

# Testes de rastreamento x testes de diagnóstico: atualidades no contexto da atuação fonoaudiológica\*\*\*

## Screening versus diagnostic tests: an update in the speech, language and hearing pathology practice

Bárbara Niegia Garcia de Goulart\*  
Brasília Maria Chiari\*\*

\* Fonoaudióloga. Doutora em Ciências pela Universidade Federal de São Paulo - Escola Paulista de Medicina. Professora Titular da Feevale. Endereço para correspondência: Rua São Manoel, 2107/302 - Porto Alegre - RS - CEP 90620-110 (bgoulart@via-rs.net).

\*\*Fonoaudióloga. Professora Titular do Departamento de Fonoaudiologia da Universidade Federal de São Paulo - Escola Paulista de Medicina.

\*\*\*Trabalho Realizado no Departamento de Fonoaudiologia da Universidade Federal de São Paulo.

Artigo de Atualização

Artigo Submetido a Avaliação por Pares

Conflito de Interesse: não

Recebido em 10.03.2006.  
Revisado em 03.06.06; 15.03.2007;  
25.05.2007.  
Aceito para Publicação em 25.05.2007.

### Abstract

**Background:** evaluation instruments, properties, selection indicators, application and validation of screening and diagnostic tests. **Aim:** to present some concepts concerning screening and diagnostic tests and their application according to a specific purpose. To present a few practical examples of the application of these instruments related to human communication, as well as to present validation criteria of tests in the population and criteria used for the rational selection of screening or diagnostic instruments in health programs and health services based on epidemiological concepts indexed in Scielo, Lilacs or Medline up to January 2007. **Conclusion:** diagnostic instruments differ from screening instruments in their objectives and eligibility criteria. Sensibility and specificity are two important indicators to be considered when choosing an instrument for screening or diagnosis. Reproducibility, time required to complete the evaluation and previous preparation of patients, if needed, are also indicators to be considered when choosing an instrument. Publication and information exchange regarding the properties of evaluation instruments, used for diagnosis or screening, related to the Speech, Language and Hearing Sciences must be systematically stimulated. Besides that, improving the knowledge about methodologies and evaluation instruments under different perspectives contribute to the better use of human and financial resources. Furthermore, the elaboration of studies that promote the correct validation of screening and diagnostic instruments used in human communication disorders contributes to the increase in knowledge in the field of Speech, Language and Hearing Sciences and, indirectly, to the acknowledgement of this science, based on technical-scientific evidence, in health promotion.

**Key Words:** Diagnostic; Screening; Sensitivity; Specificity; Evaluation.

### Resumo

**Tema:** instrumentos diagnósticos, propriedades e uso de indicadores para seleção, aplicação e validação de instrumentos de diagnóstico e de rastreamento. **Objetivo:** apresentar conceitos ligados aos instrumentos de avaliação e sua aplicação de acordo com o objetivo, seja rastreamento ou diagnóstico. Também são apresentados alguns exemplos práticos de aplicação de instrumentos de avaliação ligados à comunicação humana, bem como critérios de validação de testes na população e critérios utilizados para sua escolha e aplicação racional em serviços e programas de saúde a partir de pressupostos epidemiológicos pesquisados em artigos indexados nas bases de dados Scielo, Lilacs ou Medline até janeiro de 2007. **Conclusão:** instrumentos de avaliação e diagnóstico clínico diferem dos instrumentos de rastreamento em relação a seus objetivos e critérios de elegibilidade. São indicativos da precisão de um instrumento de avaliação, seja para rastreamento ou diagnóstico, a sensibilidade e a especificidade de tal instrumento. Questões como reprodutibilidade, tempo para realização do teste ou exame e preparação prévia do paciente também devem ser considerados quando da seleção de instrumentos de avaliação clínica. O conhecimento e disseminação de informações ligadas às propriedades dos instrumentos de avaliação ligados a Fonoaudiologia devem ser incentivados sistematicamente. Além disso, a ampliação da gama de conhecimentos acerca das diferentes perspectivas ligadas às metodologias e instrumentos diagnósticos contribuem com a melhor racionalização de recursos humanos e financeiros. A elaboração de estudos que promovam a validação dos instrumentos correntemente utilizados para rastreamento e diagnóstico dos distúrbios da comunicação humana colabora com o avanço do conhecimento ligado a Fonoaudiologia e, indiretamente, para com o reconhecimento da ciência fonoaudiológica, baseada em evidências técnico-científicas, na promoção da saúde.

**Palavras-Chave:** Diagnóstico; Rastreamento; Sensibilidade; Especificidade; Avaliação.

Referenciar este material como:



GOULART, B. N. G.; CHIARI, B. M. Testes de rastreamento x testes de diagnóstico: atualidades no contexto da atuação fonoaudiológica. *Pró-Fono Revista de Atualização Científica*, Barueri (SP), v. 19, n. 2, p. 223-232, abr.-jun. 2007.

## Introdução

Aliadas ao imenso avanço tecnológico, principalmente das últimas décadas, as biociências vêm evoluindo muito nos últimos anos.

Neste processo, a Fonoaudiologia foi reconhecida como profissão e a pesquisa ligada à comunicação humana desenvolveu-se progressivamente, contribuindo com grande parte dos conhecimentos hoje utilizados. Na atualidade, a preocupação com estudos que demonstrem evidências científicas mais fortes, trazendo mais subsídios ligados aos impactos da atuação fonoaudiológica na prevenção e reabilitação de questões ligadas à comunicação humana ganha cada vez mais espaço na agenda da pesquisa fonoaudiológica a fim de que seja possível ganhar e manter espaço para atuação do fonoaudiólogo nas equipes de saúde.

Neste sentido, é importante que se ampliem as discussões sobre saberes e práticas já difundidas em outras áreas do conhecimento para que estes contribuam ainda mais com o avanço dos conhecimentos ligados à avaliação, ao diagnóstico e ao tratamento da comunicação humana e de seus distúrbios.

Uma reflexão que se faz necessária é a de que, constantemente, o profissional de saúde trabalha com algum tipo de medida, seja durante o processo diagnóstico, seja durante o processo terapêutico. Exemplos são citados nas publicações de Nobre et al. (2004) e Gomes (2005), pois, quando se pesquisa a história clínica do paciente e de sua queixa medimos o tempo decorrido desde o início dos sinais e sintomas, bem como sua variação e intensidade neste período, entre outros. Além disso, a tomada dos tempos fonatórios, medidas faciais, acompanhamento do tempo decorrido em cada estágio da deglutição, medição da acuidade auditiva e avaliação das alterações de fala são exemplos práticos do uso de medidas, testes e instrumentos na clínica fonoaudiológica.

Tais dados contribuem para o diagnóstico fonoaudiológico e auxiliam no planejamento terapêutico, sendo que, a partir de algumas abordagens conceituais, o diagnóstico é condição primordial para que o planejamento terapêutico seja feito e tenha mais chances de êxito.

Outro exemplo, descrito por Oliveira et al. (2002), nos remete ao diagnóstico de surdez. Os autores afirmam que hoje em dia é consensual que a simples observação médica e a suspeita familiar não são suficientes para a detecção de surdez no primeiro ano de vida. Há que contar com avaliações

mais precisas, utilizando instrumentos padronizados, tais como as emissões otoacústicas e o BERA (*Brain Electrical Response Audiometry*), entre outros.

Cabe ressaltar que o conhecimento mais apurado das propriedades dos instrumentos de avaliação, seja de diagnóstico ou rastreamento, exames ou avaliações clínicas, está igualmente ligado à racionalização dos recursos tecnológicos, cada vez mais disponíveis em quantidade no mercado e nem sempre com os parâmetros que contribuem para o processo avaliativo devidamente ponderados (Krauss-Silva, 2004; Potvin, 2005).

As decisões sobre avaliações complementares e uso de exames que variam da baixa até a alta complexidade devem estar atreladas, entre outros, às características de cada teste e/ou procedimento e a sua capacidade de contribuir para o diagnóstico, ou mesmo, trazer mais subsídios para o planejamento terapêutico.

Tomando por exemplo a avaliação da fala de uma criança, a fim de determinar se apresenta desvio fonético-fonológico. Sabe-se que tal avaliação pode ser feita a partir da coleta da fala espontânea, utilizando figuras que representem todos os fonemas do Português ou mesmo a partir da emissão dos vocábulos pelo avaliador para posterior repetição pela criança avaliada (Castro e Wertzner, 2006; Goulart e Ferreira, 2006).

Na atualidade, contamos com diversos instrumentos no Brasil para o rastreamento e/ou diagnóstico deste e de outros distúrbios da comunicação humana, mas é essencial que estejam validados entre sujeitos com e sem as alterações que se pretende diagnosticar, a fim de avaliar com maior precisão as potencialidades do instrumento utilizado. A tradução das características dos testes e/ou exames de diagnóstico e/ou rastreamento pode ser caracterizada pelo conhecimento da sensibilidade, especificidade e valores preditivos de cada método, teste ou exame utilizado.

Este artigo objetiva apresentar alguns conceitos, reflexões e aplicações de testes de rastreamento (também denominados *screening*) e instrumentos para diagnóstico no âmbito dos distúrbios da comunicação humana. Além disso, abordaremos algumas questões pertinentes quanto à validação de testes, seja de rastreamento ou diagnóstico, sob a ótica da Epidemiologia.

Estas questões serão discutidas a partir da literatura publicada até fevereiro de 2007 nas bases de dados Scielo, Lilacs e/ou Medline. Para a busca

do referencial teórico foram utilizados os termos: rastreamento, rastreio, *screening*, *screening tests*, *screening test*, *screening procedures*, *screening method*, teste diagnóstico, diagnóstico, teste estandardizado, testes de rastreamento, validação de testes, avaliação, sensibilidade, especificidade, validade, reprodutibilidade dos resultados, avaliação da tecnologia biomédica, promoção e prevenção em saúde, promoção da saúde e avaliação de linguagem.

#### Características e propriedades dos instrumentos de avaliação

Costa et al. (2004), Jeckel et al. (2005a) e Hochman et al. (2005) citam dois objetivos distintos que cabe ressaltar quando da coleta e análise de dados a partir de instrumento de avaliação, seja em atividades clínicas ou de pesquisa: a acurácia e a precisão. A acurácia se refere à capacidade de uma medida ser correta na média. Quando a medida não é acurada, é viciada. A precisão, também conhecida como reprodutibilidade ou confiabilidade, é a capacidade de a medida dar o mesmo resultado (ou um resultado muito semelhante) nas medições repetidas de um mesmo fato. Ambas são qualidades importantes, caso uma destas esteja ausente, os dados tornam-se inúteis, pois não são, necessariamente, fidedignos.

Newman et al. (2003), Nobre et al. (2004) e Potvin (2005) referem que o acesso aos dados referentes a reprodutibilidade, a acurácia, o grau de importância do teste para tomada de decisão, custos, riscos e a aceitabilidade do instrumento e contribuição do teste na melhora do agravo/doença ou na produção de efeitos adversos também devem ser considerados quando da decisão sobre o uso destes.

Tais conceitos estão intimamente relacionados à demanda por uma atuação baseada em evidências, conceito introduzido nas práticas clínicas na América do Norte desde a década de 70, principalmente na Medicina, e que na atualidade vem tendo destaque nas discussões ligadas às práticas clínicas fonoaudiológicas no Brasil (Goulart, 2002b; Goulart e Chiari, 2006).

Desta forma, quando considerados os achados de estudos clínicos ou de base populacional para a tomada de decisão em ações coletivas para promoção da saúde ou área clínica, há de se considerar toda a situação que envolveu a pesquisa consultada, ou seja, a seleção dos participantes, instrumentos utilizados e treinamento ou habilidades requeridas pelos responsáveis pela

aplicação dos protocolos estudados (Potvin, 2005; Goulart, 2002b). Também é imperativo ressaltar que somente podem ser extrapolados dados de estudos para populações semelhantes àquelas pesquisadas e, quando há seleção demasiada dos indivíduos que irão compor as amostras estudadas, com inúmeros critérios de exclusão, estes sujeitos podem não representar a população em geral.

Esses fatores podem fazer com que testes, instrumentos, métodos de diagnóstico ou rastreamento que demonstram resultados satisfatórios em ambiente de pesquisa podem não apresentar respostas semelhantes em outras oportunidades.

O grau de importância do instrumento utilizado para a tomada de decisão em relação à conduta terapêutica, incluindo-se a relação de custo da aplicação do instrumento, os riscos ao qual o avaliador e o paciente estarão submetidos por conta do uso de tal instrumento e a aceitabilidade do procedimento pelo paciente (e seus familiares, quando for o caso) também devem ser considerados quando da indicação e uso de instrumentos para rastreamento ou diagnóstico. Estes critérios são largamente descritos e discutidos pela comunidade científica desde a década de 70 e mais recentemente juntam-se nestas discussões gestores do setor saúde, usuários dos serviços de saúde e os profissionais que utilizam tais instrumentos em sua rotina (McKinlay e Marceau, 2002; Zuckerman, 2003; Georg et al., 2005).

Krauss-Silva (2004) refere que a cobertura de tecnologias e procedimentos, assim como os processos de elaboração de diretrizes clínicas, planejamento, gerência e avaliação de serviços e programas por parte de governos dos países desenvolvidos cada vez mais dependem de avaliações tecnológicas formais. Em relação ao Brasil, a autora destaca que as dificuldades operacionais para a elaboração e utilização de avaliações tecnológicas dependem da adequação e disponibilização de dados e da formação de pesquisadores e tomadores de decisão nesse campo.

Krauss-Silva (2004) utiliza o conceito de avaliação de tecnologias em saúde (ATS) para designar os métodos utilizados para seleção dentre as alternativas existentes, sejam tecnologias individuais ou conjuntos tecnológicos, para lidar com uma mesma necessidade/problema de saúde. Refere que também são úteis, embora com mais limitações, para selecionar conjuntos tecnológicos na forma de programas (um programa pode corresponder a vários conjuntos de tecnologias)

destinados a diferentes problemas de saúde, como etapa técnica do processo de seleção de prioridades, bem como para o planejamento e a gerência de serviços de saúde enquanto subsídio crítico à elaboração de diretrizes clínicas baseadas em evidências.

#### Características e aplicação dos instrumentos de rastreamento e de diagnóstico

O rastreamento é caracterizado como a identificação de uma doença ou fator de risco não reconhecido, por meio da história clínica, do exame físico, de um exame laboratorial ou de outro procedimento que possa ser aplicado rapidamente (Bhopal, 2002; Newman et al., 2003; Toscano, 2004). Os testes de rastreamento separam as pessoas que estão aparentemente bem, mas que apresentam uma doença ou um fator de risco para uma doença, daquelas que não os apresentam.

O rastreamento de pacientes na fase pré-clínica da doença já é uma constante na prática médica e, atualmente, o rastreamento para detecção de agravos ou doenças diversas vem sendo ampliado para a população em geral, sem queixas (Georg et al., 2005; Fletcher e Fletcher, 2006).

Entretanto, é importante ressaltar que as recomendações atuais em favor da prevenção se dão no sentido de submeter a população à testes específicos de acordo com a prevalência de distúrbios característicos a determinadas faixas etárias, sexo e características clínicas (Bhopal, 2002; Newman et al., 2003; Fletcher e Fletcher 2006).

Coutinho et al. (2003), Toscano (2004) e Georg et al. (2005) afirmam que a aplicação de testes de rastreamento deve ser norteada pela premissa de que o diagnóstico precoce efetivamente possa contribuir para a melhora do prognóstico da doença. Estas questões são corroboradas nos estudos de Cardoso et al. (2004), Fiorindo et al. (2004) e Toscano (2004), ligados a diferentes especialidades relacionadas às ciências da saúde.

Toscano (2004) e Fletcher e Fletcher (2006) referem que um teste de rastreamento deve detectar o maior número de casos com o menor custo, além de não acarretar reações adversas ou efeitos colaterais. Os autores também afirmam que há menor grau de benefício agregado ao teste de rastreamento em resposta aos sintomas, se comparado à avaliação diagnóstica, devendo este ser mais seguro que os testes clínicos correntemente empregados. Isso se traduz em menor precisão em relação ao diagnóstico, visto que os instrumentos diagnósticos tendem a ser mais específicos.

Um teste ideal para rastreamento deveria levar apenas alguns minutos para ser realizado, além de requerer um mínimo de preparação prévia do paciente e não depender de agendamento especial, além de ser de baixo custo (Toscano, 2004; Fletcher e Fletcher, 2006). Deve ser buscado um teste de rastreamento barato, de fácil aplicabilidade e (que cause) desconfortos e/ou transtornos mínimos aos pacientes (Toscano, 2004; Jeckel et al., 2005b). Todos os autores citados ainda referem que os resultados de um teste de rastreamento devem ser válidos, confiáveis e reproduzíveis.

Em resumo, a validade de um instrumento para rastreamento ou diagnóstico é medida pela sua habilidade em fazer aquilo que se propõe, ou seja, categorizar adequadamente os indivíduos com sintomas "pré-clínicos" da doença como teste positivo e aqueles sujeitos sem sintomas pré-clínicos da doença como teste negativo. Estas relações entre doença e teste positivo, não-doença e teste positivo, doença e teste negativo e não-doença e teste negativo são expressas através de sensibilidade, especificidade e valores preditivos dos testes, tanto de rastreamento quanto de diagnósticos (Fletcher e Fletcher, 2006).

A acurácia do instrumento utilizado depende, não apenas da sensibilidade e da especificidade, mas, também, da prevalência da doença na população na qual o mesmo está sendo utilizado. Quanto mais rara a doença, mais específico deve ser o teste para ser útil, ou seja, para ser passível de detecção dos casos. Porém, se uma doença é freqüente (tomando-se como um exemplo atual o diabete *mellitus* e a perda auditiva ocorrida a partir da sexta década de vida), o teste deve ser muito sensível para que seja útil ao clínico, pois caso contrário, o resultado negativo expressará um resultado falso-negativo.

Um bom teste de rastreamento deve ter alta sensibilidade para não perder os casos da doença presentes na população testada, bem como alta especificidade a fim de reduzir o número de pessoas com resultados falso-positivos que necessitem de posterior investigação (Bhopal, 2002; Fletcher e Fletcher, 2006).

Os instrumentos para diagnóstico tendem a ser mais precisos e trazer maior detalhamento em relação ao agravo ou doença ao qual estão relacionados. Tanto quanto os instrumentos de rastreamento, os testes para diagnóstico carecem de validação para as populações às quais se destinam, incluindo sujeitos saudáveis com as mesmas características que os sujeitos acometidos por desordem ou agravo à saúde aos quais serão submetidos. Além disso, quando da tomada de decisão pela utilização de um instrumento para diagnóstico deve ser considerada sua relação custo-

benefício, sua capacidade de agregar informações que contribuam com o planejamento terapêutico e complementem, quando for o caso, o diagnóstico realizado a partir dos achados de anamnese e exame clínico.

Para que os instrumentos de diagnóstico efetivamente agreguem informações a outros dados e contribuam com o diagnóstico de um agravo ou doença é importante que sua especificidade seja alta, assim, raramente o teste será positivo na ausência de doença.

Krauss-Silva (2004) refere que os estudos internacionais frequentemente focalizam populações cujas características genéticas, demográficas e culturais diferem daquelas da maioria da população brasileira, o que é importante para avaliar a aplicação direta de tecnologias para a saúde na realidade brasileira. A autora destaca que tais diferenças podem modificar, de forma significativa, parâmetros importantes como a acurácia, a eficácia e a utilidade ou preferência de tais instrumentos de acordo com as diferentes realidades encontradas nas diversas regiões do Brasil. O comprometimento da validade das medidas relativas às propriedades dos instrumentos pode influenciar, inclusive, a adesão a estas tecnologias, bem como sua efetividade.

#### Caracterização e validação de instrumentos de rastreamento ou diagnóstico

Para que um instrumento de diagnóstico clínico ou de rastreamento possa ser utilizado correntemente de forma a produzir dados confiáveis, há que proceder sua validação.

Os instrumentos são validados de acordo com a população à qual se destinam (adultos, crianças, com ou sem deficiências), devendo testá-los para verificar sua sensibilidade e especificidade em sujeitos saudáveis e em sujeitos portadores do agravo/doença para a qual tais instrumentos se destinam a diagnosticar ou rastrear.

Além disso, estudos para validação de instrumentos de medida devem contar com reavaliação com um instrumento para mesmo fim já devidamente validado ou, no caso de impossibilidade de uso de um instrumento com estas características, através de avaliação clínica com profissional experiente.

Para estimar a sensibilidade de um instrumento de medida há que contar com um número de pessoas que apresentam o evento no qual estamos interessados (desordem, alteração, doença) e utilizá-lo neste grupo. A partir disso, é necessário verificar em quantos destes o instrumento detectou

o evento em estudo (resultado positivo), como descreve Gomes (2005).

Neste caso, a sensibilidade do instrumento é medida a partir da proporção de casos em que os resultados desse teste, quando positivos, concordam com os resultados de um teste considerado para referência e a especificidade é a proporção de casos em que o teste, quando negativos, concordam com os do teste considerado para referência (Gomes, 2005; Hochman et al., 2005; Fletcher e Fletcher, 2006).

Tomando-se como exemplo, temos o uso de um instrumento de rastreamento para desordens de fala em crianças a partir de sete anos (Teste de Rastreamento de Distúrbios Articulatorios de Fala - TERDAF - Goulart, 2002a), ao qual fossem submetidas 200 crianças com tais desordens. Se o teste detecta tal desordem em 163 crianças, a sensibilidade do teste será de 81,5%, ou seja, esta é a possibilidade de o TERDAF ser positivo em uma criança com desordem de fala.

A sensibilidade de um teste ou exame para diagnóstico ou rastreamento mede a capacidade deste em detectar algo que existe (Gomes, 2005; Hochman et al., 2005), ou sejam, no caso de um instrumento de diagnóstico sensível para surdez (levando-se em conta o conjunto equipamento e examinador) é aquele que, com alta probabilidade, detecta surdez em um paciente surdo.

Destá forma, um instrumento é considerado mais sensível quanto maior for a probabilidade de detectar um evento (desfecho, uma desordem, uma doença).

A especificidade, por sua vez, é caracterizada pela capacidade de um instrumento de medida não detectar eventos inexistentes (Gomes, 2005; Hochman et al., 2005). No caso de sujeitos normotensos, um instrumento altamente específico detectaria que a tensão arterial está dentro dos limites de normalidade. Tomando-se um exemplo ligado a comunicação humana, um instrumento suficientemente específico, não detectaria alterações vocais em um sujeito sem disфонia.

Em relação à especificidade, há que submeter um certo número de sujeitos que não apresentam a característica de interesse ao instrumento estudado para verificar quantas apresentarão resultado negativo para o evento estudado.

Mantendo-se o uso do exemplo referido anteriormente, se submetêssemos 220 crianças sem desordens de fala ao TERDAF e dentre essas o referido instrumento apresentasse resultado negativo em 99 delas, a especificidade do instrumento será de 45%. Isso significa que para

cada dez sujeitos sem desordem de fala, submetidos ao teste, aproximadamente quatro apresentariam resultado negativo.

Outra questão a ser considerada diz respeito ao ponto de corte (ou limiar diagnóstico do teste). Há que considerar qual será o grau (ou nível) de alteração detectada que positivar o teste. Esta escolha envolve uma decisão entre aumentar a sensibilidade à custa de redução da especificidade ou vice-versa. A maioria dos pesquisadores deve avaliar cuidadosamente a importância relativa da sensibilidade e especificidade do teste para estabelecer o ponto de transição diagnóstica mais adequada.

Como estratégia geral, quando a principal preocupação for evitar o resultado falso-positivo (por exemplo, o resultado do teste pode levar a indicação de realização de um procedimento arriscado para o paciente), então o ponto de corte deve objetivar o máximo de especificidade e se a preocupação maior é evitar resultado falso-negativo (o resultado do teste em paciente com suspeita de disfagia), o ponto de corte deve objetivar o máximo de sensibilidade (Fletcher e Fletcher, 2006). Os autores ainda referem que a curva ROC (*Receive Operator Characteristic Curve*) é a melhor maneira de estabelecer o ponto de corte, otimizando a sensibilidade e especificidade do teste diagnóstico.

Para construir a curva ROC o pesquisador deve selecionar vários pontos ou níveis de alteração do teste (graus de intensidade variados, por exemplo) e determinar a sensibilidade e especificidade em cada ponto. A partir disso, é construído um gráfico da sensibilidade em função da proporção dos resultados falsos-positivos. O teste ideal é aquele que alcança a extremidade mais superior e esquerda do gráfico. Uma das vantagens deste método é que as curvas de diferentes testes diagnósticos podem ser comparadas.

Newman et al. (2003) ressalta que quanto melhor o teste mais perto estará sua curva do canto superior esquerdo do gráfico. Os autores também destacam que os elementos necessários, para que um artigo de estudo do tipo diagnóstico ou de rastreamento seja considerado consistente devem ser passíveis de resposta três questões principais: a) existência de comparação cega e independente do teste diagnóstico com o "padrão ouro" usado no reconhecimento da doença. Os autores descrevem que os pacientes em estudo devem ser submetidos aos dois procedimentos diagnósticos, aquele sendo testado e o "padrão-ouro" (de referência), reconhecidos como provas de que o

paciente tem ou não tem uma determinada doença. Além disso, destacam que os resultados de um procedimento não podem ser conhecidos por quem aplica ou interpreta o teste que está sendo comparado, visto que esses dois elementos são fundamentais para que sejam afastados os vícios de interpretação, conscientes ou inconscientes, dos investigadores. O teste diagnóstico (em estudo) deve ser aplicado em uma amostra apropriada de pacientes semelhante àquela encontrada na prática clínica diária. Os pacientes em estudo devem apresentar os achados clínicos comumente presentes na doença em questão, bem como as características clínicas presentes em doenças que entram no diagnóstico diferencial. E, finalmente, destacam os autores que o teste diagnóstico deve ser validado em um segundo grupo independente de pacientes. O primeiro estudo (com o teste diagnóstico) serve para se fazer uma predição da sua acurácia. A confirmação dos níveis semelhantes de desempenho com a aplicação em uma segunda amostra, independente, de pacientes permite a estimativa real da acurácia do teste diagnóstico em estudo.

O teste utilizado como referência, quando possível, deve ser um instrumento bastante conhecido e, idealmente, o melhor de que dispomos no momento. Comumente é atribuído a este o termo padrão-ouro ou *gold standard*. Pode ser um exame do mesmo tipo daquele que está sendo estudado ou mesmo um instrumento de outra natureza. Por exemplo, há como comparar a sensibilidade e especificidade da otoemissão acústica com o BERA na detecção da surdez em neonatos.

Didaticamente há como sintetizar a comparação entre um instrumento em estudo e um teste padrão-ouro em uma Tabela 2x2 (ou Tabela de contingências), conforme apresentado na Tabela 1.

TABELA 1. Disposição dos dados na Tabela 2x2.

	Padrão-Ouro +	Padrão-Ouro -	TOTAL
teste em estudo +	a	b	a + b
teste em estudo -	c	d	c + d
TOTAL	a + c	b + d	N = a + b + c + d

A Tabela 2x2 é largamente utilizada em estudos epidemiológicos e suas possibilidades de uso com exemplos ligados à comunicação humana foram descritas por Goulart em (2002b).

Desta forma, (a+c) é o total de sujeitos com teste padrão (padrão-ouro) positivo, ou seja, inclui os sujeitos que sabidamente apresentam a característica em estudo. A soma (b+d) representa o grupo de pessoas que não apresentam a característica em estudo.

Na casela (a) estão representados os sujeitos que tanto no instrumento em estudo, quanto no padrão-ouro apresentam resultado positivo, ou seja, são os chamados "verdadeiros positivos". Na casela (b) estão alocados os sujeitos que não apresentam o evento estudado de acordo com o padrão-ouro e apresentam resultado positivo quando submetidos a avaliação pelo instrumento em estudo. O grupo de sujeitos dispostos na casela (b) é denominado "falso positivo".

Da mesma forma, são considerados "falsos negativos" os sujeitos dispostos na casela (c) e "verdadeiros negativos" os sujeitos dispostos na casela (d).

Como descreve Gomes (2005), a sensibilidade é calculada pela razão entre o número de verdadeiros positivos e o total de positivos de acordo com o teste padrão-ouro e a especificidade é caracterizada pela razão entre o número de verdadeiros negativos e o total de negativos do teste padrão-ouro (vide esquema abaixo):

$$\text{Sensibilidade} = a / (a+c) \quad \text{Especificidade} = d / (b+d)$$

A reprodutibilidade dos testes (instrumentos) diagnósticos e/ou de rastreamento também deve ser considerada. Esta propriedade está atrelada à possibilidade de que o teste seja aplicado por profissionais (ou não, de acordo com a sua natureza) e pacientes em momentos diversos possibilitando um mesmo resultado.

Uma baixa reprodutibilidade pode ser resultante de problemas com o próprio instrumento, tais como dificuldade técnica de aplicação do mesmo (procedimentos complexos ou complicados que dependem muito do examinador) ou mesmo alterações do teste a partir de interpretações variadas (por exemplo, piora do BERA, a partir da análise dos traçados ou mesmo análise de videonasofaringoscopia pré e pós fonoterapia como indicador de melhora do fechamento velofaríngeo).

Um artigo que disserta sobre a validação de instrumentos de avaliação deve trazer descrito o

grau de reprodutibilidade que pode ser esperado. Esta descrição é especialmente importante quando da análise dos resultados do instrumento, principalmente quando requerem algum grau de experiência, como nos casos de análise videolaringoscópica, videofluoroscópica e videonasofaringoscópica. Assim, é possível confirmar o grau de concordância entre os achados (ou análises) nos casos em que há divergências entre os responsáveis por interpretar tais achados. Deve ser avaliado o grau de reprodutibilidade intra e inter-avaliadores, conforme descrito no estudo realizado para a validação do *Tinnitus Handicap Inventory* (THI) para o português brasileiro por Ferreira et al. (2005).

Nos casos em que se observa que a reprodutibilidade de um teste em situação de pesquisa é boa, mas ocorre discordância inter-observadores, há que considerar em conjunto se a sensibilidade do instrumento é boa, a fim de que discrimine adequadamente aqueles com e sem o agravo ou doença em estudo. Ainda assim, este instrumento pode ser útil na atuação clínica, pois, se a reprodutibilidade de um teste é muito alta e a variação inter-observador é baixa, o teste continua a ser simples e confiável para aqueles que o interpretam (Jaeschke et al., 1995). Os autores também descrevem que se dois tipos de instrumento oferecem respostas muito próximas (ou muito relacionadas, mesmo tipo de resposta), a aplicação do segundo teste oferecerá muito pouca ou nenhuma contribuição diagnóstica. Esta questão deve ser considerada quando da necessidade constante de racionalização de recursos que vivenciamos na sociedade brasileira contemporânea, conforme discutido por Goulart (2003) e Nobre et al. (2004).

Citando como exemplo instrumentos diagnósticos de surdez, Oliveira et al. (2002) referem que o rastreamento para crianças em risco é também alvo de controvérsia quanto aos métodos a utilizar. Assim, enquanto alguns autores defendem a realização de emissões otoacústicas seguidas de BERA, outros advogam a utilização dos BERA como testes de primeira escolha. Os autores ainda referem que em qualquer das situações é necessária atenção às constantes evoluções do conhecimento e da técnica, pois novos esquemas e novos métodos, como os Produtos de Distorção e os BERA automáticos, podem brevemente assumir posições de destaque.

Outro exemplo intimamente ligado à atuação fonoaudiológica em relação à escolha de instrumentos de avaliação é descrita por Hage et al. (2004). Os autores citam em seu artigo

instrumentos com características e objetivos diferentes para avaliação do desenvolvimento infantil. Destacam que as escalas de desenvolvimento, as quais refletem os principais ganhos ao longo do desenvolvimento e têm o objetivo de determinar o nível evolutivo específico da criança. A partir destas, o nível de evolução da criança é obtido através de dados relatados sobre seu desenvolvimento e a partir da observação direta de seu comportamento. Algumas escalas avaliam especificamente o desenvolvimento da linguagem, como é o caso da *Early Language Milestone Scale* (ELM) e a *Reynell Developmental Language Scale* (RDLs) e outras avaliam vários aspectos do desenvolvimento, dentre estas os autores destacam a *Denver Developmental Screening Test* (DDST) e a Escala de Desenvolvimento Comportamental de Gesell e Amatruda (EDCGA).

Moreira e Ferreira-Junior (2004) realizaram estudo comparativo entre portadores de perda auditiva induzida por ruído (PAIR) e indivíduos com audição normal. Os autores constataram que no grupo pesquisado o IPRF (Índice Percentual de Reconhecimento de Fala) não se mostraram úteis no diagnóstico da PAIR.

Estudo publicado por Marteleto e Pedromônico (2005), no qual examinaram a validade de critério do Inventário de Comportamentos Autísticos (ICA) detectou que o Inventário de Comportamentos Autísticos identificou corretamente 81,6% das crianças com autismo pesquisadas. As autoras também descrevem que o Inventário de Comportamentos Autísticos (ICA) mostrou baixa sensibilidade (57,89%) e alta especificidade (94,73%) quando utilizados como ponto de corte 68 pontos. Quando a nota de corte foi diminuída para 49 pontos, a sensibilidade da escala aumentou para 92,1% e a especificidade se manteve alta (92,6%). Desta forma, as autoras sugerem que o ICA é um instrumento promissor para identificar crianças com autismo, especialmente com ponto de corte 49 pontos, tanto na clínica, como em contextos educacionais.

Em estudo publicado em 2005 por Ferreira et al., os autores descrevem detalhadamente a validação do THI para uso na população brasileira, além das aplicações e resultados encontrados na validação do instrumento na população pesquisada.

## Conclusão

A importância de conhecer as demandas fonoaudiológicas nos diferentes estratos da população (faixa etária, sexo, de acordo com patologias de base, entre outros) é essencial para que os esforços, recursos humanos e financeiros, sejam mais efetivos e traduzam em retorno para a saúde coletiva. Pois, como descrito por Krauss-Silva (2003) as incipiências crônicas das atividades e da aplicação da avaliação das tecnologias de saúde têm expressões bastante visíveis na política de financiamento e cobertura de procedimentos, sejam em relação à saúde pública ou no rol de procedimentos da saúde suplementar (planos e seguros de saúde). Então, para que a Fonoaudiologia amplie suas inserções nos serviços de saúde é necessário que cada vez mais todos os profissionais tenham conhecimentos e saibam utilizar as ferramentas disponíveis para a atuação baseada em evidências.

Mas, para que isso ocorra, os centros de estudos e universidades devem contribuir ainda mais fortemente na validação e ampliação de conhecimentos sobre as aplicações das tecnologias atualmente disponíveis na área da Comunicação Humana.

Além disso, a produção destes conhecimentos deve somar-se a sua ampla divulgação em associados aos nossos veículos de comunicação científica, nos eventos, encontros e periódicos, tanto para os fonoaudiólogos já formados, quanto para acadêmicos de graduação através da inclusão destes nos projetos pedagógicos dos cursos de graduação e programas de pós-graduação, como já ocorre em diversas áreas.

A importância de levar em conta as populações testadas para validação de instrumentos em relação à comunicação humana deve considerar os regionalismos em relação à avaliação de fala e linguagem, faixa etária e ocupação ou atividade profissional em relação à voz e questões ligadas a motricidade orofacial, entre outros. Desta forma, a descrição detalhada da população testada para validação do instrumento e critérios diagnósticos empregados a partir das respostas gerados através da aplicação do instrumento são informações essenciais para o profissional que o utiliza em suas práticas de rotina, ou mesmo nas pesquisas clínicas. A adequada validação de instrumentos e conhecimento de suas propriedades técnicas na

população brasileira também deve contribuir com a seleção de prioridades e delimitação do conteúdo dos programas de saúde pelos tomadores de decisão em diferentes níveis de gestão.

Do mesmo modo, conhecidos os dados epidemiológicos da população que demanda por atenção fonoaudiológica em suas diferentes formas e de acordo com as realidades regionais, são necessários estudos de eficácia, de análise de decisão e de análise de custo-efetividade, bem como revisões sistemáticas que congreguem estudos diversos, mas com metodologia semelhante. Tais estudos podem agregar evidências científicas ainda mais consistentes para o avanço da atuação fonoaudiológica para a população, a partir do avanço da ciência em relação à comunicação humana.

Em contrapartida, é necessário considerar que nem todos os problemas de saúde, mesmo ligados à comunicação humana, são passíveis de prevenção e que somente as opiniões de usuários dos serviços de fonoaudiologia não constituem

evidências suficientes, tampouco satisfatórias do ponto de vista metodológico-científico, para a tomada de decisão sobre as tecnologias e recursos a serem empregados na atenção fonoaudiológica.

Garber (2001) e Krauss-Silva (2004) citam alguns estudos sobre o uso irracional, do ponto de vista da população, de tecnologias e procedimentos, apontando larga utilização de tecnologias sem evidência de benefício e vice-versa, utilização de tecnologias fora das condições nas quais apresentam eficácia/accurácia significativa e grande variação no uso de tecnologias sem variação correspondente dos resultados. Também são descritos o grande fascínio exercidos por novas tecnologias, nem sempre melhores e quase sempre mais caras, além de comumente complementarem e sobreporerem às tecnologias já existentes.

É importante que o fonoaudiólogo esteja atento a estas questões e, sempre que necessário, busque subsídios que denotem evidências para contribuir com a decisão sobre o uso de ferramentas e instrumentos na sua prática.

## Referências Bibliográficas

BHOPAL, R. S. Natural history, spectrum, iceberg, population patterns and screening: interrelated concepts in the epidemiology of disease. In: \_\_\_\_\_. *Concepts of Epidemiology*. Oxford, Oxford University Press, 2002. 317 p. Disponível em: <<http://www.pitt.edu/~super1/lecture/lec18491/001.htm>>. Acesso em: 05 maio 2006.

CARDOSO, A. A.; MAGALHÃES, L. C.; AMORIM, R. H. C.; PAIXÃO, M. L.; MANCINI, M. L.; ROSSI, L. D. F. Validade preditiva do movement assessment of infants para crianças pré-termo brasileiras. *Arq. Neuro-Psiquiatr.*, São Paulo, v. 62, n. 4, p. 1052-1057, dez. 2004.

CASTRO, M. M.; WERTZNER, H. F. Estimulabilidade e tipos de erro de fala. *R. Soc. Bras. Fonoaudiol.*, São Paulo, v. 11, n. 1, p. 1-9, jan.-mar. 2006.

COSTA, R. M. S.; VELASCO, T. R.; COSTA, L. S.; RODRIGUES, M. L. V. Eficácia da associação de métodos diagnósticos no rastreamento do glaucoma. *Arq. Bras. Oftalmol.*, São Paulo, v. 67, n. 1, p. 37-41, jan.-fev. 2004.

COUTINHO, E. S. F.; HUF, G.; BLOCH, K. V. Ensaio clínicos pragmáticos: uma opção na construção de evidências em saúde. *Cad. Saúde Públ.*, Rio de Janeiro, v. 19, n. 4, p. 1189-1193, ago. 2003.

FERREIRA, P. É. A.; CUNHA, F.; ONISHI, E. T.; BRANCO-BARREIRO, F. C. A.; GANANÇA, F. F. Tinnitus Handicap Inventory: adaptação cultural para o Português Brasileiro. *Pró-Fono R. Atual. Cient.*, Barueri (SP), v. 17, n. 3, p. 303-310, set.-dez. 2005.

FLETCHER, R. H.; FLETCHER, S. W. Prevenção. In: FLETCHER, R.; FLETCHER, S. *Epidemiologia clínica: elementos essenciais*. 4. ed. Porto Alegre: Artes Médicas, 2006. p. 179-201.

GARBER, A. M. Evidence-based coverage policy. *Health Aff.*, Bethesda (USA), v. 20, n. 5, p. 62-82, sep.-oct. 2001.

GEORG, A. E.; DUNCAN, B. B.; TOSCANO, C. M.; SCHMIDT, M. I.; MENGUE, S.; DUARTE, C.; POLANCZYK, C. A.; GRUPO DE TRABALHO DE AVALIAÇÃO DA CAMPANHA NACIONAL DE DETECÇÃO DE DIABETES MELLITUS. Análise econômica de programa para rastreamento do diabetes mellitus no Brasil. *R. Saúde Públ.*, São Paulo, v. 39, n. 3, p. 452-460, jun. 2005.

GOMES, U. A. Características dos instrumentos de medida. In: FRANCO, L. J.; PASSOS, A. D. C. (Org). *Fundamentos de Epidemiologia*. Barueri (SP): Manole, 2005. p. 151-164.

- GOULART, B. N. G.; FERREIRA, J. Articulatory Speech Disorders Screening (TERDAF) - validation in to 1st grade public school students population. *Ci. Saúde Col.*, Rio de Janeiro, supl. esp., p. 5636-5636, 2006.
- GOULART, B. N. G. A fonoaudiologia e suas inserções no Sistema Único de Saúde: análise prospectiva. *R. Fonoaud. Bras.*, Brasília, v. 2, n. 4, p. 29-34, dez. 2003.
- GOULART, B. N. G. *Teste de rastreamento de distúrbios articulatórios de fala em crianças da 1ª série do ensino fundamental público*. 2002a. 70 f. Dissertação (Mestrado em Epidemiologia) - Departamento de Medicina Social, Universidade Federal do Rio Grande do Sul, Porto Alegre.
- GOULART, B. N. G. Contribuições da epidemiologia para a pesquisa e atuação clínica em fonoaudiologia. *R. Fono Atual*, São Paulo, v. 21, n. 3, p. 60-63, jul.-set. 2002b.
- HAGE, S. R. V.; JOAQUIM, R. S. S.; CARVALHO, K. G.; PADOVANI, R.; GUERREIRO, M. M. Diagnóstico de crianças com alterações específicas de linguagem por meio de escala de desenvolvimento. *Arq. Neuro-Psiquiatr.*, v. 62, n. 3a, p. 649-653, set. 2004.
- HOCHMAN, B.; NAHAS, F. X.; OLIVEIRA FILHO, R. S.; FERREIRA, L. M. Desenhos de Pesquisa. *Acta Cir. Bras.*, São Paulo, v. 20, supl. 2, p. 2-9, 2005.
- JAESCHKE, R.; GUYATT, G. H.; SACKETT, D. L.; THE EVIDENCE BASED MEDICINE WORKING GROUP. How to Use an Article About a Diagnostic Test. Based on the Users' Guides to Evidence-based Medicine and reproduced with permission from JAMA. *Jama*, v. 271, n. 5, p. 389-391, 1994 (artigo previamente publicado) and v. 271, n. 9, p. 703-707, 1994. Copyright 1995, American Medical Association. Disponível em < <http://www.cche.net/text/usersguides/diagnosis.asp>>. Acesso em: 25 fev. 2007.
- JECKEL, J. F.; ELMORE, J. G.; KATZ, D. L. Delineamentos comuns de pesquisa usados em epidemiologia. In: \_\_\_\_\_. *Epidemiologia, bioestatística e medicina preventiva*. 2. ed. Porto Alegre: Artmed, 2005a. p. 80-112.
- JECKEL, J.; ELMORE, J.; KATZ, D. Métodos de prevenção secundária. In: \_\_\_\_\_. *Epidemiologia, bioestatística e medicina preventiva*. 2. ed. Porto Alegre: Artmed, 2005b. p. 236-247.
- KRAUSS-SILVA, L. Avaliação tecnológica e análise custo-efetividade em saúde: a incorporação de tecnologias e a produção de diretrizes clínicas para o SUS. *Ci. Saúde Col.*, Rio de Janeiro, v. 8, n. 2, p. 501-520, 2003.
- KRAUSS-SILVA, L. Avaliação tecnológica em saúde: questões metodológicas e operacionais. *Cad. Saúde Públ.*, Rio de Janeiro, v. 20, supl. 2, p. S199-S207, 2004.
- McKINLAY, J. B.; MARCEAU, L. D. The end of the golden age of doctoring. *Int. J. Health Serv.*, Amityville (USA), v. 32, n. 2, p. 379-416, 2002.
- MARTELETO, M. R. F.; PEDROMÔNICO, M. R. M. Validade do inventário de comportamentos autísticos (ICA): estudo preliminar. *R. Bras. Psiquiatr.*, São Paulo, v. 27, n. 4, p. 295-301, dez. 2005.
- MOREIRA, R. R.; FERREIRA-JUNIOR, M. Testes de fala: aplicação em portadores de perda auditiva induzida por ruído. *Pró-Fono R. Atual. Cient.*, Barueri (SP), v. 16, n. 3, p. 293-300, set.-dez. 2004.
- NEWMAN, T. B.; BROWNER, W. S.; CUMMINGS, S. R. Delineando estudos de testes médicos. In: HULLEY, S. B. et al. *Delineando a pesquisa clínica: uma abordagem epidemiológica*. 2. ed. Porto Alegre: Artmed, 2003. p. 203-224.
- NOBRE, M. R. C.; BERNARDO, W. M.; JATENE, F. B. A prática clínica baseada em evidências: parte III - avaliação crítica das informações de pesquisas clínicas. *R. Assoc. Med. Bras.*, São Paulo, v. 50, n. 2, p. 221-228, jan.-abr. 2004.
- OLIVEIRA, P.; CASTRO, F.; RIBEIRO, A. Surdez infantil. *R. Bras. Otorrinolaringol.*, São Paulo, v. 68, n. 3, p. 417-423, maio 2002.
- POTVIN, L. Why we should be worried about evidence-based practice in health promotion. *R. Bras. Saúde Mater. Infant.*, Recife, v. 5, supl. 1, p. S93-S97, dez. 2005.
- TOSCANO, C. M. As campanhas nacionais para detecção das doenças crônicas não-transmissíveis: diabetes e hipertensão arterial. *Ci. Saúde Col.*, Rio de Janeiro, v. 9, n. 4, p. 885-895, out.-dez. 2004.
- ZUCKERMAN, D. Hype in health reporting: "checkbook science" buys distortion of medical news. *Int. J. Health Serv.*, Amityville (USA), v. 33, n. 2, p. 383-389, 2003.