

Análise da confiabilidade do Sistema de Triagem de Manchester: concordância interna e entre observadores¹

Cristiane Chaves de Souza²
Tânia Couto Machado Chianca³
Welfane Cordeiro Júnior⁴
Maria do Carmo Paixão Rausch⁵
Gabriela Fontoura Lana Nascimento⁶

Objetivo: analisar a confiabilidade do Sistema de Triagem de Manchester para determinar o grau de prioridade de pacientes em serviços de urgência. Método: trata-se de estudo de confiabilidade com amostra de 361 enfermeiros. Os dados foram coletados em três etapas e os questionários aplicados utilizando *software* eletrônico. A concordância foi mensurada pela avaliação de casos clínicos. Os desfechos avaliados foram: concordância com o padrão ouro e intraobservadores na indicação do fluxograma, discriminador e nível de risco. Os dados foram submetidos a análises uni e bivariadas. A concordância foi mensurada pelo cálculo do índice Kappa. Resultados: a confiabilidade externa e interna do protocolo variou de moderada a substancial (Kappa:0,55–0,78). Tempo de experiência profissional como enfermeiro, em serviços de urgência e emergência e na classificação de risco, foi associado à confiabilidade externa e interna. A escolha certa do discriminador influenciou mais a correta indicação do nível de risco ($R^2=0,77$; $p<0,0001$) do que a escolha correta do fluxograma ($R^2=0,16$; $p<0,0001$). Conclusão: a confiabilidade do Sistema de Triagem de Manchester variou de moderada a substancial e foi influenciada pela experiência clínica do enfermeiro. O protocolo é seguro para definição das prioridades clínicas utilizando diferentes fluxogramas de classificação.

Descritores: Enfermagem; Serviços Médicos de Emergência; Triagem; Avaliação em Enfermagem; Reprodutibilidade dos Testes; Validade dos Testes.

¹ Artigo extraído da tese de doutorado "Análise da confiabilidade do Sistema de Triagem de Manchester para determinar o grau de prioridade de pacientes em serviços de urgência", apresentada à Escola de Enfermagem, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil.

² PhD, Professor Adjunto, Departamento de Medicina e Enfermagem, Universidade Federal de Viçosa, Viçosa, MG, Brasil.

³ PhD, Professor Titular, Escola de Enfermagem, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil.

⁴ MSc, Presidente, Grupo Brasileiro de Classificação de Risco, Belo Horizonte, MG, Brasil.

⁵ Especialista em Epidemiologia em Serviços de Saúde, Diretor Administrativo, Grupo Brasileiro de Classificação de Risco, Belo Horizonte, MG, Brasil.

⁶ Enfermeira, Grupo Brasileiro de Classificação de Risco, Belo Horizonte, MG, Brasil.

Como citar este artigo

Souza CC, Chianca TCM, Cordeiro Junior W, Rausch MCP, Nascimento GFL. Reliability analysis of the Manchester Triage System: inter-observer and intra-observer agreement. Rev. Latino-Am. Enfermagem. 2018;26:e3005. [Access]; Available in: . DOI: <http://dx.doi.org/10.1590/1518-8345.2205.3005>.

mês dia ano

URL

Introdução

O processo de triagem configura-se como elemento intuitivo da prática clínica de enfermeiros que atuam em serviços de urgência e emergência. Atribuir um grau de risco ao paciente é um complexo processo de tomada de decisão. Segundo a Teoria do Continuum Cognitivo, para tomar decisões, os enfermeiros utilizam-se do raciocínio clínico que envolve um pensamento ordenado e intencional, fundamentado nos conhecimentos teóricos e práticos e na experiência pessoal e profissional⁽¹⁾. Na triagem, a tomada de decisão é influenciada por sistemas de apoio ao julgamento clínico, pelo julgamento intuitivo e reflexivo e pela avaliação por pares envolvendo enfermeiro e paciente⁽²⁾.

Os sistemas de apoio ao julgamento referem-se ao uso das escalas ou sistemas de triagem que norteiam o enfermeiro na avaliação da queixa apresentada pelo paciente. O julgamento intuitivo e o reflexivo são fortemente influenciados pelo tempo de experiência profissional do enfermeiro e pelo uso das experiências anteriores para julgar novos casos e tomar decisões. A avaliação por pares envolvendo enfermeiro e paciente implica na certificação da credibilidade da queixa apresentada pelo paciente, envolvendo-o no processo de avaliação e confirmação dos achados clínicos que sustentam a tomada de decisão acerca do nível de prioridade⁽²⁾.

No que se refere aos sistemas de apoio, sistemas ou protocolos de triagem têm sido desenvolvidos para guiar a avaliação do enfermeiro⁽³⁾. Dentre estes, destaca-se o Sistema de Triagem de Manchester (STM), que estratifica em cinco os níveis de gravidade e atribui, a cada nível, cor e tempo-alvo para atendimento médico. É estruturado em fluxogramas com discriminadores que orientam a coleta e análise de informações para a definição da prioridade clínica do paciente⁽⁴⁾.

Em um contexto de demanda por serviços maior que a oferta e com recursos limitados de atendimento, a triagem dos pacientes deve acontecer de forma acurada para garantir o cuidado de acordo com a real necessidade do paciente⁽⁵⁾. Assim, a tomada de decisão do enfermeiro na classificação de risco deve ser guiada por um sistema de apoio confiável, capaz de mensurar com precisão o grau de prioridade do paciente.

A confiabilidade é considerada o principal critério para testar a qualidade dos instrumentos de medida. Trata-se da habilidade do instrumento mensurar, de forma consistente e acurada, aquilo que pretende medir e reproduzir um resultado de forma consistente no tempo e no espaço ou com observadores diferentes⁽⁶⁾.

Apesar do STM ter sido desenvolvido há duas décadas e ser amplamente utilizado em diferentes países

como sistema de apoio ao enfermeiro para tomada de decisão na triagem, sua confiabilidade tem sido pouco questionada⁽⁷⁾. Uma revisão integrativa apontou que a confiabilidade do STM variou de moderada a quase perfeita, o que mostra a necessidade de novos estudos para identificar modificações necessárias no protocolo e aumentar a segurança na gestão do risco clínico dos pacientes em serviços de urgência⁽⁸⁾.

Verificam-se variações nos estudos sobre a confiabilidade do STM que indicam lacuna sobre a real confiabilidade desta escala de triagem⁽⁷⁻⁸⁾. Sendo assim, questiona-se: Qual a confiabilidade do STM para determinar o grau de prioridade do paciente, considerando a concordância interna e entre enfermeiros que utilizam este protocolo? Assim, delineou-se este estudo com objetivo de analisar a confiabilidade do STM para determinar o grau de prioridade de pacientes em serviços de urgência.

Método

Trata-se de estudo de confiabilidade, realizado em parceria com o Grupo Brasileiro de Classificação de Risco (GBCR), que é o responsável pela capacitação dos enfermeiros brasileiros no uso e na auditoria do STM.

A população do estudo foi composta por enfermeiros habilitados pelo GBCR entre janeiro de 2008 e agosto de 2014 para usar o STM na prática clínica (N=11.711). A habilitação foi obtida no curso de classificador ministrado por médicos e enfermeiros do GBCR, na modalidade presencial, com duração de 12 horas. Ao final, os enfermeiros realizaram uma avaliação teórica e os que atingiram índice de aproveitamento maior ou igual a 60%, foram considerados aptos a utilizar o STM na prática clínica. Foram incluídos no estudo os enfermeiros que possuíam endereço eletrônico válido e habilitados no curso de classificador (n= 6.227 enfermeiros).

A amostra foi calculada considerando intervalo de confiança de 95%, margem de erro de 5%, estimativa de acertos entre observadores e padrão ouro de 75% e estimativa de acertos intraobservadores de 85%. A estimativa de acertos utilizada foi baseada em estudo de referência previamente realizado⁽⁹⁾. Assim, para mensurar a confiabilidade externa do STM, a amostra mínima necessária foi de 273 enfermeiros e para mensurar a confiabilidade interna de 152 enfermeiros.

Entretanto, cabe ressaltar que a amostra obtida no presente estudo foi maior do que o mínimo necessário. A amostra final atingida foi de 361 enfermeiros para mensuração da confiabilidade externa do STM e de 153 enfermeiros para avaliação da confiabilidade interna. Estes eram enfermeiros de 21 diferentes estados

brasileiros, sendo a maior parcela dos respondentes dos estados de Minas Gerais (93–25,76%), São Paulo (49–13,58%), Distrito Federal (45–12,46%), Espírito Santo (45–12,46%), Rio Grande do Sul (31–8,59%), Ceará (26–7,20%), Santa Catarina (23–6,37%) e o restante (49–13,58%) dos estados de Alagoas, Amazonas, Bahia, Goiás, Mato Grosso, Mato Grosso do Sul, Pará, Paraíba, Paraná, Pernambuco, Piauí, Rio de Janeiro, Rio Grande do Norte e Sergipe.

A confiabilidade do STM foi mensurada pelo parâmetro da estabilidade, que consiste na administração do mesmo instrumento aos mesmos sujeitos sob condições semelhantes em duas ou mais ocasiões⁽¹⁰⁾. Casos clínicos foram avaliados por enfermeiros em dois momentos distintos (teste e reteste).

A coleta dos dados foi realizada entre os meses de setembro de 2014 e agosto de 2015 e envolveu três etapas. A primeira consistiu na obtenção e validação dos casos clínicos. Foi obtido, com o GBCR, um total de 40 casos, que correspondia ao total de casos utilizados à época para capacitação dos enfermeiros brasileiros no uso do STM. Os casos foram validados quanto ao conteúdo por um grupo de três especialistas habilitados no uso do STM na prática clínica e com publicações sobre o STM, dois com experiência profissional no uso do protocolo na prática clínica e um com experiência no ensino e orientação de projetos de pesquisa envolvendo o STM.

Para validação dos casos clínicos foi utilizada a versão II da Escala de Acurácia de Diagnósticos de Enfermagem⁽¹¹⁾, adaptada para o estudo após consentimento das autoras a fim de identificar o grau em que um conjunto de pistas descritas em cada caso clínico permitia a identificação do nível de risco do paciente, seguindo o STM. Para cada caso clínico, os especialistas avaliaram a presença, relevância, especificidade e coerência das pistas de acordo com o desfecho clínico (nível de risco do paciente). Os escores de acurácia da escala variam de zero a 13,5, indicando desde acurácia nula até alta acurácia. Foram considerados válidos os casos clínicos em que os especialistas concordaram em 100% que as pistas descritas eram de alta acurácia (n= 28).

Na segunda etapa da coleta dos dados foi feito o teste para avaliação da confiabilidade externa do STM, mensurada pela concordância entre os enfermeiros e o padrão ouro na indicação, para cada caso clínico, do fluxograma, discriminador e nível de risco. O padrão ouro do estudo foi o gabarito dos casos clínicos fornecido pelo GBCR. Para a coleta dos dados, utilizou-se o *software* eletrônico SurveyGizmo® com questionários eletrônicos autoaplicados usando como plataforma computadores conectados à internet. O *link*

para acesso aos instrumentos de coleta de dados foi enviado para o *e-mail* dos enfermeiros participantes do estudo (n=6.227) através da ferramenta MailChimp®. Ao acessar o link, o enfermeiro era direcionado ao Termo de Consentimento Livre e Esclarecido e, caso aceitasse participar do estudo, era direcionado para a página inicial, contendo instruções sobre como proceder para responder às questões. Foram configurados no sistema SurveyGizmo® dois instrumentos de coleta de dados intitulados “Casos clínicos” e “Perfil profissional dos enfermeiros”.

O instrumento “Casos clínicos” continha os 28 casos clínicos validados. Para cada caso clínico, o enfermeiro deveria indicar o fluxograma, discriminador e nível de risco utilizando o STM. Todas as respostas foram pré-formatadas e não foram permitidas respostas discursivas. Ao clicar no campo “Fluxograma” era disponibilizada a listagem dos 50 fluxogramas existentes no STM. Foram excluídos os dois fluxogramas do STM relacionados ao atendimento a catástrofes. Ao clicar no fluxograma escolhido, era disponibilizado o organograma específico daquele fluxograma, bem como a definição dos discriminadores, tal como consta na versão impressa do STM. Uma vez escolhido o discriminador da classificação, o enfermeiro indicava o nível de risco do paciente.

O instrumento “Perfil profissional dos enfermeiros” continha 13 questões agrupadas em dois blocos: o de perguntas referentes ao perfil demográfico e de formação do enfermeiro e o de perguntas referentes à experiência clínica do enfermeiro na enfermagem, em serviços de urgência e emergência, e na classificação de risco.

Na terceira etapa da coleta dos dados foi feito o reteste para avaliação da confiabilidade interna do STM, mensurada pela concordância intraobservadores. Foi utilizada a mesma estratégia de coleta de dados descrita na etapa dois do estudo. Todos os enfermeiros que responderam ao questionário na segunda etapa de coleta de dados (n=361) foram convidados a participar desta etapa do estudo respondendo novamente ao instrumento “Casos clínicos”, sendo-lhes ocultada a primeira classificação.

Os dados coletados foram tabulados e submetidos a análises uni e bivariadas utilizando o programa estatístico SPSS (Statistical Package for Social Sciences) versão 19.0.

Para traçar o perfil profissional dos enfermeiros, utilizou-se estatística descritiva com distribuição de frequência. Para avaliar se o número de acertos na indicação do fluxograma e do discriminador influencia a escolha correta do nível de risco, empregou-se análise de regressão linear. Para esta análise, a variável dependente foi “número de acertos na escolha do nível

de risco” e as variáveis independentes “número de acertos na escolha do fluxograma” e “número de acertos na escolha do discriminador”. Avaliou-se também o quanto a escolha correta do fluxograma influenciou na indicação correta do discriminador. Para esta análise, a variável dependente foi “número de acertos na escolha do discriminador” e a variável independente foi “número de acertos na escolha do fluxograma”.

Os valores de *undertriage* e *overtriage* que traduzem, respectivamente, a porcentagem de pacientes triados pelos enfermeiros para níveis de menor e maior gravidade, quando comparados ao padrão ouro, foram mensurados utilizando estatística descritiva com cálculo do % de concordância entre os enfermeiros e o padrão ouro na escolha do nível de risco.

A confiabilidade externa e interna do STM foi avaliada através do índice de concordância Kappa linear e não ponderado, que mede a concordância intraobservadores ou entre observadores além do esperado pelo acaso em que sejam submetidos o mesmo número de sujeitos. A concordância foi considerada: nula ($k=0$); pobre (0,01 – 0,19); fraca (0,20 – 0,39); moderada (0,40 – 0,59); substancial (0,60 – 0,79); e quase perfeita (0,80 – 1)⁽¹²⁾.

Para avaliar a associação entre a confiabilidade externa e interna do STM e as variáveis do perfil profissional, utilizou-se o teste não paramétrico de Kruskal-Wallis e, para os casos de mais de 2 grupos em que foram encontradas diferenças significativas, o teste não paramétrico de Mann-Whitney com correção de Bonferroni.

O estudo foi aprovado pelo Comitê de Ética e Pesquisa (Parecer Etic Nº 816.372) e seguiu o disposto na Resolução nº 466/12 do Conselho Nacional de Saúde sobre Pesquisa envolvendo seres humanos.

Resultados

Dos 361 enfermeiros que participaram do estudo, a maioria (294 – 81,44%) era do sexo feminino, com idade entre 23 e 62 anos (média: 34,16 anos + 8,17 anos). A Tabela 1 mostra a caracterização dos enfermeiros do estudo, segundo as variáveis do perfil profissional investigadas.

Tabela 1 – Perfil profissional dos enfermeiros do estudo (n=361). Belo Horizonte, MG, Brasil, 2015

Variável	n	%
Tempo de graduação em Enfermagem		
Menor que um ano	2	0,55
Entre um e cinco anos	144	39,89
Entre cinco e dez anos	142	39,34
Mais de dez anos	73	20,22
Maior grau de formação obtido ou em andamento		
Graduação	40	11,08
Especialização	276	76,46
Mestrado	33	9,14
Doutorado	12	3,32
Formação relacionada à classificação de risco durante a graduação		
Nenhum conteúdo	240	66,48
Apenas conteúdo teórico	90	24,93
Conteúdo teórico e prático	31	8,59
Formação relacionada ao STM* durante a graduação		
Nenhum conteúdo	282	78,12
Apenas conteúdo teórico	60	16,62
Conteúdo teórico e prático com pouca ou nenhuma aplicação na prática clínica	16	4,43
Conteúdo teórico e prático com muita aplicação na prática clínica	3	0,83
Tempo de experiência profissional como enfermeiro		
Menor que um ano	30	8,31
Entre um e cinco anos	164	45,43
Entre cinco e dez anos	96	26,59
Mais de dez anos	71	19,67
Tempo de experiência em serviços de Urgência e Emergência		
Menor que um ano	82	22,71
Entre um e cinco anos	188	52,08
Entre cinco e dez anos	58	16,07
Mais de dez anos	33	9,14
Tempo de experiência na classificação de risco como enfermeiro		
Nunca	40	11,08
Menor que um ano	111	30,75
Entre um e cinco anos	198	54,85
Entre cinco e dez anos	11	3,05
Mais de dez anos	1	0,27

*STM: Sistema de Triagem de Manchester

A Tabela 2 mostra os valores da média e variações dos acertos dos enfermeiros em relação ao padrão ouro na escolha do fluxograma, discriminador e nível de risco, considerando os 28 casos avaliados.

Tabela 2 – Acerto dos enfermeiros com o padrão ouro na indicação do fluxograma, discriminador e nível de risco (n=361). Belo Horizonte, MG, Brasil, 2015

Variável	Número de acertos			
	Média	Desvio Padrão	Valor máximo	Valor mínimo
Fluxograma	20,60	2,58	27	11
Discriminador	16,40	3,97	26	4
Nível de Risco	21,72	3,24	28	11

A análise de regressão linear mostrou que há relação direta entre o número de acertos do fluxograma e do discriminador e a indicação do nível de risco ($p < 0,0001$). A variação de acertos na escolha do fluxograma explicou 16% da variação de acertos do nível de risco ($R^2=0,16$; $p < 0,0001$; IC 95%: 0,39-0,62), ao passo que a indicação correta do discriminador explicou 77% dos acertos do nível de risco ($R^2=0,77$; $p < 0,0001$; IC 95%: 0,67-0,76). A variação de acertos na escolha do fluxograma explicou 23% da variação no acerto dos discriminadores ($R^2=0,23$; $p < 0,0001$; IC 95%: 0,60-0,88), comprovando a segurança do protocolo na determinação do nível de prioridade a partir de diferentes fluxogramas de apresentação, já que há queixas que podem ser avaliadas utilizando diferentes fluxogramas de apresentação.

Tanto na avaliação da confiabilidade externa como interna do STM, a *overtriage* foi mais frequente no nível V de gravidade (cor azul), ocorrendo em 17% a 18% dos

casos. A *undertriage* foi mais frequente no nível II de gravidade (cor laranja), ocorrendo em 27% dos casos. Em sua maioria, os pacientes foram triados para um nível acima e um nível abaixo da gravidade real do caso.

A confiabilidade externa e interna do STM foi mensurada pelo cálculo do índice kappa para a escolha do fluxograma, discriminador e nível de risco (Tabela 3).

Para a escolha do fluxograma e do nível de risco, o STM apresenta uma concordância substancial. Para a escolha do discriminador, a concordância é moderada. Procedeu-se à investigação para verificar a associação entre os valores de kappa e as variáveis do perfil profissional. A Tabela 4 apresenta as variáveis em que foram encontradas associações significativas entre os valores de Kappa e a escolha do fluxograma e discriminador de classificação na avaliação da confiabilidade externa do STM. Não foi encontrada associação entre as variáveis do perfil profissional e a concordância na escolha do nível de risco na avaliação da confiabilidade externa do STM.

Tabela 3 – Confiabilidade externa (n=361) e interna (n=153) do Sistema de Triagem de Manchester: valores de Kappa. Nível de significância: $p < 0,001$. Belo Horizonte, MG, Brasil, 2015

Variável	Confiabilidade externa		Confiabilidade interna	
	Kappa: enfermeiros e padrão ouro	Kappa intraobservadores	Kappa: enfermeiros e padrão ouro	
Fluxograma	0,72	0,78	0,73	
Discriminador	0,55	0,57	0,59	
Nível de Risco	0,69	0,70	0,72	

Tabela 4 – Confiabilidade externa do Sistema de Triagem de Manchester: fatores associados à concordância entre enfermeiros e padrão ouro na escolha do fluxograma e discriminador (n=361). Belo Horizonte, MG, Brasil, 2015

Variável	Categorias (Tempo)	Índice Kappa			P-valor*	
		Média	Desvio Padrão	Mediana		
Escolha do Fluxograma	Experiência em urgência e emergência [‡]	Menor que 1 ano (a [†])	0,70	0,10	0,70	0,04
		Entre 1 e 5 anos (a [†] ,b [†])	0,73	0,09	0,74	
		Entre 5 e 10 anos (b [†])	0,74	0,10	0,74	
		Mais de 10 anos (a [†] ,b [†])	0,71	0,08	0,70	
Escolha do Discriminador	Experiência como enfermeiro [‡]	Menor que 1 ano (a [†])	0,45	0,15	0,44	<0,001
		Entre 1 e 5 anos (a [†] ,b [†])	0,58	0,14	0,62	
		Entre 5 e 10 anos (b [†])	0,56	0,15	0,58	
		Mais de 10 anos (a [†] ,b [†])	0,52	0,13	0,55	
	Experiência em urgência e emergência [‡]	Menor que 1 ano (a [†])	0,48	0,13	0,51	<0,001
		Entre 1 e 5 anos (a [†] ,b [†])	0,52	0,17	0,51	
		Entre 5 e 10 anos (b [†])	0,59	0,13	0,59	
		Mais de 10 anos (a [†] ,b [†])	0,60	0,07	0,58	
Experiência na classificação de risco [§]	Nunca atuou (b [†])	0,50	0,12	0,53	<0,001	
	Menor que 1 ano (b [†])	0,55	0,15	0,57		
	Entre 1 e 5 anos (c [†])	0,61	0,12	0,64		
	Entre 5 e 10 anos (c [†])	0,62	0,06	0,60		
	Mais de 10 anos (a [†] ,c [†])	0,40	--	0,32		

*P-valor: valores de p obtidos com o Teste de Kruskal-Wallis. Nível de significância: $p < 0,05$.

†a, b, c: As letras "a", "b" e "c" foram utilizadas para nomear os grupos comparados no teste de Mann-Whitney com correção de Bonferroni. Letras iguais simbolizam igualdade entre grupos e letras diferentes refletem as diferenças encontradas entre os grupos.

‡Experiência em urgência e emergência, Experiência como enfermeiro: Nível de significância adotado para o Teste de Mann-Whitney com correção de Bonferroni para comparação dos grupos: $p < 0,006$.

§ Experiência na classificação de risco: Nível de significância adotado para o Teste de Mann-Whitney com correção de Bonferroni para comparação dos grupos: $p < 0,0001$.

Na avaliação da confiabilidade externa do STM para escolha do fluxograma de classificação, enfermeiros que possuem entre cinco e dez anos de experiência em serviços de urgência e emergência apresentaram maior concordância com o padrão ouro, quando comparados àqueles com menos de um ano de experiência, mas, para ambos os grupos, a concordância foi substancial (Tabela 4).

Para a escolha do discriminador, enfermeiros com menos de um ano de experiência profissional apresentaram menor concordância com o padrão ouro, quando comparados aos demais grupos. De modo semelhante, enfermeiros com menos de um ano de experiência em serviços de urgência e emergência concordaram menos com o padrão ouro na escolha do discriminador, quando comparados àqueles que possuem entre um e cinco e entre cinco e dez anos de experiência e, quanto maior o tempo de experiência, maiores os valores de kappa encontrados. Não houve diferença entre os que nunca atuaram e os que atuam há menos de

um ano na classificação de risco e a concordância com o padrão ouro na escolha do discriminador de classificação. No entanto, quem atua há menos de um ano obteve maior concordância com o padrão ouro do que os que atuam há mais de dez anos. Enfermeiros com tempo de experiência na classificação de risco entre cinco e dez anos apresentaram os maiores valores de kappa com o padrão ouro na escolha do discriminador (Tabela 4).

A confiabilidade interna do STM foi avaliada por 153 (42,4%) dos 361 enfermeiros que avaliaram a confiabilidade externa do STM e a concordância encontrada variou de moderada a substancial (Tabela 3). A maior parcela dos enfermeiros que participaram desta etapa do estudo possuía entre um e cinco anos de experiência profissional como enfermeiro (43,8%), em serviços de urgência e emergência (58,2%), e na classificação de risco (59,5%). A Tabela 5 mostra as variáveis em que foram encontradas associações significativas entre os valores de Kappa e a escolha do fluxograma, discriminador e nível de risco.

Tabela 5 – Confiabilidade interna do Sistema de Triagem de Manchester: fatores associados à concordância na escolha do fluxograma, discriminador e nível de risco (n=153). Belo Horizonte, MG, Brasil, 2015

Variável	Categorias (Tempo)	Índice Kappa			P-valor*	
		Média	Desvio Padrão	Mediana		
Escolha do Fluxograma	Experiência na classificação de risco [‡]	Nunca atuou (a [†])	0,73	0,09	0,73	0,001
		Menor que 1 ano (a [†])	0,75	0,09	0,77	
		Entre 1 e 5 anos (b [†])	0,80	0,09	0,81	
		Entre 5 e 10 anos (a [†] ,b [†])	0,79	0,06	0,77	
Escolha do Discriminador	Experiência em urgência e emergência [‡]	Nunca atuou (a [†])	0,48	0,14	0,51	0,001
		Menor que 1 ano (a [†])	0,60	0,14	0,59	
		Entre 1 e 5 anos (b [†])	0,58	0,11	0,55	
		Entre 5 e 10 anos (a [†] ,b [†])	0,51	0,04	0,51	
	Experiência na classificação de risco [‡]	Nunca atuou (a [†])	0,52	0,14	0,53	<0,001
		Menor que 1 ano (a [†])	0,51	0,12	0,53	
		Entre 1 e 5 anos (b [†])	0,61	0,13	0,62	
		Entre 5 e 10 anos (a [†] ,b [†])	0,50	0,08	0,47	
Escolha do nível de risco	Experiência em urgência e emergência [‡]	Nunca atuou (a [†])	0,60	0,17	0,59	<0,001
		Menor que 1 ano (a [†])	0,75	0,14	0,75	
		Entre 1 e 5 anos (b [†])	0,69	0,11	0,70	
		Entre 5 e 10 anos (a [†] ,b [†])	0,60	0,16	0,64	
	Experiência na classificação de risco [‡]	Nunca atuou (a [†])	0,63	0,15	0,65	<0,001
		Menor que 1 ano (a [†])	0,62	0,15	0,61	
		Entre 1 e 5 anos (b [†])	0,74	0,15	0,76	
		Entre 5 e 10 anos (a [†] ,b [†])	0,68	0,07	0,66	

*P-valor: valores de p obtidos com o Teste de Kruskal-Wallis. Nível de significância adotado: p<0,05.

[†]a, b: As letras "a" e "b" foram utilizadas para nomear os grupos comparados no teste de Mann-Whitney com correção de Bonferroni. Letras iguais simbolizam igualdade entre grupos e letras diferentes refletem as diferenças encontradas entre os grupos.

[‡]Experiência na classificação de risco, Experiência em urgência e emergência: Nível de significância adotado para o Teste de Mann-Whitney com correção de Bonferroni para comparação dos grupos: p<0,006.

O tempo de experiência profissional influenciou na concordância interna do STM. Na escolha do fluxograma de classificação, enfermeiros que possuem entre um e cinco anos de experiência prática na classificação

de risco apresentaram concordância quase perfeita, quando comparados àqueles com menos de um ano ou que nunca atuaram na classificação de risco. Na escolha do discriminador, enfermeiros que possuem entre um

e cinco anos de experiência em serviços de urgência e emergência apresentaram valores médios de kappa maiores do que aqueles que possuem menos de um ano de experiência. De forma semelhante, enfermeiros que possuem entre um e cinco anos de experiência na classificação de risco apresentaram kappa médio maior do que aqueles que nunca atuaram e do que aqueles com menos de um ano de experiência neste cenário de prática (Tabela 5).

Discussão

Dos enfermeiros que participaram deste estudo, a maioria (79,23%) possui entre um e dez anos de graduação, 66,48% afirmam não ter tido nenhum conteúdo sobre classificação de risco durante a formação e 78,12% não tiveram nenhum conteúdo sobre o STM durante a graduação. Para tomar a decisão na classificação de risco, o enfermeiro integra a avaliação da queixa apresentada pelo paciente com o seu conhecimento adquirido durante a formação e a vida profissional, bem como o ambiente de cuidado no qual está inserido⁽⁵⁾. Assim, embora seja necessário ser habilitado pelo GBCR para utilizar o STM na prática clínica, recomenda-se que conteúdos sobre a classificação de risco e escalas de triagem, em especial o STM, sejam incluídos nas disciplinas obrigatórias de formação do enfermeiro desde a graduação.

As escalas ou sistemas de triagem compõem os sistemas de apoio à tomada de decisão do enfermeiro na triagem⁽²⁾. Neste estudo, foi avaliada a confiabilidade do STM. A escolha correta do fluxograma explicou 16% da variação na indicação correta do nível de risco, enquanto a escolha correta do discriminador de classificação explicou 77% da variação na indicação correta do nível de risco ($p < 0,0001$). Este achado mostra que, assim como prevê o STM, o enfermeiro pode escolher fluxogramas parecidos na avaliação de uma queixa que irão conduzir ao mesmo nível de risco, o que garante a segurança do protocolo. Isso porque muitas queixas podem levar à escolha de mais de um fluxograma de apresentação. Desconhecem-se estudos anteriores nacionais e internacionais que tenham comprovado esta segurança por meio de análises envolvendo testes estatísticos.

Tanto na avaliação da confiabilidade externa quanto na avaliação da confiabilidade interna do STM, ocorreram casos de *overtriage* e de *undertriage*, sendo mais frequentes os casos de *overtriage*. Resultado semelhante foi encontrado em estudo de meta-análise, em que a frequência de *overtriage* utilizando o STM foi de 46,65 e a de *undertriage* foi de 12,86%⁽⁵⁾. O STM é útil na triagem de pacientes em serviços de urgência, mas

a classificação de pacientes para níveis acima ou abaixo do real ainda ocorre, sendo mais frequentes os casos de *overtriage*⁽¹³⁾. A triagem de pacientes acima do nível de prioridade correto pode levar ao uso desnecessário de recursos em serviços de urgência e emergência⁽¹⁴⁾. Já uma subestimação do nível de risco pode aumentar o risco de consequências adversas aos pacientes, como atraso no atendimento e não utilização de recursos necessários à sua gravidade em tempo adequado⁽¹⁵⁾.

Os valores de kappa encontrados apontam que a confiabilidade externa e interna do STM variou de moderada a substancial, com valores ligeiramente mais elevados de kappa para a concordância intraobservadores (Tabela 3). Estudo de meta-análise conduzido para avaliar a confiabilidade do STM encontrou que a concordância global do STM foi substancial ($k=0,75$). No entanto, os autores ressaltaram que nos estudos avaliados foi utilizado o kappa ponderado, que superestima a confiabilidade. Sendo assim, é mais prudente considerar que a confiabilidade atual do STM é moderada. Nesse estudo, a concordância intraobservadores também foi maior e quase perfeita do que a concordância interobservadores. A confiabilidade entre enfermeiros foi substancial ($k=0,78$) e entre enfermeiros e especialista *expert* foi quase perfeita ($k=0,86$). A concordância foi quase perfeita nos estudos que a mensuraram pela avaliação de pacientes na prática clínica ($k=0,86$) e substancial nos estudos que avaliaram a concordância pela avaliação de casos clínicos ($k=0,76$)⁽⁷⁾.

Em estudo de revisão integrativa da literatura, a concordância entre enfermeiros e o padrão ouro utilizando o STM variou de moderada a quase perfeita ($k: 0,40 - 0,81$) e a concordância intraobservadores variou de substancial a quase perfeita ($k: 0,65 - 0,84$)⁽⁸⁾. Observou-se larga variação na confiabilidade entre observadores utilizando o STM, com valores de kappa entre 0,31 (concordância pobre) e 0,81 (concordância quase perfeita), com prevalência de estudos que apontam para uma concordância de boa a muito boa. A variação observada nos valores de kappa foi atribuída às diferenças nas populações estudadas, entre os avaliadores que utilizam o STM, e no modo como o STM é aplicado nos locais estudados⁽¹⁴⁾.

Estudo realizado na Alemanha mostrou que a concordância entre enfermeiros para a versão alemã do STM foi quase perfeita (Kappa = 0,95)⁽¹⁶⁾. Esta é a maior concordância encontrada em todos os estudos disponíveis na literatura que avaliaram a confiabilidade do STM. Cabe ressaltar que, diferentemente da versão portuguesa do STM utilizada no Brasil, a versão alemã foi submetida a processo de adaptação cultural que resultou em alterações de linguagem nos fluxogramas de apresentação e na definição dos discriminadores. Assim,

sugere-se a realização de estudo que trate da adaptação cultural e validação do STM para uso no Brasil, de modo a aumentar a confiabilidade do protocolo.

O julgamento intuitivo e reflexivo, componentes que envolvem a tomada de decisão na triagem, é fortemente influenciado pela experiência profissional do enfermeiro⁽²⁾. Corroborando com esta teoria, neste estudo, as variáveis tempo de experiência profissional como enfermeiro, experiência como enfermeiro em serviços de urgência e emergência e experiência como enfermeiro na classificação de risco foram associadas à confiabilidade externa e interna do STM. De modo geral, enfermeiros que possuem entre um e cinco anos e entre cinco e dez anos de experiência obtiveram maiores níveis de concordância entre si e com o padrão ouro.

No Brasil, embora haja recomendação informal de que o enfermeiro deva ter experiência prévia em serviços de urgência para atuar na classificação de risco, esta não é uma exigência regulamentada pelo conselho da classe. Na Itália, é exigido que os enfermeiros tenham experiência mínima de seis meses para realizarem a triagem de pacientes em serviços de urgência⁽¹⁷⁾.

A experiência profissional tem sido apontada na literatura como fator que influencia a tomada de decisão do enfermeiro na triagem. Os enfermeiros utilizam o conhecimento e experiências anteriores para fazer inferências e triar novos casos^(2,5,18). A correta classificação de risco depende do treinamento e da experiência do enfermeiro na aplicação do STM⁽¹⁵⁾.

Cabe ressaltar que a literatura disponível não é conclusiva sobre a quantidade de tempo de experiência necessária para garantir a competência do enfermeiro da triagem. Este é o primeiro estudo brasileiro que investigou e encontrou associação entre a experiência profissional e a confiabilidade externa e interna do STM. Em estudo que avaliou a experiência do enfermeiro na triagem e a habilidade no uso correto do STM, os valores de Kappa entre os enfermeiros e o padrão ouro foi maior quanto à experiência do enfermeiro com o STM, mas não foi encontrada diferença significativa entre as categorias de tempo de experiência na triagem analisadas⁽⁹⁾. A combinação entre o uso do STM, a experiência do enfermeiro na avaliação de pacientes criticamente enfermos e fatores organizacionais representaram 65% da segurança na triagem correta do paciente. A experiência dos enfermeiros contribuiu para maior segurança do paciente do que o próprio sistema de triagem utilizado, que não pode substituir completamente as habilidades clínicas que o enfermeiro experiente desenvolve ao longo dos anos na profissão⁽¹⁹⁾.

Assim, recomenda-se a realização de novas pesquisas que visem verificar a associação entre o tempo de experiência profissional como enfermeiro, de

experiência em serviços de urgência e emergência e de experiência na classificação de risco com a habilidade na utilização do STM na prática clínica. Também faz-se necessário investigar a relação entre o tempo de experiência no uso do STM e a habilidade na utilização correta do mesmo na prática clínica.

Outros estudos têm mostrado que o STM extrapola seu objetivo central que é o de estabelecer prioridade para tratamento imediato^(13,20-21).

Estudo realizado mostrou que a sensibilidade do STM para identificar pacientes que evoluíram para óbito ou que necessitaram de internação em unidades de terapia intensiva variou de 0,80 a 0,86 em adultos. Já a especificidade para mensurar os mesmos desfechos variou de 0,84 a 0,91⁽¹³⁾. Pacientes classificados como vermelho, laranja ou amarelo apresentaram risco 4,86 vezes maior de internação, 5,58 vezes maior de morrer e maior necessidade de realizar eletrocardiograma e exames laboratoriais quando comparados aos classificados como verde ou azul⁽²⁰⁾. Na Alemanha, estudo mostrou que quanto maior o nível de risco do paciente, maior o tempo de enfermagem despendido para assistência ao paciente, mostrando que o STM pode ser utilizado para subsidiar o cálculo do dimensionamento de pessoal de enfermagem em serviços de urgência e emergência⁽²¹⁾. Assim, percebe-se que o STM é capaz de prever, de forma precoce, as necessidades de cuidados nos serviços de urgência e emergência, bem como de direcionar a organização do cuidado e a gestão da clínica.

Uma limitação do estudo foi a utilização de casos clínicos em *software*, em detrimento da avaliação de pacientes na prática clínica ou em laboratório de simulação realística com pacientes padronizados. Entretanto, o atendimento a pacientes com instabilidade clínica requer avaliação clínica rápida e intervenções em tempos determinados de acordo com o nível de prioridade clínica, dificultando, assim, a apreensão dos dados de interesse do estudo. Sabe-se que a avaliação utilizando pacientes padronizados proporcionaria maior realismo e aproximação com o que ocorre na prática clínica. Entretanto, a coleta de dados utilizando a avaliação de pacientes na prática clínica ou com pacientes padronizados restringiria o número de enfermeiros participantes da pesquisa e os reduziria a uma amostra local. Assim, optou-se por trabalhar com uma amostra de enfermeiros maior, na tentativa de permitir a generalização dos resultados, e formatou-se o instrumento de coleta dos dados, de forma a refletir o processo de tomada de decisão do enfermeiro utilizando o STM, aproximando ao máximo do cenário da prática clínica. Não obstante, sugere-se a realização de estudos prospectivos utilizando a metodologia de simulação com pacientes padronizados no sentido de elucidar melhor

os motivos que levam às discordâncias na classificação, bem como os elementos que influenciaram a tomada de decisão do enfermeiro.

Por fim, salienta-se que a atuação do enfermeiro na classificação de risco é complexa e que a tomada de decisão envolve, além dos elementos cognitivos, aspectos como a gestão do fluxo de atendimento e a organização da rede assistencial, que extrapolam o poder de governabilidade do enfermeiro. No entanto, o uso de um instrumento confiável é importante para a segurança do enfermeiro, tendo em vista que o protocolo é o seu sistema de apoio na tomada de decisão.

Conclusão

A confiabilidade externa e interna do STM variou de moderada a substancial, com valores de Kappa, respectivamente, entre 0,55 e 0,72 ($p < 0,001$), e entre 0,57 e 0,78 ($p < 0,05$). Este foi o primeiro estudo que avaliou a confiabilidade externa e interna do STM para indicação do fluxograma e do discriminador de classificação. A escolha correta do fluxograma explicou 16% da variação na indicação correta do nível de risco ($R^2: 0,16$; $p < 0,0001$), enquanto que a escolha correta do discriminador explicou 77% da escolha correta do nível de risco ($R^2 = 0,77$; $p < 0,0001$), comprovando a segurança do STM na determinação do nível de prioridade a partir de diferentes fluxogramas, já que algumas queixas apresentadas pelos pacientes podem conduzir à escolha de diferentes fluxogramas de apresentação.

Tanto na avaliação da confiabilidade externa quanto interna do STM ocorreram casos de *overtriage* e de *undertriage*, e as discordâncias ocorreram, respectivamente, para um nível acima e para um nível abaixo do nível de risco estabelecido como correto pelo padrão ouro. Recomenda-se a realização de estudos futuros que busquem compreender os motivos das discordâncias, de modo a traçar estratégias direcionadas ao aumento da confiabilidade da avaliação do enfermeiro no uso do STM.

Apenas as variáveis tempo de experiência profissional como enfermeiro, de experiência como enfermeiro em serviços de urgência e emergência e de experiência na classificação de risco foram associadas à confiabilidade externa e interna do STM. De modo geral, enfermeiros que possuem entre um e cinco anos e entre cinco e dez anos de experiência obtiveram maiores níveis de concordância entre si e com o padrão ouro. Estes achados apontam que a inserção do enfermeiro na prática clínica e a experiência prévia em serviços de urgência e na classificação de risco são importantes para a confiabilidade externa e interna do STM. No entanto, tendo em vista a escassez de literatura disponível que

avaliar a associação entre a experiência profissional e a confiabilidade externa e interna do STM, é precoce estabelecer relação de causalidade entre a experiência profissional e a acurácia do enfermeiro na utilização do STM. Assim, recomenda-se a realização de novas pesquisas que visem consolidar os achados deste estudo.

Agradecimento

Ao Grupo Brasileiro de Classificação de Risco, pela parceria e incentivo à pesquisa.

Referências

1. Standing M. Clinical judgment and decision-making in nursing – nine modes of practice in a revised cognitive continuum. *J Adv Nurs*. 2008;62 (1):124-34. doi: <http://dx.doi.org/10.1111/j.1365-2648.2007.04583.x>
2. Smith A. Using a theory to understand triage decision making. *Int Emerg Nurs*. 2013;21:113-7. doi: <http://dx.doi.org/10.1016/j.ienj.2012.03.003>
3. Farrokhnia N, Castrén M, Ehrenberg A, Lind L, Oredsson S, Jonsson H, et al. Emergency Department Triage Scales and Their Components: a systematic review of the scientific evidence. *Scand J Trauma Resusc Emerg Med*. 2015;19(42):2-13. doi: <http://dx.doi.org/10.1186/1757-7241-19-42>
4. Mackway-Jones K, Marsden J, Windle J. *Emergency triage*. 3th ed. Manchester: BMJ Books; 2014. doi: <http://dx.doi.org/10.1002/9781118299029>
5. Stanfield LM. Clinical decision making in triage: an integrative review. *J Emerg Nurs*. 2015;41(5):396-403. doi: <http://dx.doi.org/10.1016/j.jen.2015.02.003>
6. DeVon HA, Block ME, Moyle-Wright P, Ernst DM, Hayden SJ, Lazzara DJ, et al. A psychometric toolbox for testing validity and reliability. *J Nurs Scholarsh*. [Internet]. 2007 May 23 [cited Feb 20, 2017];39(2):155-64. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/17535316>
7. Mirhaghi A, Mazlom R, Heydari A, Ebrahimi M. The Reliability of the Manchester Triage System (MTS): A Meta-analysis. *J Evid Based Med*. 2016;Sep 9. doi: <http://dx.doi.org/10.1111/jebm.12231>
8. Souza CC, Araújo FA, Chianca TC. Scientific literature on the reliability and validity of the Manchester triage system (MTS) Protocol: a integrative literature review. *Rev Esc Enferm USP*. 2015;49(1):144-51. doi: <http://dx.doi.org/10.1590/S0080-623420150000100019>
9. Storm-Versloot MN, Ubbink DT, Chin a Choi V, Luitse JS. Observer agreement of the Manchester Triage System and the Emergency Severity Index: a simulation study. *Emerg Med J*. 2009;26(8):556-60. doi: <http://dx.doi.org/10.1136/emj.2008.059378>

10. Alexandre NMC, Gallasch CH, Lima MHM, Rodrigues RCM. Reliability in the development and evaluation of measurement instruments in the health field. *Rev Eletr Enferm*. 2013;15(3):802-9. doi: <http://dx.doi.org/10.5216/ree.v15i3.20776>
11. Matos FGOA, Cruz DALM. Development of an instrument to evaluate diagnosis accuracy. *Rev Esc Enferm USP*. 2009;43:1088-97. doi: <http://dx.doi.org/10.1590/S0080-62342009000500013>
12. Cohen JA. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. 1960 Apr 1 [cited Feb 2, 2017];20(1):37-46. Available from: <http://journals.sagepub.com/doi/abs/10.1177/001316446002000104?journalCode=epma>
13. Zachariasse JM, Seiger N, Rood PPM, Alves CF, Freitas P, Smit FJ, et al. Validity of the Manchester Triage System in emergency care: a prospective observational study. *PLoS One*. 2017; 12(2):e0170811. doi: <http://dx.doi.org/10.1371/journal.pone.0170811>
14. Parenti N, Reggiani ML, Iannone P, Percudani D, Dowding D. A systematic review on the validity and reliability of an emergency department triage scale, the Manchester Triage System. *Int J Nurs Stud*. 2014;51(7):1062-9. doi: <http://dx.doi.org/10.1016/j.ijnurstu.2014.01.013>
15. Coutinho AAP, Cecílio LCO, Mota JAC. Risk classification in emergency care departments: A discussion of the literature on the Manchester Triage System. *Rev Med Minas Gerais*. [Internet].2012 Jun [cited Mar 13, 2017];22(2):188-98. Available from: <http://rmmg.org/artigo/detalhes/101>
16. Gräff I, Goldschmidt B, Glien P, Bogdanow M, Fimmers R, Hoeft A, et al. The German Version of the Manchester Triage System and its quality criteria: first assessment of validity and reliability. *PLoS ONE*. 2014;9(2):e88995. doi: <http://dx.doi.org/10.1371/journal.pone.0088995>
17. Palma E, Antonaci D, Colì A, Cicolini G. Analysis of emergency medical services triage and dispatch errors by registered nurses in Italy. *J Emerg Nurs*. 2014;40(5):476-83. doi: <http://dx.doi.org/10.1016/j.jen.2014.02.009>
18. Clarke DE, Boyce-Gaudreau K, Sanderson A, Baker JA. ED triage decision-making with mental health presentations: a "think aloud" study. *J Emerg Nurs*. 2015;41(6):496-502. doi: <http://dx.doi.org/10.1016/j.jen.2015.04.016>
19. Forsman B, Forsgren S, Carlström ED. Nurses working with Manchester triage – The impact of experience on patient security. *Australas Emerg Nurs J*. 2012;15(2):100-7. doi: <http://dx.doi.org/10.1016/j.aenj.2012.02.001>
20. Santos AP, Freitas P, Martins HMG. Manchester Triage System version II and resource utilisation in the emergency department. *Emerg Med J*. 2014. 31:148-52. doi: <http://dx.doi.org/10.1136/emmermed-2012-201782>
21. Gräff I, Goldschmidt B, Glien P, Klockner S, Erdfelder F, Schiefer JL, et al. Nurse Staffing Calculation in the Emergency Department - Performance-Oriented Calculation Based on the Manchester Triage System at the University Hospital Bonn. *PLoS ONE*. 2016;11(5):e0154344. doi: <http://dx.doi.org/10.1371/journal.pone.0154344>

Recebido: 27.04.2017

Aceito: 09.01.2018

Correspondência:

Cristiane Chaves de Souza
 Universidade Federal de Viçosa. Departamento de Medicina e Enfermagem
 Av. Peter Henry Rolfs, s/n
 Bairro: Campus Universitário
 CEP: 36570-000, Viçosa, MG, Brasil
 E-mail: souzac.cris@gmail.com

Copyright © 2018 Revista Latino-Americana de Enfermagem

Este é um artigo de acesso aberto distribuído sob os termos da Licença Creative Commons CC BY.

Esta licença permite que outros distribuam, remixem, adaptem e criem a partir do seu trabalho, mesmo para fins comerciais, desde que lhe atribuam o devido crédito pela criação original. É a licença mais flexível de todas as licenças disponíveis. É recomendada para maximizar a disseminação e uso dos materiais licenciados.