

Mariana Miranda Autran
Sampaio¹

Cláudia Medina Coeli¹

Nair Navarro de Miranda^{II}

Eduardo Faerstein¹

Guilherme Loureiro Werneck¹

Dora Chor^{III}

Cláudia S Lopes¹

Confiabilidade interobservador da Classificação Internacional de Cuidados Primários

Interobserver reliability of the International Classification of Primary Care

RESUMO

OBJETIVO: A Classificação Internacional de Cuidados Primários foi desenvolvida para suprir as limitações da Classificação Estatística Internacional de Doenças e Problemas Relacionados à Saúde, 10ª revisão, quando aplicada na atenção primária de saúde. O objetivo do estudo foi avaliar a confiabilidade interobservador da Classificação Internacional de Cuidados Primários na codificação dos motivos de saúde para interrupção de atividades habituais.

MÉTODOS: Foram analisados dados relativos a 801 participantes da Fase 2 (2001) do Estudo Pró Saúde, realizado com funcionários de uma universidade localizada no estado do Rio de Janeiro, que afirmaram ter ficado impedidos de realizar alguma de suas atividades habituais (trabalho, estudo ou lazer) por motivo de saúde nas duas semanas anteriores à coleta dos dados. Os motivos de saúde relatados em resposta à pergunta aberta foram codificados de maneira independente por duas classificadoras. A confiabilidade interobservador em relação ao número de motivos foi avaliada por meio do coeficiente kappa ponderado; para as demais análises (capítulos e códigos completos), empregou-se o coeficiente kappa simples.

RESULTADOS: Foram codificados 1641 motivos pela primeira classificadora e 1629 pela segunda. A confiabilidade interobservador em relação ao número de motivos codificados foi substancial (kappa ponderado=0,94; IC 95%: 0,93;0,94). Em relação aos códigos de capítulos e os códigos completos, foram encontradas confiabilidade substancial (kappa= 0,89; IC 95%: 0,88;0,90) e moderada (0,76; IC 95%: 0,76;0,78), respectivamente.

CONCLUSÕES: Os resultados sugerem que a Classificação Internacional de Cuidados Primários é adequada para a codificação dos motivos de saúde na interrupção de atividades habituais.

DESCRITORES: Atenção Primária à Saúde, classificação. Variações Dependentes do Observador. Classificação Internacional de Doenças. Questionários, utilização. Estudos de Validação.

¹ Instituto de Medicina Social. Universidade Estadual do Rio de Janeiro. Rio de Janeiro, RJ, Brasil

^{II} Escola Politécnica de Saúde Joaquim Venâncio. Fundação Oswaldo Cruz (Fiocruz). Rio de Janeiro, RJ, Brasil

^{III} Escola Nacional de Saúde Pública. Fiocruz. Rio de Janeiro, RJ, Brasil

Correspondência | Correspondence:
M.M.A. Sampaio
Instituto de Medicina Social
Universidade do Estado do Rio de Janeiro
R. São Francisco Xavier 524, 7º andar
20559-900 Rio de Janeiro, RJ, Brasil
E-mail: m.autran@gmail.com

Recebido: 23/11/2006
Revisado: 1/10/2007
Aprovado: 8/11/2007

ABSTRACT

OBJECTIVE: The International Classification of Primary Care was developed as an attempt to overcome the limitations of the International Statistical Classification of Diseases and Related Health Problems, 10th revision, when used for primary health care. The aim of the study was to evaluate the interobserver reliability of the International Classification for Primary Care when coding reasons for health-related interruption of daily activities.

METHODS: Data analyzed pertained to 801 subjects from Phase 2 of the Pró-Saúde Study, involving the employees of a Rio de Janeiro university who reported having been prevented from carrying out any of their usual activities (work, study, or leisure) for health-related reasons in the two weeks prior to data collection. Health problems reported in response to an open question were separately coded by two classifiers. Interobserver reliability with respect to number of health problems was calculated by weighted kappa; for the remaining analyses (chapters and full codes), crude kappa coefficients were used.

RESULTS: A total of 1,641 health problems were coded by the first classifier, and 1,629 by the second. Interobserver reliability with respect to the number of health problems coded was substantial (weighted kappa=0.94; 95% CI: 0.93;0.94). Chapter and full codes showed substantial (kappa=0.89; 95% CI: 0.88;0.90) and moderate (0.76; 95% CI: 0.76;0.78) reliability, respectively.

CONCLUSIONS: The results suggest that the International Classification of Primary Care is adequate for the coding of health-related reasons for interruption of daily activities.

DESCRIPTORS: Primary Health Care, classification. Observer Variation. International Classification of Diseases. Questionnaires, utilization. Validation Studies.

INTRODUÇÃO

Os inquéritos de morbidade referida vêm sendo empregados na avaliação em saúde, principalmente por permitirem o acesso ao perfil de morbidade da população geral. A população que busca serviços de saúde e consegue atendimento, avaliada por estudos ambulatoriais e hospitalares, pode apresentar perfil distinto da população geral.⁶ Uma das questões a ser abordada na condução de inquéritos de morbidade diz respeito à forma como os dados são coletados e codificados. Podem ser utilizadas tanto listas de problemas de saúde ou de sintomas comuns como uma pergunta geral aberta. Os estudos que optam pela pergunta aberta, sem a apresentação de lista de problemas de saúde ou sintomas, demandam um esquema de codificação. A Classificação Estatística Internacional de Doenças e Problemas Relacionados à Saúde, 10ª revisão (CID-10), é a classificação adotada internacionalmente para codificar diagnósticos e elaborar estatísticas de saúde. Todavia, ela é considerada inadequada para situações

em que não há condições bem definidas nem a possibilidade de diagnósticos precisos, como os inquéritos ou a assistência primária em saúde.^{5,8,a}

A Classificação Internacional de Cuidados Primários,¹⁵ do original *International Classification of Primary Care* (ICPC), surgiu para complementar a utilização da CID no contexto da assistência primária e tem como principal característica a inclusão de queixas expressas pelos pacientes e dos problemas de natureza social.³ Para garantir a comparabilidade dos seus códigos com os da CID-10, ela sofreu uma revisão que ficou conhecida como ICPC-2. Essa nova versão tem estrutura biaxial: o primeiro eixo apresenta capítulos que compreendem os sistemas orgânicos, psicológicos e sociais aos quais os relatos se referem (geral e inespecífico, sangue, digestivo, olho, orelha, entre outros); o segundo eixo apresenta os componentes sobre o tipo de relato (sinais e sintomas, procedimentos, diagnósticos e doenças). O código de um motivo de consulta

^a Jamouille M, Roland M. The WONCA Classification Committee, 1972-1997, 25 years in the service of family practice. 1997 [acesso em 9 nov 2004]. Disponível em: <http://www.ulb.ac.be/esp/wiccc/history.html>

é composto por uma letra, que representa o capítulo, e dois dígitos numéricos, que representam os componentes. Em função de ter sido criada para a assistência primária e considerar a fala do paciente da forma como é enunciada, é possível supor que a ICPC tenha bom desempenho em perguntas abertas de inquéritos de saúde. No entanto, essa classificação é recente e tem seu uso pouco explorado no Brasil. Em uma revisão da literatura⁶ foi encontrado apenas um estudo que avalia o uso dessa classificação no Brasil.⁷ Além disso, apenas um outro estudo utilizou a ICPC-2 para a codificação de morbidade referida em questionários, porém sem apresentar avaliação de sua confiabilidade.¹⁴

O presente artigo teve por objetivo avaliar a confiabilidade interobservador da codificação independente das respostas a uma pergunta aberta sobre os motivos de interrupção de atividades habituais, utilizando a ICPC-2.

MÉTODOS

Os dados analisados fazem parte de um projeto mais amplo, Estudo Pró-Saúde, cujas características e metodologia foram publicadas em 2005 (Faerstein et al²). Trata-se de uma coorte de funcionários técnico-administrativos do quadro efetivo de uma universidade localizada no estado do Rio de Janeiro, cujo objetivo principal foi estudar a associação entre determinantes sociais e diversos desfechos de saúde. Até o momento foram realizadas três fases de coleta de dados no Estudo Pró-Saúde, em 1999, em 2001 e em 2006/2007. No presente estudo, foram avaliados os funcionários que: a) responderam ao questionário auto-preenchível da Fase 2 (2001); b) responderam de forma válida à pergunta sobre interrupção de atividades habituais por motivo de saúde; c) afirmaram ter ficado impedidos de realizar alguma de suas atividades habituais por motivos de saúde nas duas semanas anteriores ao preenchimento do questionário. Do total de 812 pessoas que informaram ter interrompido suas atividades habituais por problemas de saúde, 11 não preencheram o motivo que levou a tal interrupção e foram excluídas da análise. Portanto, a população de estudo foi composta por 801 funcionários.

A pergunta que avaliou a interrupção de atividades habituais por motivo de saúde foi: “nas duas últimas semanas, qual foi ou quais foram esses problemas de saúde que você teve ou tem que o(a) impediram de realizar alguma dessas suas atividades habituais (por exemplo, trabalho, estudo, lazer ou tarefas domésticas)?”

A codificação foi realizada por duas classificadoras: uma especialista em classificações, em especial na CID, porém sem experiência no uso da ICPC-2; e outra que não possuía experiência no uso de qualquer

tipo de classificação. Ambas classificadoras realizaram treinamento que envolveu a leitura da classificação e sessões subseqüentes para discussão de sua lógica antes do início da codificação principal. Posteriormente, a título de teste, foram codificados os motivos de consulta de 29 atendimentos de uma unidade primária de saúde e de 59 questionários do Estudo Pró-Saúde, que não compuseram a amostra estudada. Em seguida, foram realizadas reuniões para chegar ao consenso das codificações, com a participação de uma terceira classificadora, e para estabelecer diretrizes para a codificação dos dados da população de estudo.⁸

Como a queixa de uma pessoa podia ser interpretada com um ou mais códigos, houve divergência no número de motivos codificados por cada classificadora. Dessa forma, foi realizada uma avaliação da confiabilidade, entre as duas classificadoras, em relação ao número de motivos codificados, por meio do kappa ponderado.¹³ Além disso, foi analisada a confiabilidade da ICPC-2, segundo capítulos (por exemplo, se uma classificadora codificou P01 e a outra P02, elas concordaram no capítulo), segundo os códigos completos dentro de cada capítulo (analisando separadamente o capítulo P, as classificadoras deveriam concordar no código completo, por exemplo, P01 e P01) e segundo os códigos completos em geral (as classificadoras deveriam concordar no código completo, no entanto, o cálculo da confiabilidade foi feito para o conjunto dos códigos, i.e., não levando em consideração os capítulos). Posteriormente, a análise de capítulos e de códigos completos foi estratificada por sexo, escolaridade e ocupação. A estratificação foi feita porque o sujeito, dependendo de suas características, tende a se expressar de maneira diferente, com mais ou menos detalhes, usando determinadas peculiaridades linguísticas. Portanto, como em um questionário auto-preenchível não existe a possibilidade de esclarecimentos, a qualidade do relato pode representar dificuldades distintas na hora da codificação. Em todos os casos, a confiabilidade entre as duas classificadoras foi estimada por meio da estatística kappa.¹ Os intervalos de 95% de confiança (IC 95%) do kappa foram calculados empregando-se a rotina *kapci*¹⁰ desenvolvida para o *software* Stata.

Para a interpretação do kappa, foi empregada classificação proposta por Shrout:¹¹ $k < 0,10$ confiabilidade virtualmente ausente; 0,10 a 0,40 – confiabilidade fraca; 0,41 a 0,60 – confiabilidade discreta; 0,61 a 0,80 – confiabilidade moderada; 0,81 a 1,0 – confiabilidade substancial. Essa classificação vem sendo mais recentemente empregada por alguns autores¹² e representa um desenvolvimento da classificação proposta por Landis & Koch.⁴

⁸ Sampaio MMA. Avaliação da Classificação Internacional de Cuidados Primários na codificação dos motivos de interrupção de atividades habituais no Estudo Pró-Saúde [dissertação de mestrado]. Rio de Janeiro: Instituto de Medicina Social, Universidade do Estado do Rio de Janeiro; 2006.

RESULTADOS

A primeira classificadora codificou um total de 1.641 motivos, enquanto a segunda 1.621, com ambas identificando uma mediana de dois motivos por participante. A análise em relação ao número de motivos codificados por participante apresentou concordância simples de 82,4%, com kappa simples de 0,78 (IC 95%: 0,77;0,78) e kappa ponderado de 0,94 (IC 95%: 0,93;0,94), indicando confiabilidade substancial.

A Tabela 1 mostra que as estimativas de confiabilidade dos códigos de capítulos foram substanciais, tanto globalmente como em cada estrato. Quando se considerou o código completo, as estimativas de confiabilidade foram moderadas, tendendo a substanciais. A pequena diferença entre ocupações na análise de códigos completos deixa de existir quando são considerados os códigos de capítulos. A confiabilidade entre os sexos também foi semelhante, mas houve diferença em relação à escolaridade: a confiabilidade entre participantes com até o primeiro grau foi menor que a observada para os demais, apesar da confiabilidade nesse estrato ter sido considerada moderada.

A Tabela 2 mostra a análise de concordância de códigos completos, dentro de cada capítulo da ICPC-2. Os motivos relacionados a alguns capítulos apareceram com uma baixa frequência, dificultando a avaliação da confiabilidade. É possível perceber que os capítulos N (sistema nervoso), R (aparelho respiratório) e P (psicológico) apresentaram confiabilidade substancial. Houve concentração de códigos referentes a poucos diagnósticos como cefaléia (178 de 248), no capítulo referente ao sistema nervoso; infecção aguda do aparelho respiratório superior (35 de 183), sinusite crônica/aguda (28 de 183) e gripe (52 de 183), no

capítulo referente ao aparelho respiratório; e sensação de ansiedade/nervosismo/tensão (71 de 263), reação aguda ao stress (31 de 263) e perturbações depressivas (99 de 263), no capítulo referente ao psicológico. O capítulo K (aparelho circulatório) apresentou confiabilidade fraca. Analisando-se detalhadamente as discordâncias do capítulo K, foi possível perceber que em 19 das 41 discordâncias, uma das pesquisadoras atribuiu o código K85 (tensão arterial elevada) enquanto a outra, o código K86 (hipertensão sem complicações), de significados semelhantes. Caso as classificadoras tivessem atribuído o mesmo código a esse motivo, a concordância simples seria de 74,4% com um kappa simples de 0,70 (IC 95%: 0,63;0,70), classificado como moderado.

DISCUSSÃO

O presente estudo encontrou confiabilidade substancial entre as duas classificadoras para o número de motivos codificados e capítulos, e confiabilidade moderada tendendo a substancial em relação aos códigos completos. A confiabilidade alta foi inesperada, considerando-se que uma das classificadoras não tinha qualquer tipo de experiência prévia em codificação de morbidade.

A diferença da confiabilidade para os códigos completos entre os motivos relatados por funcionários com escolaridade até primeiro grau quando comparados com os de maior escolaridade pode indicar uma dificuldade de compreensão da linguagem empregada por esse grupo por parte das classificadoras. Por outro lado, também é possível pensar que a classificação poderia englobar uma gama maior de expressões, facilitando a generalização de sua aplicação.

Tabela 1. Concordância entre as classificadoras em relação aos códigos completos e aos códigos dos capítulos dos motivos de consulta mencionados, segundo sexo, escolaridade e ocupação do respondente, usando a ICPC-2. Estudo Pró-Saúde, Estado do Rio de Janeiro, 2001.

| Variável | Capítulo | | Código completo | |
|-----------------------|------------------------------------|------------------|------------------------------------|------------------|
| | Percentual de concordância simples | kappa (IC 95%) | Percentual de concordância simples | kappa (IC 95%) |
| Global | 90,6% | 0,89 (0,89;0,90) | 76,5% | 0,76 (0,76;0,78) |
| Sexo | | | | |
| Masculino | 91,5% | 0,90 (0,89;0,91) | 78,0% | 0,77 (0,74;0,80) |
| Feminino | 90,3% | 0,89 (0,88;0,89) | 75,9% | 0,75 (0,72;0,79) |
| Escolaridade | | | | |
| Até 1º grau | 90,0% | 0,88 (0,86;0,92) | 70,7% | 0,70 (0,68;0,72) |
| Até 2º grau | 90,8% | 0,89 (0,89;0,91) | 76,7% | 0,76 (0,72;0,78) |
| 3º grau | 90,4% | 0,89 (0,88;0,89) | 78,6% | 0,78 (0,75;0,80) |
| Ocupação | | | | |
| Profissional de saúde | 89,9% | 0,88 (0,87;0,89) | 75,2% | 0,74 (0,72;0,79) |
| Outros | 91,2% | 0,90 (0,90;0,92) | 77,7% | 0,77 (0,76;0,78) |

Tabela 2. Concordância interobservador em relação aos códigos completos dos motivos de consulta mencionados em cada capítulo, usando a ICPC-2. Estudo Pró-Saúde, Estado do Rio de Janeiro, 2001.

| Capítulo | N | Percentual de concordância simples | kappa | IC 95% |
|--------------------------------------|-----|------------------------------------|-------|-----------|
| N (sistema nervoso) | 248 | 96,4 | 0,92 | 0,86;0,94 |
| R (aparelho respiratório) | 183 | 83,6 | 0,81 | 0,79;0,85 |
| P (psicológico) | 263 | 83,7 | 0,78 | 0,75;0,79 |
| H (ouvidos) | 18 | 83,3 | 0,74 | 0,53;0,80 |
| U (aparelho urinário) | 21 | 76,2 | 0,7 | 0,62;0,75 |
| D (aparelho digestivo) | 153 | 71,9 | 0,69 | 0,67;0,73 |
| L (sistema musculoesquelético) | 407 | 67,1 | 0,65 | 0,61;0,67 |
| X (aparelho genital feminino) | 56 | 69,6 | 0,65 | 0,61;0,66 |
| A (geral e inespecífico) | 184 | 67,4 | 0,63 | 0,62;0,65 |
| B (sangue) | 3 | 66,7 | 0,57 | 0,40;1,00 |
| T (endócrino e metabólico) | 18 | 61,1 | 0,57 | 0,44;0,68 |
| F (olhos) | 28 | 60,7 | 0,56 | 0,53;0,63 |
| K (aparelho circulatório) | 86 | 52,3 | 0,45 | 0,41;0,52 |
| Z (problemas sociais) | 7 | 57,1 | 0,43 | 0,00;0,47 |
| W (gravidez e planejamento familiar) | 11 | 36,4 | 0,34 | 0,15;0,38 |
| S (pele) | 23 | 30,4 | 0,29 | 0,24;0,33 |
| Y (aparelho genital masculino) | 0 | - | - | - |

Quando considerou-se o nível do capítulo, a confiabilidade entre profissionais de saúde e outros profissionais foi próxima, sendo até um pouco mais alta para os outros profissionais. Seria esperado que profissionais de saúde tivessem melhor capacidade para informar sobre problemas de saúde e, portanto, suas respostas resultariam em maior confiabilidade interobservador. No entanto, isso não ocorreu. Uma das possíveis explicações para esse resultado pode ser relacionada ao fato de que uma informação muito detalhada também pode trazer dificuldades para a classificação, como por exemplo, “hérnias discais na coluna cervical com comprometimento do canal”. Esse relato foi codificado como síndrome da coluna cervical por uma das classificadoras e síndrome vertebral com irradiação de dores pela outra classificadora.

A concentração da classificação em alguns códigos mais frequentes pode ter contribuído para que houvesse maior confiabilidade nesses capítulos. A fraca confiabilidade no capítulo do aparelho circulatório pode ser atribuída à presença de códigos diferentes para a classificação de problemas de saúde bastante semelhantes. A distinção entre pressão alta e hipertensão sem complicações poderia ser relevante em outros contextos, mas não nos casos de assistência primária ou de questionários, uma vez que o paciente ou o respondente não tem necessariamente clareza da diferença entre os termos e provavelmente usaria um ou outro indistintamente. No entanto, como se trata de importante aspecto da codificação, é necessário que este detalhe seja enfatizado no treinamento de codificadores.

O único artigo similar ao presente estudo foi conduzido por Letriliart et al.⁹ Os autores estudaram a confiabilidade da classificação de códigos atribuídos por médicos generalistas treinados na ICPC-2, que classificaram diretamente a partir da fala do paciente, e epidemiologistas, que classificaram indiretamente, a partir de informações médicas registradas em uma base de dados. Os autores encontraram kappa ponderado de 0,65 (IC 95%: 0,52;0,77) para a confiabilidade relativa ao número de motivos. A concordância simples entre códigos (considerado apenas o nível do capítulo) foi de 69,2% (83 de 120) e o kappa simples, 0,84 (IC 95%: 0,78;0,91). Ainda que seja questionável a comparação entre estatísticas kappa observadas em populações de estudo diversas (uma vez que a medida sofre a influência da prevalência do fenômeno em estudo), observaram-se valores inferiores aos encontrados no presente estudo. É possível supor que essas discrepâncias sejam explicadas, ao menos parcialmente, pela utilização de fontes diferentes para a captação dos dados (primária vs. secundária) e profissionais distintos (clínicos vs. epidemiologistas).

No artigo de Van der Heyden et al,¹⁴ a ICPC-2 foi aplicada em pergunta aberta de questionário auto-preenchível e foram observados alguns problemas no uso da classificação, como o caráter pouco específico de algumas respostas, respostas inadequadas à pergunta, casos em que não ficou claro o código que deveria ser empregado e elevado consumo de tempo para a codificação das respostas. No presente estudo, também

foi observado o caráter pouco específico de algumas respostas. Provavelmente essa é uma dificuldade do uso da ICPC-2 em questionários auto-preenchíveis, uma vez que não há possibilidade de esclarecer ou detalhar pontos obscuros na fala do paciente. Na assistência primária ou no caso da aplicação de um questionário em que o entrevistador estivesse preenchendo o formulário, seria possível fazer mais perguntas no caso de respostas pouco específicas. Não foram observadas respostas inadequadas à pergunta, nem casos em que não ficou claro o código que deveria ser empregado.

Observou-se que, à medida que as classificadoras adquiriam mais experiência com a classificação, o processo se tornou mais rápido. Entretanto, o tempo de codificação pode ter sido aumentado por alguns problemas da ICPC-2, como o fato de o índice não comportar expressões, tais como alergia, distensão muscular, enjôo, gripe, hérnia, resfriado, dentre muitas outras. A classificação de procedimentos também apresenta algumas falhas, pois contém termos bastante especifi-

cos, como excisão e exerése, mas não contém termos como cirurgia. Uma classificação cuja utilização não é restrita aos médicos⁷ poderia conter um leque mais vasto de opções para que pessoas que não conheçam os procedimentos estejam aptas a codificar.

Em conclusão, a ICPC-2 apresentou boa confiabilidade interobservador na codificação dos problemas de saúde que levaram à interrupção de atividades habituais. O fato de a população acompanhada no presente estudo ter sido composta por funcionários de uma universidade pública faz com que seu perfil seja, em vários aspectos, diferente da população em geral. Sendo assim, os resultados encontrados devem apenas ser generalizados para populações com perfil semelhante ao do Estudo Pró-Saúde. Entretanto, as análises segundo sexo, escolaridade e ocupação revelaram desempenho similar da classificação nesses diferentes estratos, sugerindo que a ICPC-2 possa também apresentar desempenho adequado em outros contextos.

REFERÊNCIAS

- Cohen J. A coefficient of agreement for nominal scales. *Educ Psych Meas*. 1960;20(1):37-46.
- Faerstein E, Chor D, Lopes CS, Werneck GL. Estudo Pró-Saúde: características gerais e aspectos metodológicos. *Rev Bras Epidemiol*. 2005;8(4):454-66.
- Lamberts H, Wood M. The birth of the International Classification of Primary Care (ICPC): Serendipity at the border of Lac Léman. *Fam Pract*. 2002;19(5):433-5.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-74.
- Laurenti R. Décima revisão da Classificação Internacional de Doenças e Problemas Relacionados à Saúde (CID-10): a revisão do final do século. *Bol Oficina Sanit Panam*. 1995;118(3):273-80.
- Lebrão ML, Carandina L, Magaldi C. Análise das condições de saúde e de vida da população urbana de Botucatu, São Paulo (Brasil). IV – Morbidade referida em entrevistas domiciliares, 1983-1984. *Rev Saude Publica*. 1991;25(6):452-60.
- Lebrão ML. Classificação internacional de motivos de consulta para assistência primária: teste em algumas áreas brasileiras. *Rev Saude Publica*. 1985;19(1):69-78.
- Lebrão ML. Estudos de morbidade. São Paulo: Edusp; 1997.
- Letrilliat L, Guiguet M, Flahault A. Reliability of report coding of hospital referrals in primary care versus practice-based coding. *Eur J Epidemiol*. 2000;16(7):653-9.
- Reichenheim ME. Confidence intervals for the kappa statistic. *Stata J*. 2004;4(4):421-8.
- Shrout PE. Measurement reliability and agreement in psychiatry. *Stat Methods Med Res*. 1998;7(3):301-17.
- Simões SEM, Reichenheim ME. Confiabilidade das informações de causa básica nas declarações de óbito por causas externas em menores de 18 anos no Município de Duque de Caxias, Rio de Janeiro, Brasil. *Cad Saude Publica*. 2001;17(3):521-31.
- Szklo M, Nieto FJ. Quality assurance and control. In: Szklo M, Nieto FJ. *Epidemiology: beyond the basics*. 2nd ed Jones and Bartlett: Boston; 2007. p.326-30.
- Van der Heyden J, Van Oyen H, Gisle L. The classification of health problems in health interview surveys: using the International Classification of Primary Care (ICPC). *Soz Praventivmed*. 2004;49(2):161-3.
- Comissão de Classificações da Organização Mundial de Ordens Nacionais, Academias e Associações Acadêmicas de Clínicos Gerais/Médicos de Família (WONCA). *Classificação internacional de cuidados primários*. 2. ed. Oxford: Oxford University Press; 1999.