

# Confiabilidade da “Entrevista Clínica Estruturada para o DSM-IV – Versão Clínica” traduzida para o português

## Reliability of the Structured Clinical Interview for DSM-IV – Clinical Version translated into Portuguese

Cristina Marta Del-Ben<sup>a</sup>, José Antônio A Vilela<sup>b</sup>, José Alexandre de S Crippa<sup>b</sup>, Jaime Eduardo C Hallak<sup>a</sup>, Cybelli M Labate<sup>c</sup> e Antonio W Zuardi<sup>a</sup>

<sup>a</sup>Departamento de Neurologia, Psiquiatria e Psicologia Médica da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo (FMRP/USP). <sup>b</sup>FMRP/USP. <sup>c</sup>Hospital das Clínicas da FMRP/USP. Ribeirão Preto, SP, Brasil

**Resumo** **Objetivos:** Verificar a confiabilidade da “Entrevista Clínica Estruturada para o DSM-IV – Versão Clínica (SCID-CV)” traduzida para o português.

**Métodos:** Foram submetidos, a duas entrevistas independentes (teste-reteste), 45 pacientes psiquiátricos em seguimento no Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo (HC-FMRP/USP). Os dados foram analisados pelo Coeficiente Kappa (K).

**Resultados:** O Kappa ponderado foi excelente (Kw=0,83). A confiabilidade foi estatisticamente significativa em transtorno do humor (K=0,87); transtornos psicóticos (K=0,90); transtornos relacionados ao uso de substância (K=0,76); transtornos de ansiedade (K=0,61); e nas categorias diagnósticas específicas analisadas, exceto em agorafobia sem história de transtorno do pânico (K=-0,04).

**Conclusões:** A SCID-CV traduzida e adaptada para o português apresenta, em geral, boa confiabilidade, mas a ausência de questões e critérios diagnósticos específicos no próprio instrumento em diagnósticos, como agorafobia sem história de transtorno de pânico, diminuiu sua confiabilidade.

**Descritores** Entrevista psiquiátrica padronizada. Confiabilidade e validade. Diagnóstico.

**Abstract** **Objectives:** To assess the reliability of a Portuguese version of the Structured Clinical Interview for DSM-IV – Clinical Version (SCID-CV).

**Methods:** Forty-five psychiatric patients at the university hospital of Ribeirão Preto, Brazil, were assessed twice in two independent interviews (test-retest). Inter-rater agreement analysis was done using the kappa coefficient (K).

**Results:** The weighted Kappa was excellent (Kw=0.83). Reliability was statistically significant for affective disorders (K=0.87); psychotic disorders (K=0.90); substance-related disorders (K=0.76); anxiety disorder (K=0.61); and for all specific diagnostic categories analyzed, except for agoraphobia without history of panic attacks (K=-0.04).

**Conclusions:** The present Portuguese version of the SCID-CV showed good inter-rater reliability, but the lack of specific questions in the interview and specific diagnostic criteria for some disorders, such as agoraphobia without history of panic attacks, resulted in poor reliability.

**Keywords** Mental status schedule. Reproducibility of results. Diagnosis.

### Introdução

A “Entrevista Clínica Estruturada para o DSM-III-R – Manual Diagnóstico e Estatístico de Distúrbios Mentais, 3ª ed. revisada” (SCID),<sup>1</sup> desenvolvida durante a década de 80, tem sido largamente utilizada, mostrando-se um instrumento útil para o aprimoramento da confiabilidade do diagnóstico psiquiátrico.<sup>2-4</sup>

A ampla utilização da SCID provavelmente se deve a algumas

características do instrumento que facilitam sua aplicação. A SCID inicia-se por uma seção de revisão geral, que segue o roteiro de uma entrevista clínica não-estruturada, conduzida por um profissional experiente. Em seguida, é dividida em módulos que correspondem às categorias diagnósticas maiores. Os critérios diagnósticos estão presentes no próprio corpo do instrumento, facilitando a elaboração do diagnóstico conforme

a entrevista progride. Há também a possibilidade de que questões remanescentes sejam ignoradas, caso critérios essenciais para o diagnóstico não sejam preenchidos (*skip-out*), o que permite o descarte rápido de diagnósticos irrelevantes.<sup>5,6</sup>

No entanto, as mesmas características que tornam a aplicação do instrumento mais acessível poderiam ser fonte de erro e diminuir a confiabilidade e a validade das avaliações diagnósticas. Uma característica essencial da SCID é que, embora as perguntas sejam estruturadas, a pontuação se refere ao julgamento clínico do entrevistador, com relação à presença ou não de determinado critério, e não à resposta dada pelo paciente. Daí a necessidade de treinamento dos entrevistadores e a avaliação da confiabilidade da SCID para os diferentes serviços e profissionais que se dispõem a utilizar o instrumento. No Brasil, obtiveram-se bons índices de concordância com a utilização de uma versão traduzida e adaptada para o português da SCID para o DSM-III-R,<sup>7</sup> por meio da metodologia de entrevistas conjuntas.<sup>8</sup>

Em 1994, foi publicada, pela Associação Americana de Psiquiatria, a 4ª ed. do Manual de Diagnóstico e Estatística de Transtornos Mentais (DSM-IV),<sup>9</sup> que introduziu algumas modificações em relação à versão anterior com o objetivo de aumentar a validade dos critérios diagnósticos propostos.

As questões da SCID foram adaptadas para os critérios propostos pelo DSM-IV,<sup>9</sup> não havendo mudanças na sua estrutura básica. Além da versão de pesquisa, também foi publicada uma outra voltada para a aplicação na prática clínica, independente dos objetivos de pesquisa – “Structured Clinical Interview for DSM-IV Axis I Disorders – Clinician Version” (SCID-CV).<sup>10</sup>

A SCID-CV<sup>10</sup> diferencia-se da versão de pesquisa por não apresentar uma avaliação completa de algumas categorias diagnósticas específicas, como transtornos somatoformes, transtornos alimentares e alguns transtornos de ansiedade, sendo sugerida, nesses casos, a consulta ao DSM-IV<sup>9</sup> para confirmação do diagnóstico. A SCID-CV<sup>10</sup> também não apresenta questões para avaliação de gravidade e de especificadores dos transtornos mentais.

## Objetivo

O objetivo do presente trabalho foi verificar a aplicabilidade da SCID-CV<sup>10</sup> na prática clínica com a realização de um estudo de confiabilidade, pela metodologia teste-reteste,<sup>11</sup> de uma versão do instrumento traduzido e adaptado para o português.

## Método

### Adaptação da SCID-CV para o português

A SCID-CV<sup>10</sup> foi traduzida para o português por dois psiquiatras familiarizados com o inglês e treinados na versão anterior da SCID para o DSM-III-R.<sup>7</sup>

Essa tradução inicial foi submetida à avaliação independente de dois psiquiatras familiarizados com a língua inglesa e com os critérios diagnósticos do DSM-IV,<sup>9</sup> que compararam a entrevista original com o material traduzido e fizeram modificações e adaptações que buscavam manter o mesmo

sentido das perguntas, sem comprometer a compreensão pela população a que se destinavam.

## Sujeitos

Os sujeitos do estudo foram recrutados entre os pacientes atendidos na Enfermaria de Psiquiatria, no Hospital-Dia e no Setor de Psiquiatria da Unidade de Emergência do HC-FMRP/USP, em um período de um ano.

Para a formação da amostra estudada, um profissional não envolvido com as entrevistas de aplicação da SCID, e vinculado a um dos serviços em estudo, selecionou os pacientes e os distribuiu entre os avaliadores. A amostra foi composta por 45 pacientes de ambos os sexos.

O presente estudo foi aprovado pela Comissão de Ética Médica do HC-FMRP/USP. Os pacientes selecionados para a pesquisa foram convidados a participar pessoalmente junto a profissionais vinculados ao serviço no qual faziam seguimento, sendo garantida a preservação da privacidade das pessoas envolvidas no projeto. Todos os pacientes foram informados previamente dos objetivos e dos procedimentos adotados e assinaram um termo de consentimento.

## Entrevistas diagnósticas

As entrevistas foram realizadas por cinco psiquiatras com experiência na aplicação da versão anterior da SCID, treinados por *role playing* e aplicação conjunta em sala com espelho unidimensional em pacientes internados. Também foi realizado um treinamento na realização de diagnósticos, de acordo com os critérios diagnósticos do DSM-IV,<sup>9</sup> pelo estudo sistemático de casos clínicos.

Os entrevistadores foram distribuídos em pares para que cada um dos participantes formasse par com todos os demais, chegando-se a um total de dez pares. Em média, cada entrevistador realizou 18 entrevistas (mínimo de 15, máximo de 20).

Foram realizadas duas entrevistas independentes para cada paciente, com um intervalo mínimo de dois dias e máximo de um mês. Os avaliadores se alternaram na função de primeiro e segundo avaliador e não receberam qualquer informação a respeito dos pacientes antes das entrevistas. A entrevista com o paciente foi a única fonte de informação utilizada para a elaboração do(s) diagnóstico(s).

## Análise estatística

O estudo da confiabilidade do instrumento foi realizado pela análise da concordância dos diagnósticos atribuídos pelos dois avaliadores para um mesmo paciente, por meio do coeficiente Kappa (K),<sup>12</sup> que mede a proporção de concordância diagnóstica observada entre avaliadores, corrigindo a proporção de concordância casual que normalmente seria esperada. O teste de significância estatística foi aplicado conforme proposto por Everitt<sup>13</sup> (1989), sendo considerados significativos valores de p maiores ou iguais a 0,05.

Para uma análise qualitativa, considerou-se que valores de Kappa acima de 0,75 indicariam uma concordância excelente; entre 0,40 e 0,75, uma concordância satisfatória; e abaixo de 0,40, uma concordância insatisfatória.<sup>12</sup>

Inicialmente foi calculado um índice geral de concordância por meio do Kappa ponderado,<sup>14</sup> considerando-se apenas os diagnósticos principais de cada sujeito. A determinação do diagnóstico principal foi dada pela avaliação clínica do entrevistador, considerando o diagnóstico que justificou a procura pelo atendimento psiquiátrico. Os pesos do Kappa ponderado foram estabelecidos de acordo com os seguintes critérios: peso 0 para concordância completa (diagonal da matriz k x k); peso 1 para discordância de categorias diagnósticas específicas que pertenciam à mesma categoria diagnóstica maior (exemplo: esquizofrenia e transtorno psicótico sem outra especificação); e peso 2 para discordâncias de categoria maior (exemplo: transtorno bipolar e esquizofrenia).

A concordância das categorias diagnósticas específicas foi calculada apenas naquelas condições consideradas presentes por pelo menos um dos avaliadores, em pelo menos 10% da amostra, tentando-se, com isto, contornar a instabilidade do Kappa, quando o número de observações é muito baixo.<sup>15,16</sup>

## Resultados

A idade média dos pacientes foi de 34,9±11,8 anos, com um predomínio do sexo feminino (60%). Em cerca de metade da amostra (53,3%), os pacientes receberam, de pelo menos um dos avaliadores, mais de um diagnóstico, sendo incluídos tanto diagnósticos atuais quanto passados.

Os diagnósticos principais levantados na primeira e na segunda entrevista foram, em ordem decrescente de frequência: transtorno bipolar; esquizofrenia; episódio depressivo maior; transtorno de pânico; transtorno esquizoafetivo; transtorno obsessivo-compulsivo; transtornos do humor devido ao uso de substância ou à condição médica geral; transtorno psicótico sem outras especificações; e agorafobia sem história de transtorno de pânico.

O Kappa ponderado para o diagnóstico principal foi 0,83, denotando uma concordância excelente e estatisticamente significativa (p<0,001).

A Tabela 1 mostra que se obteve uma concordância estatisticamente significativa para todas as categorias diagnósticas maiores. A concordância foi excelente para os transtornos de humor, psicóticos e relacionados ao uso de substâncias psicoativas (álcool e não-álcool) e satisfatória para os transtornos de ansiedade.

**Tabela 1 - Concordância (Kappa) para categorias diagnósticas maiores (n=45); proporção do diagnóstico na amostra (%) e significância estatística (p).**

Diagnóstico	(%)	Kappa	p
Transtornos de Humor	55,6	0,87	<0,001
Transtornos Psicóticos	33,3	0,90	<0,001
Transtornos do Uso de Substâncias Psicoativas	28,9	0,76	<0,001
Transtornos de Ansiedade	42,2	0,61	<0,001

Considerando-se o critério de ocorrência mínima de 10% na amostra, foi possível aplicar o tratamento estatístico em oito categorias diagnósticas específicas. A Tabela 2 demonstra que a concordância foi significativa em todas as categorias, com exceção de agorafobia sem história de transtorno de pânico, que apresentou uma concordância insatisfatória (K=-0,04).

**Tabela 2 - Concordância (Kappa) para categorias diagnósticas específicas (n=45); proporção do diagnóstico na amostra (%) e significância estatística (p).**

Diagnóstico	(%)	Kappa	p
Depressão Maior	22,2	0,93	<0,001
Transtorno Bipolar	26,7	0,88	<0,001
Esquizofrenia	24,4	0,80	<0,001
Transtorno esquizoafetivo	11,1	0,73	<0,001
Dependência ao álcool	13,3	0,90	<0,001
Transtorno obsessivo-compulsivo	15,5	0,81	<0,001
Transtorno de pânico	20,0	0,76	<0,001
Agorafobia sem história de transtorno de pânico	11,1	-0,04	0,62

## Discussão

Os resultados obtidos sugerem que a SCID-CV<sup>10</sup> traduzida e adaptada para o português é um instrumento que permite bons índices de confiabilidade do diagnóstico psiquiátrico, já que a concordância geral da entrevista, medida pelo Kappa ponderado, foi bastante satisfatória, bem como a concordância obtida nas categorias diagnósticas maiores e na grande maioria das categorias diagnósticas específicas submetidas ao tratamento estatístico.

A exceção foi o diagnóstico de agorafobia sem história de transtorno de pânico, que não apresentou uma concordância estatisticamente significativa. Essa baixa confiabilidade poderia estar relacionada à própria validade dessa categoria diagnóstica. A ocorrência de agorafobia sem ataques de pânico foi detectada em um grande estudo epidemiológico desenvolvido no início dos anos 80, nos EUA: o *National Institute of Mental Health (NIMH) Epidemiological Catchment Area (ECA)*.<sup>17</sup> Esses resultados influenciaram a inclusão dessa categoria nas classificações diagnósticas mais recentes. Embora o DSM-IV<sup>9</sup> considere a possibilidade de um diagnóstico de agorafobia sem história de transtorno de pânico, essa condição é tratada como rara. Alguns autores<sup>18</sup> questionam a validade desse diagnóstico, porque eles foram detectados em estudos epidemiológicos, mas raramente são vistos em *settings* clínicos.

Essa discrepância entre a observação clínica e os resultados de um estudo populacional pode ser decorrente de limitações na coleta de dados para a elaboração do diagnóstico. No ECA, as informações foram colhidas pelo *Diagnostic Interview Schedule (DIS)*,<sup>19</sup> uma entrevista totalmente estruturada, administrada por entrevistadores treinados no instrumento, mas que não eram profissionais de saúde mental.

Apesar dessas discussões quanto à validade da definição de agorafobia sem história de transtorno de pânico, um fator que pode ter influenciado na baixa confiabilidade obtida neste estudo é a ausência dos critérios diagnósticos do DSM-IV<sup>9</sup> para essa categoria na SCID-CV.<sup>10</sup> Nesse instrumento, é apenas sugerido que, caso o paciente responda positivamente à questão "Você já teve medo de sair de casa sozinho, ficar sozinho no meio de muitas pessoas, ficar em uma fila ou viajar de ônibus ou de trem?", o clínico reporte-se posteriormente ao DSM-IV<sup>9</sup> para confirmação dos demais critérios diagnósticos de agorafobia sem história de transtorno de pânico.

A concordância obtida neste estudo para os transtornos de ansiedade foi satisfatória, mas menor do que aquela obtida para outras categorias diagnósticas maiores, cujos diagnósticos

específicos possuem questões para a avaliação de todos os critérios diagnósticos que os constituem. Em função do pequeno número amostral, não foi possível realizar um tratamento estatístico apropriado para outros transtornos de ansiedade, como fobias específica e social e transtornos de ansiedade generalizada, para os quais também não constam critérios específicos na estrutura da entrevista. Entretanto, para essas categorias também foram observadas discrepâncias entre os diagnósticos obtidos na primeira e na segunda entrevista.

Em geral, espera-se que pela metodologia teste-reteste obtenham-se índices de confiabilidade menores do que aqueles obtidos pelas entrevistas conjuntas.<sup>11</sup> No entanto, essa suposição não se confirma quando os resultados obtidos neste estudo são comparados com aqueles obtidos anteriormente pelo mesmo grupo por meio de entrevistas conjuntas.<sup>8</sup> Provavelmente isto se deve a um melhor treinamento do grupo de pesquisadores no instrumento e na classificação diagnóstica. Além

disso, sabe-se que geralmente a concordância é maior entre avaliadores vinculados a um mesmo serviço.<sup>20</sup>

Embora nesses dois estudos tenham sido usadas versões diferentes da SCID, essa comparação é possível, pois as modificações dos critérios do DSM-III-R<sup>7</sup> para o DSM-IV<sup>9</sup> foram pequenas. De qualquer forma, essa comparação deve ser feita com cautela, pois, apesar de muito semelhantes, tratam-se de dois instrumentos diferentes.

A confiabilidade obtida com a tradução e a adaptação para o português da SCID-CV,<sup>10</sup> por meio da metodologia teste-reteste, sugere que esse instrumento pode ser útil para aplicação no Brasil, desde que sejam consideradas apenas as categorias diagnósticas para as quais a entrevista apresenta todos os critérios diagnósticos. A ausência de critérios diagnósticos no corpo do instrumento levou a uma baixa concordância do diagnóstico, mesmo entre profissionais treinados na SCID e familiarizados com o DSM-IV.

## Referências

- Spitzer RL, Williams JR, Gibbon M, First MB. Structured clinical interview for DSM-III-R - patient edition (SCID-P, version 1.0). Washington (DC): American Psychiatric Press; 1990. p. 1-124.
- Williams JR, Gibbon M, First MB et al. The structured clinical interview for DSM-III-R (SCID). II - multisite test-retest reliability. *Arch Gen Psychiatry* 1992;49:630-6.
- Arntz A, Van Beijsterveldt B, Hoeskstra R, Hofman A, Eussen M, Sallaerts S. The interrater reliability of a Dutch version of the structured clinical interview for DSM-III-R personality disorders. *Acta Psychiatr Scand* 1992;85:394-400.
- Skre I, Onstad S, Torgersen S, Kringlen E. High interrater reliability for the structured clinical interview for DSM-III-R axis I (SCID-I). *Acta Psychiatr Scand* 1991;84:167-173.
- Spitzer RL, Williams JR, Gibbon M, First MB. Instruction manual for the structured clinical interview for DSM-III-R (SCID, 5/1/89 Revision). New York: Biometrics Research Department/ New York State Psychiatric Institute; 1989. p. 1-56.
- Spitzer RL, Williams JR, Gibbon M, First MB. The structured clinical interview for DSM-III-R. I - history, rationale, and description. *Arch Gen Psychiatry* 1992;49:624-9.
- American Psychiatric Association. manual diagnóstico e estatístico de distúrbios mentais. 3ª ed. Revisada. São Paulo: Manole; 1989. p. 1-379.
- Del-Ben CM, Rodrigues CRC, Zuairi AW. Reliability of a Portuguese version of the structured clinical interview for DSM-III-R (SCID) in a Brazilian sample of outpatient. *Braz J Med Bio Res* 1996;29:1675-82.
- American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4th ed. Washington (DC): American Psychiatric Association; 1994. p. 1-550.
- First MB, Spitzer RL, Gibbon M, Williams JBW. Structured clinical interview for DSM-IV axis I disorders - clinician version (SCID-CV). Washington (DC): American Psychiatric Press; 1997. p. 84.
- Helzer JE, Robins LN, Taibleson M, Woodruff RA Jr, Reich T, Wish ED. Reliability of psychiatric diagnosis. I. A methodological review. *Arch Gen Psychiatry* 1977;34:129-33.
- Fleiss JL. The measurement of interrater agreement. In: *Statistical methods for rates and proportions*. 2nd ed. New York: John Wiley & Sons Inc.; 1981. p. 212-36.
- Everitt BS. Measurement in medicine. In: *Statistical methods for medical investigations*. New York: Oxford university Press; 1989. p. 16-27.
- Cohen J. Weighted Kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull* 1968;70:213-20.
- Grove WM, Andreasen NC, McDonald-Scott P, Keller MB, Shapiro RW. Reliability studies of psychiatric diagnosis: theory and practice. *Arch Gen Psychiatry* 1981;38:408-13.
- Shrout PE, Spitzer RL, Fleiss JL. Quantification of agreement in psychiatric diagnosis revisited. *Arch Gen Psychiatry* 1987;44:172-7.
- Robins LN, Helzer JE, Weissman MM, Orvaschel H, Gruenberg E, Burke JD-Jr, et al. Lifetime prevalence of specific psychiatric disorders in three sites. *Arch Gen Psychiatry* 1984;41(10):949-58.
- Gorman JM, Liebowitz MR, Fyer AJ, Stein J. A neuroanatomical hypothesis for panic disorder. *Am J Psychiatry* 1989;146:148-61.
- Robins LN, Helzer JW, Croughan J, Ratcliff KS. National institute of mental health diagnostic interview schedule. Its history, characteristics, and validity. *Arch Gen Psychiatry* 1981;38:381-9.
- Flemenbaum A, Zimmermann RL. Inter- and intra-rater reliability of the brief psychiatric rating scale. *Psychol Rep* 1973;36:783-92.

---

## Correspondência

Cristina Marta Del-Ben  
Av. dos Bandeirantes, 3900  
14048-900 Ribeirão Preto, SP, Brasil  
Tel.: (0xx16) 602-2533  
Fax: (0xx16) 635-0713  
E-mail: delben@fmrp.usp.br

---