuan	83 SI-1 Shaoze	148 GB-27 Wushu
19 LU-10 Yuji	84 SI-2 Qiangou	149 GB-28 Weidao
20 LU-11 Shaoshang	85 SI-3 Houxi	150 GB-29 Juliao
21 LR-1 Dadun	86 SI-4 Wangu	151 GB-30 Huantiao
22 LR-2 Xingjian	87 SI-5 Yanggu	152 GB-31 Fengshi
23 LR-3 Taichong	88 SI-6 Yanglao	153 GB-32 Zhongdu
24 LR-4 Zhongfeng	89 SI-7 Zhizheng	154 GB-33 Xiyangguan
25 LR-5 Ligou	90 SI-8 Xiaohai	155 GB-34 Yanglingquan
26 LR-6 Zhongdu	91 SI-9 Jianshen	156 GB-35 Yangjiao
27 LR-7 Xiguan	92 SI-10 Naoshu	157 GB-36 Waigou
28 LR-8 Ququan	93 SI-11 Tianzong	158 GB-37 Guangming
29 LR-9 Yinbao	94 SI-12 Bingfeng	159 GB-38 Yangfu
30 LR-10 Zuwuli	95 SI-13 Quyuan	160 GB-39 Xuanzhong
31 LR-11 Yinlian	96 SI-14 Jianwaishu	161 GB-40 Qiuxu
32 LR-12 Jimai	97 SI-15 Jianzhongshu	162 GB-41 Zulingqi
33 LR-13 Zhangmen	98 SI-16 Tianchuang	163 GB-42 Diwuhui
34 LR-14 Qimen	99 SI-17 Tianrong	164 GB-43 Xiaxi
35 SP-1 Yinbai	100 SI-18 Quanliao	165 GB-44 Zuqiaoyin
36 SP-2 Dadu	101 SI-19 Tinggong	166 ST-1 Chengqi
37 SP-3 Taibai	102 LI-1 Shangyang	167 ST-2 Sibai
38 SP-4 Gongsun	103 LI-2 Erjian	168 ST-3 Juliao
39 SP-5 Shangqiu	104 LI-3 Sanjian	169 ST-4 Dicang
40 SP-6 Sanyinjiao	105 LI-4 Hegu	170 ST-5 Daying
41 SP-7 Lougu	106 LI-5 Yangxi	171 ST-6 Jiache
42 SP-8 Diji	107 LI-6 Pianli	172 ST-7 Xiaguan
43 SP-9 Yinlingquan	108 LI-7 Wenliu	173 ST-8 Touwei
44 SP-10 Xuehai	109 LI-8 Xialian	174 ST-9 Renying
45 SP-11 Jimen	110 LI-9 Shanglian	175 ST-10 Shuitu
46 SP-12 Chongmen	111 LI-10 Shousanli	176 ST-11 Qishe
47 SP-13 Fushe	112 LI-11 Qushi	177 ST-12 Quepen
48 SP-14 Fujie	113 LI-12 Zhouliao	178 ST-13 Qihu
49 SP-15 Daheng	114 LI-13 Shouwuli	179 ST-14 Hufang
50 SP-16 Fuai	115 LI-14 Binao	180 ST-15 Wuyi
51 SP-17 Shidou	116 LI-15 Jianyu	181 ST-16 Yingchuang
52 SP-18 Tianxi	117 LI-16 Jugu	182 ST-17 Ruzhong
53 SP-19 Xiongxiang	118 LI-17 Tianjing	183 ST-18 Rugen
54 SP-20 Zhourong	119 LI-18 Futu	184 ST-19 Burong
55 SP-21 Dabao	120 LI-19 Kouheliao	185 ST-20 Chengman
56 KI-1 Yongquan	121 LI-20 Yingxiang	186 ST-21 Llangmen
57 KI-2 Rangu	122 GB-1 Tongziliao	187 ST-22 Guanmen
58 KI-3 Taixi	123 GB-2 Tinghui	188 ST-23 Taiyi
59 KI-4 Dazhong	124 GB-3 Shangguan	189 ST-24 Huaroumen
60 KI-5 Shuiquan	125 GB-4 Hanyan	190 ST-25 Tianshu
61 KI-6 Zhaohai	126 GB-5 Xuanlu	191 ST-26 Wailing
62 KI-7 Fulu	127 GB-6 Xuanli	192 ST-27 Daju
63 KI-8 Jiaoxin	128 GB-7 Qubin	193 ST-28 Shuidao
64 KI-9 Zhubin	129 GB-8 Shuaigu	194 ST-29 Guilai
65 KI-10 Yingu	130 GB-9 Tianchong	195 ST-30 Qichong

196	ST-31 Biguan	260	BL-50 Wiechang	324	GV-15 Yamen
197	ST-32 Futu	261	BL-51 Huangmen	325	GV-16 Fengfu
198	ST-33 Yinshi	262	BL-52 Zhishi	326	GV-17 Naohu
199	ST-34 Liangqiu	263	BL-53 Baohuang	327	GV-18 Qiangjian
200	ST-35 Dubi	264	BL-54 Zhibian	328	GV-19 Houding
201	ST-36 Zusanli	265	BL-55 Heyang	329	GV-20 Baihui
202	ST-37 Shangjuxu	266	BL-56 Chengjin	330	GV-21 Qianding
203	ST-38 Tiaokou	267	BL-57 Chengshan	331	GV-22 Xinhui
204	ST-39 Xiajuxu	268	BL-58 Feiyang	332	GV-23 Shangxing
205	ST-40 Fenglong	269	BL-59 Fuyang	333	GV-24 Shenting
206	ST-41 Jiexi	270	BL-60 Kunlun	334	GV-25 Suliao
207	ST-42 Chongyang	271	BL-61 Pucan	335	GV-26 Shuigou
208	ST-43 Xiang	272	BL-62 Shenmai	336	GV-27 Duiduan
209	ST-44 Neiting	273	BL-63 Jinmen	337	GV-28 Yinjiao
210	ST-45 Lidui	274	BL-64 Jinggu	338	CV-1 Huiyin
211	BL-1 Jingming	275	BL-65 Shugu	339	CV-2 Qugu
212	BL-2 Cuanzhu	276	BL-66 Zutonggu	340	CV-3 Zhongji
213	BL-3 Meichong	277	BL-67 Zhiyin	341	CV-4 Guanyuan
214	BL-4 Qucha	278	TE-1 Guanchong	342	CV-5 Shimen
215	BL-5 Wuchu	279	TE-2 Yemen	343	CV-6 Qihai
216	BL-6 Chengguang	280	TE-3 Zhongzhu	344	CV-7 Yinjiao
217	BL-7 Tongtian	281	TE-4 Yangchi	345	CV-8 Shenque
218	BL-8 LUoque	282	TE-5 Waiguan	346	CV-9 Shuifen
219	BL-9 Yuzhen	283	TE-6 Zhigou	347	CV-10 Xiawan
220	BL-10 Tianzhu	284	TE-7 Huizong	348	CV-11 Jianli
221	BL-11 Dazhu	285	TE-8 Sanyangluo	349	CV-12 Zhongwan
222	BL-12 Engmen	286	TE-9 Sidu	350	CV-13 Shangwan
223	BL-13 Feishu	287	TE-10 Tianjing	351	CV-14 Juque
224	BL-14 Jueyinshu	288	TE-11 Qinglengyuan	352	CV-15 Jiuwei
225	BL-15 Xinshu	289	TE-12 Xiaoluo	353	CV-16 Zhongting
226	BL-16 Dushu	290	TE-13 Naohui	354	CV-17 Danzhong
227	BL-17 Geshu	291	TE-14 Jianliao	355	CV-18 Yutang
228	BL-18 Ganshu	292	TE-15 Tianliao	356	CV-19 Zigong
229	BL-19 Danshu	293	TE-16 Tianyou	357	CV-20 Huagai
230	BL-20 Pishu	294	TE-17 Yifeng	358	CV-21 Xuanji
231	BL-21 Weishu	295	TE-18 Chimai	359	CV-22 Tiantu
232	BL-22 Sanjiaoshu	296	TE-19 Luxi	360	CV-23 Lianquan
233	BL-23 Shenshu	297	TE-20 Jiaosun	361	CV-24 Chengjiang
234	BL-24 Qihai	298	TE-21 Ermen	362	Yintang
235	BL-25 Dashangshu	299	TE-22 Erheliao	363	Taiyang
236	BL-26 Guanyuanshu	300	TE-23 Sizhukong	364	Baxie
237	BL-27 Xiaochangshu	301	PC-1 Tianchi	365	Bafeng
238	BL-28 Pangguangshu	302	PC-2 Tianquan	366	Zigong
239	BL-29 Zhonglushu	303	PC-3 Quze	367	Yuyao
240	BL-30 Baihuanshu	304	PC-4 Ximen	368	Shifeng
241	BL-31 Shangliao	305	PC-5 Jianshi	369	Shixuan
242	BL-32 Ciliao	306	PC-6 Neiguan	370	Xiyan
243	BL-33 Zhongliao	307	PC-7 Daling	371	Jinggong
244	BL-34 Xialiao	308	PC-8 Laogong	372	Huatuojiaji
245	BL-35 Huiyang	309	PC-9 Zhongchong	373	Shencong
246	BL-36 Chengfu	310	GV-1 Changqiang	374	Dingchuan
247	BL-37 Yinmen	311	GV-2 Yaoshu	375	Shiqizhuixia
248	BL-38 Fuxi	312	GV-3 Yaoyangguan	376	JianneiLing
249	BL-39 Weiyang	313	GV-4 Mingmen	377	Dannangxue
250	BL-40 Weizhong	314	GV-5 Xuanshu	378	Lanweixue
251	BL-41 Fufen	315	GV-6 Jizhong	379	Nenhum acuponto
252	BL-42 Pohan	316	GV-7 Zhongshu		
253	BL-43 Gaohuang	317	GV-8 Jinsuo		
254	BL-44 Shentang	318	GV-9 Zhiyang		
255	BL-45 Yixi	319	GV-10 Lingtai		
256	BL-46 Geguan	320	GV-11 Shendao		
257	BL-47 Hunmen	321	GV-12 Shenzhu		
258	BL-48 Yanggang	322	GV-13 Taodao		
259	BL-49 Yishe	323	GV-14 Dazhui		

Apêndice 6 – Termo de consentimento livre e esclarecido



Projeto: Efeitos do erro diagnóstico na confiabilidade interexaminadores para diferenciação de padrões e prescrição de acupontos

Apresentação: As informações contidas neste termo foram elaboradas para sua participação voluntária neste estudo, que tem como objetivo investigar se diferentes desfechos diagnósticos em medicina chinesa (correto, ausente e errado) influenciam na variabilidade com que especialistas em acupuntura realizam sua avaliação clínica e prescrevem tratamento por acupuntura. Se concordar em participar deste estudo, você preencherá um questionário com perguntas sobre algumas características pessoais e profissionais. Em seguida, o questionário apresentará casos clínicos para os quais você deverá assinalar a opção diagnóstica que julgar correta. Você precisará também sugerir pontos de acupuntura para o tratamento de cada caso do questionário.

Potenciais riscos: Os questionários são individuais e anônimos, garantindo assim a sua confidencialidade e privacidade. Entretanto, você pode se sentir desconfortável por ter dificuldades em responder às questões técnicas do questionário. Para minimizar esse risco, os resultados não serão divulgados individualmente e somente você terá acesso ao seu resultado individual.

Potenciais benefícios: Os benefícios para você incluem o seu desempenho no diagnóstico dos casos apresentados, enquanto que para a população espera-se contribuir na diminuição de erros de diagnóstico.

Se desejar desistir do estudo em qualquer momento, você tem toda liberdade de fazê-lo, garantindo que a recusa de participação não acarretará penalização no seu cuidado. As informações a serem recebidas durante o estudo serão analisadas em conjunto com as informações obtidas de outros voluntários, não sendo divulgada a identificação de nenhum participante. Tais informações serão de responsabilidade dos pesquisadores envolvidos no projeto e serão utilizadas para fins científicos e não será permitido o acesso a terceiros, garantindo assim proteção contra qualquer tipo de discriminação. Se desejar, você pode ser informado sobre os resultados parciais da pesquisa a qualquer momento. Os resultados serão submetidos à publicação em revistas científicas.

Não haverá despesas pessoais para você em qualquer fase do estudo, nem haverá compensação financeira relacionada à sua participação. Em caso de dano pessoal diretamente causado pelos procedimentos propostos neste estudo, você terá direito a tratamento médico, bem como às indenizações legalmente estabelecidas.

Em qualquer etapa do estudo, você terá acesso ao profissional responsável, Dra. Ingrid Jardim de Azeredo Souza Oliveira (CREFITO-2 nº 49658-F) que pode ser encontrada no telefone (21) 98318-0221. Se tiver alguma consideração ou dúvida sobre a ética da pesquisa, entre em contato com o Comitê de Ética em Pesquisa (CEP): Praça das Nações, nº 34 - Bonsucesso, Rio de Janeiro – RJ, Tel.: (21) 3882-9797 (ramal 1015), e-mail: comitedeetica@unisiam.edu.br.

Declaração: Acredito ter sido suficientemente informado sobre o estudo acima citado que li ou que foram lidas para mim. Ficaram claros para mim quais são os propósitos do estudo, os procedimentos a serem realizados, seus desconfortos e riscos, as garantias de confidencialidade e de esclarecimentos permanentes. Ficou claro também que minha participação é isenta de despesas e que tenho garantia de acesso a tratamento hospitalar se necessário em decorrência desse estudo. Concordo voluntariamente em participar deste estudo

e poderei retirar o meu consentimento a qualquer momento, antes ou durante o mesmo, sem penalidades ou prejuízo ou perda de qualquer benefício que eu possa ter adquirido.

Nome/Assinatura do participante

Data:
____/____/____

Ingrid Jardim de Azeredo Souza Oliveira

Data:
____/____/____

Apêndice 7 – Comprovantes de submissão dos manuscritos

De: Monica Toma monica.toma@hindawi.com
Assunto: Re: Invitation to Submit an Invited Contribution to Evidence-Based Complementary and Alternative Medicine
Data: 12 de janeiro de 2015 05:11
Para: Ferreira AS arthur_sf@ig.com.br

MT

Dear Dr. de Sá Ferreira,

This is to confirm that the manuscript titled "Effects of Diagnostic Errors in Pattern Differentiation and Acupuncture Prescription: A Single-Blinded, Interrater Agreement Study" has been submitted to Evidence-Based Complementary and Alternative Medicine and assigned the number 469675. For further correspondence, my colleague, Ms. Aya Momen would be glad to receive your inquiries and/or comments through her email address aya.momen@hindawi.com.

Best regards,

Monica Toma

--

Monica Toma
 Editorial Office
 Hindawi Publishing Corporation
<http://www.hindawi.com>
