

Rev. Latino-Am. Enfermagem
mar.-abr. 2014;22(2):234-40
DOI: 10.1590/0104-1169.3262.2407
www.eerp.usp.br/rlae

Artículo Original

Fiabilidad de indicadores de calidad del cuidado de enfermería: comprobando la concordancia y fiabilidad entre evaluadores¹

Dagmar Willamowius Vituri²
Yolanda Dora Martinez Évora³

Objetivo: comprobar la concordancia y fiabilidad entre evaluadores de quince indicadores de calidad de la asistencia de enfermería. **Método:** investigación cuantitativa, metodológica, experimental y aplicada, realizada en un hospital terciario, de gran porte, público de enseñanza del estado del Paraná. En el tratamiento de los datos fue utilizada la estadística Kappa para el análisis de las variables categóricas – indicadores de 1 a 11 y 15; y el Coeficiente de Correlación Intraclase para las variables continuas – indicadores 12, 13 y 14, considerando el Intervalo de Confianza de 95%. Los datos categóricos fueron tratados utilizando el software Lee, del Laboratorio de Epidemiología y Estadística del Instituto Dante Pazzanese de Cardiología – Brasil y los datos continuos por el software BioEstat 5.0. **Resultados:** los resultados por la aplicación de la estadística Kappa demostraron un excelente grado de concordancia, estadísticamente significativo; los valores obtenidos para el Coeficiente de Correlación Intraclase denotaron reproductibilidad/concordancia excelente y estadísticamente significativo para los indicadores evaluados. **Conclusión:** los indicadores presentan excelente confiabilidad y reproducibilidad, lo que evidencia que el desarrollo de instrumentos de evaluación válidos y fidedignos es posible, además de imprescindible para la administración de la asistencia de enfermería

Descriptor: Enfermería; Auditoria de Enfermería; Indicadores de Calidad de la Atención de Salud; Estudios de Validación como Asunto; Reproducibilidad de Resultados.

¹ Artículo parte de la tesis de doctorado "Evaluación como principio de la Gestión de Calidad: probando la confiabilidad de "intervaladores" de indicadores para evaluar la calidad de la asistencia de enfermería" presentada a la Escola de Enfermagem de Ribeirão Preto, Universidade de São Paulo, Centro Colaborador de la OMS para el Desarrollo de la Investigación en Enfermería, Ribeirão Preto, SP, Brasil.

² Estudiante de doctorado, Escola de Enfermagem de Ribeirão Preto, Universidade de São Paulo, Centro Colaborador de la OMS para el Desarrollo de la Investigación en Enfermería, Ribeirão Preto, SP, Brasil. Enfermera, Hospital Universitário, Universidade Estadual de Londrina, Londrina, PR, Brasil.

³ PhD, Profesor Titular, Escola de Enfermagem de Ribeirão Preto, Universidade de São Paulo, Centro Colaborador de la OMS para el Desarrollo de la Investigación en Enfermería, Ribeirão Preto, SP, Brasil.

Correspondencia:

Dagmar Willamowius Vituri
Universidade Estadual de Londrina. Hospital Universitário
Gerência de Riscos
Av. Robert Koch, 60
Vila Operária
CEP: 86038-350, Londrina, PR, Brasil
E-mail: dagmar@uel.br

Copyright © 2014 Revista Latino-Americana de Enfermagem

Este es un artículo de acceso abierto distribuido bajo los términos de la Licencia Creative Commons Reconocimiento-No Comercial (CC BY-NC). Esta licencia permite a otros distribuir, remezclar, retocar, y crear a partir de tu obra de modo no comercial, y a pesar de que sus nuevas obras deben siempre mencionarte y mantenerse sin fines comerciales, no están obligados a licenciar sus obras derivadas bajo las mismas condiciones.

Introducción

La preocupación con la calidad de los productos y servicios ha sido objeto de aprehensión en los diversos tipos de organización, principalmente en el sector salud, ya que, además de la influencia directa de estos servicios sobre la economía, sus clientes están cada vez más exigentes en cuanto a la atención de sus necesidades con calidad⁽¹⁾.

Un gran aliado de las organizaciones de salud en la búsqueda por la calidad de las acciones es el enfermero y, en este sentido, existe expectativa de los gestores en cuanto al papel de este profesional en la administración del cuidado en las instituciones hospitalarias, ya que, una de las características de la práctica del cuidar en enfermería es el contacto directo con el cliente, que posibilita conocer sus necesidades y expectativas⁽¹⁾, además del hecho de la profesión tener un papel fundamental en el resultado de la atención.

Administrar implica en conocer el desempeño de la institución en función de su objetivo y en relación a sus metas⁽²⁾ y eso requiere la implementación de sistemas de evaluación e indicadores adecuados, que permitan la (re)formulación de directrices⁽³⁾. Sin embargo, la cuantificación de la calidad de la atención en salud ha sido un gran desafío, que se inicia con la decisión de lo que debe ser medido y, después, en encontrar medidas estadísticas de calidad⁽⁴⁾.

Un indicador es una unidad de medida de una actividad, que puede ser utilizada para mensurar calidad y cantidad en organizaciones de salud, evaluando aspectos de estructura, procesos y resultados⁽⁵⁻⁶⁾. La posibilidad de desarrollarlos es ilimitada, no obstante, la dificultad está en encontrar un indicador de alta validez para el dominio en estudio⁽⁷⁾, o sea, para el aspecto del cuidado a ser medido.

El uso de medidas válidas y fidedignas propicia monitorizar la calidad del cuidado dispensado a los pacientes, identificar situaciones de riesgos evitables, así como, subsidiar la planificación de acciones correctivas, además de orientar estrategias y reajuste de metas por medio de acciones educativas y de valorización profesional. Se destaca que el uso de instrumentos válidos y fidedignos puede contribuir para el avance del conocimiento de la profesión y para el desarrollo de la teoría que ella sustenta⁽⁸⁾.

Confiabilidad implica fiabilidad y concordancia, que son aspectos importantes en el desarrollo de instrumentos de medida. Sus resultados ofrecen informaciones sobre la cantidad de error inherente a esta, lo que refleja la calidad de la medición⁽⁹⁾.

Concordancia es el grado en que los puntajes o clasificaciones son idénticos⁽⁹⁾ y fiabilidad el grado en que el resultado medido refleja el resultado verdadero, o sea, cuan libre una medida está de la variancia de los errores aleatorios⁽¹⁰⁾. Puede también ser definida como la proporción de la variancia en los puntajes de medición, que se deben a las diferencias en el puntaje verdad, no al error aleatorio⁽¹¹⁻¹²⁾. Siendo así, la fiabilidad estima la consistencia y estabilidad de la medida y aumenta a medida que el componente de error disminuye⁽¹¹⁾.

Considerando la preocupación histórica de la enfermería con la búsqueda por la calidad de la atención, señalada desde las acciones de cuidado de Florence Nightingale, la crisis de credibilidad actualmente asociada a los servicios de salud brasileños y el potencial del profesional enfermero en contribuir para la cambio de esta realidad, por medio de la mensuración de la calidad del cuidado prestado, queda evidente la necesidad de que las medidas de la calidad de la asistencia sean confiables.

El desarrollo de instrumentos evaluativos es complejo y objeto de estudio de varias disciplinas. En el contexto de la práctica de la enfermería se observan muchas críticas en cuanto a su uso en función de pretender mensurar constructos y conceptos considerados abstractos y subjetivos.

Sin embargo, se cree que esas críticas se deben a la falta de conocimiento sobre el proceso que participa de la concepción y validación de medidas evaluativas, lo que resulta en desmotivación para su desarrollo y, consecuentemente, impide el avance de la ciencia en el área.

Delante de ese escenario y dando continuidad al estudio que se inició con la validación de contenido de indicadores de calidad de la asistencia de enfermería⁽¹³⁾, se tiene, como pregunta de investigación: ¿El instrumento constituido por indicadores de calidad de cuidados básicos de enfermería atiende al requisito de confiabilidad para evaluación de la calidad de la asistencia prestada a pacientes adultos, internados en unidad médica quirúrgica de un hospital público de enseñanza del norte del Estado de Paraná?

La presente investigación tiene como objetivo comprobar la concordancia y fiabilidad entre evaluadores de quince indicadores de calidad de la asistencia de enfermería.

Metodología

Se trata de una investigación cuantitativa, metodológica, experimental y aplicada, realizada en tres fases (Validación Aparente, Test Piloto y Test de

Confiabilidad), desarrollada en un hospital terciario, de gran porte, público de enseñanza, del estado del Paraná, en Brasil. Se destaca que en el presente estudio el enfoque de la presentación de los resultados y discusión es el Test de Confiabilidad de la medida propuesta.

El estudio se inició con la Validación Aparente de un instrumento compuesto por quince indicadores de calidad, considerando el porcentaje mínimo de concordancia⁽¹⁴⁾ de 80%, a partir del instrumento propuesto por Vituri⁽¹³⁾, validado por la estrategia de validación de contenido por expertos y reformulado, con base en las necesidades sentidas por medio de su aplicación de rutina en la práctica diaria.

Para el procedimiento de Validación Aparente se seleccionó una muestra intencional de jueces, estudiantes en práctica del Servicio de Evaluación de la Calidad de la Asistencia de Enfermería del hospital en estudio. Este servicio funciona como campo de práctica curricular no obligatorio para alumnos del tercer y cuarto año del curso de graduación en enfermería, que desarrollan actividades de auditoría operacional y retrospectiva, evaluando la calidad de la asistencia de enfermería, por medio de la aplicación sistemática del instrumento de Vituri⁽¹³⁾.

El Test Piloto del instrumento de medida fue realizado por una muestra intencional de tres jueces, siendo dos enfermeros de la institución (uno asistencial y otro gerente en el nivel de dirección) más la investigadora. Este número fue definido a partir de los estudios de Crocker, Llabre y Miller⁽¹⁵⁾, ya que, cuanto mayor es el número de jueces, mayor es la heterogeneidad y, consecuentemente, menor la confiabilidad y concordancia⁽¹⁶⁾.

La comprensión de los indicadores, metodología de evaluación y aplicabilidad del instrumento fue comprobada por medio de la evaluación de una muestra aleatoria de quince pacientes adultos, internados en una unidad de clínica médica y quirúrgica.

Para determinación de la confiabilidad, entre las tres categorías de estimación de concordancia (estimativa de consenso, estimativa de consistencia y estimativa de evaluación/medición), se utilizó la estimativa de consenso, que tiene como base la suposición de que debe existir perfecta concordancia entre los jueces en el uso de escalas de puntuación para comportamientos observables, lo que caracteriza la misma comprensión del constructo⁽¹⁷⁾.

El método utilizado fue el de equivalencia, que es usado para evaluación de la confiabilidad mediante la aplicación del mismo instrumento por observadores diferentes, para medir los mismos fenómenos (fiabilidad entre evaluadores o precisión del evaluador)⁽¹¹⁾.

Este tipo de test es indicado, principalmente, para instrumentos clínicos, en que se depende del juzgamiento

del evaluador. Siendo así, es fundamental estimar la precisión por la correlación de los resultados obtenidos de cada sujeto, ya que, es evidente la variancia existente entre los evaluadores⁽¹¹⁾.

La concordancia y la confiabilidad de los indicadores fueron comprobadas por los mismos jueces que participaron del Test Piloto, ya que, ya estaban capacitados para el uso del instrumento, que fue aplicado de forma concomitante e independiente a una muestra de 33 pacientes, considerándose los quince pacientes ya evaluados anteriormente en el Test Piloto.

Para la evaluación de los indicadores de 1 la 11 y 15, el juez leyó el descriptor que define el estándar de calidad y señaló, en el instrumento de evaluación, si el aspecto del cuidado observado estaba adecuado o inadecuado en relación al estándar. Para los indicadores 12, 13 y 14, con base en la explicitación del estándar de calidad, cada juez registró la cantidad de chequeos y registros de señales vitales realizadas de forma adecuada y la cantidad inadecuada.

Para el tratamiento de los datos fue utilizada la estadística Kappa (k) para análisis de las variables categóricas – indicadores de 1 la 11 y 15 y el Coeficiente de Correlación Intraclase (CCI) para las variables continuas – indicadores 12, 13 y 14, considerando el Intervalo de Confianza (IC) de 95%^(12,18).

El Coeficiente de Kappa es una medida de asociación, que mide el grado de concordancia más allá de lo que sería esperado por el acaso, con pesos iguales para las discordancias⁽¹⁷⁾; el Fleis Kappa es una extensión de la estadística Kappa para evaluar la concordancia entre más de dos evaluadores, sin aplicación de ponderación⁽¹⁹⁾.

El Coeficiente de Correlación Intraclase (CCI) o coeficiente de reproducibilidad (R) es una estimativa de la fracción de la variabilidad total de medidas, debido a las variaciones entre los individuos⁽²⁰⁻²¹⁾.

A pesar de no ofrecer informaciones detalladas sobre la estructura de la concordancia y discordancia⁽²²⁾ la estadística Kappa es mejor que los porcentajes simple de concordancia⁽¹⁷⁾. Autores afirman que para la estimación de la confiabilidad entre observadores, la estadística Kappa y el Coeficiente de Correlación Intraclase (CCI) son los métodos más apropiados⁽⁹⁾.

Paralelamente a la estadística Kappa y CCI, también fueron calculados los porcentajes simple de concordancia para tener una idea detallada de la confiabilidad y concordancia⁽⁹⁾.

Los valores de Kappa fueron interpretados de la siguiente forma: menor que 0 como indicativo de *no agreement*; entre 0 y 0,19 *poor agreement*; 0,20 a 0,39 *fair agreement*; 0,40 a 0,59 *moderate agreement*;

0,60 a 0,79 *substantial agreement* y 0,80 a 1,00 *almost perfect agreement*⁽²³⁾. Para el CCI fueron utilizados los valores de 0,4 a 0,59 razonable reproducibilidad; 0,6 a 0,74 buena reproducibilidad y arriba de 0,74 excelente reproducibilidad⁽²⁴⁾.

Los datos categóricos (estadística k) fueron tratados utilizando el *software Lee*, del Laboratorio de Epidemiología y Estadística del Instituto Dante Pazzanese de Cardiología – Brasil y los datos continuos (CCI) por el *software BioEstat 5.0*. El *p-value* de <0.05 fue considerado estadísticamente significativo.

El estudio fue conducido en concordancia con todos los preceptos éticos, habiendo sido aprobado por la dirección de la institución y por el Comité de Ética en Investigaciones en que participan Seres Humanos de la Universidad, bajo el parecer n° 126/10, CAAEE n° 0113.0.268.000-10. Todos los sujetos de la

investigación, jueces y pacientes, firmaron el Término de Consentimiento Libre Esclarecido.

Resultados

Los resultados del procedimiento de Validación Aparente apuntaron para la validez aparente de todos los indicadores. En cuanto al Test Piloto, el parecer de los jueces fue favorable en lo que se refiere a la adecuación del instrumento en las cuestiones comprensión y aplicabilidad. Considerando que no hubo dudas ni sugerencias, el instrumento fue considerado apropiado para continuidad del Test de Confiabilidad Entre Evaluadores.

La Tabla 1, a continuación, presenta los resultados de la aplicación de la estadística Kappa para los indicadores de 1 a 11 y 15, además del Porcentaje de Concordancia simple para todo el conjunto de los indicadores.

Tabla 1 - Determinación de la concordancia entre evaluadores por la aplicación de la estadística Kappa a los indicadores de 1 a 11 y 15, Londrina, PR, Brasil, 2013

Indicador*	N†	% concordancia	Fleiss k	95% CI	p-valor	Concordancia
01	33	100	1,000	0,803-1,000	<0,001	Excelente
02	25	100	1,000	0,835-1,000	<0,001	Excelente
03	22	100	1,000	0,852-1,000	<0,001	Excelente
04	25	97	0,970	0,830-1,000	<0,001	Excelente
05a	10	97	0,956	0,795-1,000	<0,001	Excelente
05b	23	94	0,969	0,829-1,000	<0,001	Excelente
05c	04	100	1,000	0,847-1,000	<0,001	Excelente
06a	10	100	1,000	0,845-1,000	<0,001	Excelente
06b	23	97	0,969	0,829-1,000	<0,001	Excelente
06c	02	100	1,000	0,846-1,000	<0,001	Excelente
07	09	100	1,000	0,837-1,000	<0,001	Excelente
08	03	100	1,000	0,803-1,000	<0,001	Excelente
09	05	100	1,000	0,846-1,000	<0,001	Excelente
10	05	100	1,000	0,803-1,000	<0,001	Excelente
11	05	100	1,000	0,846-1,000	<0,001	Excelente
15	27	100	1,000	0,855-1,000	<0,001	Excelente

*Indicadores 1 – 11; 15

1. Identificación de la cama del paciente.
2. Identificación de riesgo para caída de la cama.
3. Identificación de accesos venosos periféricos.
4. Verificación de lesiones cutáneas post infiltrantes.
5. la – Identificación de equipos para infusión venosa (suero de mantenimiento).
b – Identificación de equipos para infusión venosa (suero para dilución de medicación).
c – Identificación de equipos para infusión venosa (suero – droga 1).
6. la – Identificación de frascos de infusión venosa – rótulo (suero de mantenimiento).
b – Identificación de frascos de infusión venosa – rótulo (suero para dilución de medicación).
c – Identificación de frascos de infusión venosa – rótulo (suero – droga 1).
7. Identificación del control de velocidad de infusión – escala con graduación.
8. Identificación de sondas gástricas – oro y naso gástricas para drenaje.
9. Fijación de la sonda vesical de demora.
10. Posición de la bolsa recolectora de diuresis, de la sonda vesical de demora.
11. Posición del prolongamiento distal de drenaje de la bolsa recolectora de diuresis – espita.
15. Elaboración de la prescripción diaria y completa por el enfermero.

†Número de pacientes en que se aplica la evaluación del indicador

Los valores de la Tabla 1 demuestran un porcentaje de concordancia simple superior a 80%, considerado adecuado para todos los indicadores evaluados⁽¹⁴⁾.

La estadística Kappa varió entre 0.956 a 1.000, o sea, los indicadores de 1 a 11 y 15 demuestran un excelente grado de concordancia, estadísticamente significativo – *p-valor* < 0.001⁽²³⁾.

Los indicadores que no obtuvieron concordancia total por la aplicación del Fleis K fueron: Indicador 4 – Verificación de lesiones cutáneas post infiltrantes ($k=0,970$, IC 95%: 0,830-1,000); 5a – Identificación de equipos para infusión venosa (suero de manutención) ($k=0,956$, IC 95%: 0,795-1,000); 5b – Identificación de equipos para infusión venosa (suero para dilución

de medicación) ($k=0,969$, IC 95%: 0,829-1,000) y 6b – Identificación de frascos de infusión venosa – rótulo ($k=0,969$, IC 95%: 0,829-1,000).

Los resultados de la aplicación del Coeficiente de Correlación Intraclase para los indicadores 12, 13 y 14 y el Porcentaje de Concordancia simple para el conjunto de los indicadores se presentan en la Tabla 2.

Tabla 2 - Determinación de la concordancia entre evaluadores por la aplicación del Coeficiente de Correlación Intraclase a los indicadores de 12, 13 y 14, Londrina, PR, 2013

Indicador*	N	% concordancia	CCI	95% CI	p-valor	Reproducibilidad†
12 - a	559	99,7	0,992	0,983-0,996	<0,001	Excelente
12 - b	81	98,1	0,980	0,959-0,990	<0,001	Excelente
13 - a	56	98,2	0,957	0,914-0,979	<0,001	Excelente
13 - b	44	97,8	0,951	0,903-0,976	<0,001	Excelente
14 - a	354	99,6	0,969	0,938-0,985	<0,001	Excelente
14 - b	64	97,5	0,859	0,732-0,929	<0,001	Excelente

*Indicadores 12 – 14

12. a – Comprobación de los procedimientos en la prescripción de enfermería (adecuado)

b – Comprobación de los procedimientos en la prescripción de enfermería (inadecuado)

13. a – Registro de verificación de las señales vitales prescritas (adecuado).

b – Registro de verificación de las señales vitales prescritas (inadecuado).

14. a – Comprobación de los procedimientos de enfermería en la prescripción médica (adecuado).

b – Comprobación de los procedimientos de enfermería en la prescripción médica (inadecuado).

†Sinónimo de confiabilidad

Para a los indicadores 12, 13 y 14 (Tabla 2), el Porcentaje de Concordancia simple varió entre 89,4% y 92,5%, lo que caracteriza una concordancia adecuada⁽¹⁴⁾. Los valores de CCI variaron entre 0,859 y 0,992 y los indicadores evaluados obtuvieron reproducibilidad/concordancia excelente y estadísticamente significativo – p-valor < 0,001⁽²⁴⁾.

El menor Coeficiente de Correlación Intraclase – CCI obtenido fue 0,859 (IC 95%: 0,732-0,929) para el indicador 14b – Comprobación de los procedimientos de enfermería en la prescripción médica (adecuado) y el mayor valor de CCI fue 0,992 (IC 95%: 0,983-0,996) para el indicador 12 la – Comprobación de los procedimientos en la prescripción de enfermería (adecuado), caracterizando reproducibilidad/concordancia excelente para los indicadores evaluados.

Discusión

La fiabilidad entre evaluadores comprobada por la estadística Kappa encontró valores que caracterizan concordancia excelente entre los jueces, en lo que se refiere al constructo, a los descriptores y a los criterios de evaluación de los indicadores comprobados.

Cabe resaltar que lo que está siendo medido no es la validez de los resultados, pero si, el grado de error en la medida, que se debe a las diferencias en el puntaje verdad^(8,10-11). Siendo así, es posible afirmar que los indicadores 4, 5-a, 5-b, 6-b, 12-a, 12-b, 13-a, 13-b, 14-a

y 14-b, que no obtuvieron resultados de Fleis K y CCI de 1,000, no son 100% precisos. Sin embargo, se resalta que, por mejor que sea un instrumento de medida, los puntajes obtenidos nunca estarán libres de errores⁽¹¹⁾.

Los errores que afectan las medidas son clasificados en dos subgrupos, los errores aleatorios, que derivan de factores que afectan la medición de la variable de forma accidental en toda la muestra, adicionando variabilidad a los datos, pero no afectando el desempeño medio para el grupo; y los errores sistemáticos, que son causados por cualquier factor que sistemáticamente afecte la medición de la variable en toda la muestra y que tienden a tener un efecto consistentemente positivo o negativo, por eso, es a veces considerado como sesgo de medición⁽¹¹⁾.

Se cree que los factores que posiblemente afectaron la precisión de la medida, contribuyendo para los errores de medición, sean factores personales transitorios⁽²⁵⁾, como prisa y fatiga, ya que uno de los jueces se encontraba participando del estudio después de la jornada diaria de trabajo. Además de eso, existió la referencia por parte de los dos jueces de dificultades en cuanto a la legibilidad de las escrituras referentes a los chequeos en las prescripciones médicas y de enfermería, lo que generó dudas en relación a la atención total al estándar de calidad adoptado.

Los valores obtenidos para el K y CCI de estos indicadores reproducen cuanto, los resultados de la aplicación de la medida, reflejan el resultado verdadero⁽¹⁰⁾. Inclusive no habiendo alcanzado un Fleis K de 1,000,

todavía así los indicadores 4, 5-a, 5-b, 6-b evidenciaron un grado de concordancia clasificado como excelente – *almost perfect agreement* – K 0,80 la 1,000(23). Los indicadores 12-a, 12-b, 13-a, 13-b, 14-a y 14-b también evidenciaron concordancia excelente, con valores de CCI arriba de 0,74⁽²⁴⁾.

Los resultados encontrados sugieren la fiabilidad entre evaluadores de los indicadores e instrumento de medida, indicando su precisión y la posibilidad de utilización para evaluación de la calidad de la asistencia de enfermería.

Cabe resaltar que confiabilidad y concordancia no son propiedades fijas de los instrumentos de medida, pero si, el producto de la interacción entre el instrumento/herramienta, los sujetos/objetos y el contexto de la evaluación⁽⁹⁾.

Para que esta propiedad no sea afectada en contextos diversos es importante establecer el control sobre las variables que interfieren en el proceso de medición. Para esto, los evaluadores deben ser capacitados sobre el constructo, descriptores, Índice de Conformidad Ideal, criterios de evaluación, así como la estandarización del procedimiento evaluativo.

Conclusión

A partir del estudio se concluyó que los quince indicadores de calidad de la asistencia de enfermería estudiados, ya validados anteriormente por la estrategia de validación de contenido, presentan excelente confiabilidad y reproducibilidad entre los ítems.

La concordancia y confiabilidad de los indicadores, estimadas por la estadística Kappa y Coeficiente de Correlación Intraclase demostraron la relevancia de este instrumento para la práctica clínica de la evaluación de la calidad de la asistencia de enfermería.

La confirmación de su confiabilidad comprueba la contribución del estudio para acrecentar evidencias de que el desarrollo de instrumentos válidos y fidedignos para la evaluación de la calidad de la asistencia de enfermería es posible e imprescindible para administración efectiva y eficaz de la asistencia, ya que, posibilita la identificación de riesgos evitables, subsidia la planificación de acciones correctivas y define estrategias para el reajuste de metas.

Delante de la evidencia de confiabilidad de los quince indicadores se cree que su utilización en otras instituciones de salud puede mejorar considerablemente la gestión de enfermería y, consecuentemente, la calidad de la asistencia y la seguridad de los pacientes.

Métodos de desarrollo y validación de sistemas de evaluación son extremadamente discutidos y empleados

en ciencias sociales y comportamentales y su aplicación en este estudio demuestra que sus principios pueden ser adaptados para la creación de instrumentos de evaluación de las prácticas de enfermería.

Las posibles limitaciones del estudio incluyen la técnica de muestreo intencional y el tamaño de la muestra de jueces. Lo que determinó la opción por esta técnica fue la inviabilidad en conducir el estudio caso fuesen sorteados enfermeros en su turno de trabajo.

La opción por el número mínimo de tres jueces se justifica por la dificultad en evaluar, de forma concomitante e independiente, a todos los pacientes acomodados en enfermarías de tres a seis camas, lo que podría resultar en tumulto, incomodidad y embarazo, además de actuar como contaminante situacional, favoreciendo el error de medición.

En relación a los indicadores comprobados, seguramente no abarcan todos los aspectos importantes de la asistencia de enfermería, pero ciertamente abarcan cuidados relevantes para la prevención de riesgos y altamente sensibles a la mejoría con inversiones simples, como estrategias educativas. Infelizmente, la selección de estos quince indicadores resultó en la exclusión de otros aspectos del cuidado relevantes, lo que puede ser una limitación del presente estudio.

Referencias

1. Rocha ESB, Trevizan MA. Quality management at a hospital's nursing service. Rev. Latino-Am. Enfermagem. 2009;17(2):240-5.
2. Kuwabara CCT, Évora YDM, Oliveira MMB. Risk Management in Technovigilance: construction and Validation of a Medical-Hospital Product Evaluation Instrument. Rev. Latino-Am. Enfermagem. 2010;18(5):943-51.
3. D'Innocenzo MANP, Cunha ICKO. O movimento pela qualidade nos serviços de saúde e enfermagem. Rev Bras Enferm. [Internet]. 2006. [acesso 23 abr 2013];59(1):84-8. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-71672006000100016&lng=en&nrm=iso&tlng=pt
4. Collier R. The challenges of quantifying quality. CMAJ. [Internet]. 2010. [acesso 23 abr 2013];182(5):E250. Disponível em: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2842849/pdf/182e250.pdf>
5. Takahashi AA, Barros ALBL, Michel JLMS, Mariana F. Difficulties and facilities pointed out by nurses of a university hospital when applying the nursing process. Acta Paul Enferm. [Internet]. 2008. [acesso 23 abr 2013];21(1):32-8. Disponível em: <http://www.scielo.br/pdf/ape/v21n1/04.pdf>

6. Vieira APM, Kurcgant, P. Quality indicators of the management of human resources in nursing: point of view of registered nurses. *Acta Paul Enferm.* [Internet]. 2010. [acesso 23 abr 2013];23(1):11-5. Disponível em: <http://www.scielo.br/pdf/ape/v23n1/02.pdf>
7. Nakrem S, Vinsnes AG, Harkless GE, Paulsen B, Seim A. Nursing sensitive quality indicators for nursing home care: international review of literature, policy and practice International. *J Nurs Stud.* 2009;46:848-57.
8. Gillespie BM, Polit DF, Hamlin L, Chaboyer W. Developing a model of competence in the operating theatre: Psychometric validation of the Perceived Perioperative Competence Scale-Revised. *Int J Nurs Stud.* [Internet]. 2012. [acesso 23 abr 2013];49(1):90-101. Disponível em: [http://www.journalofnursingstudies.com/article/S0020-7489\(11\)00301-4/fulltext](http://www.journalofnursingstudies.com/article/S0020-7489(11)00301-4/fulltext)
9. Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol.* [Internet]. 2011. [acesso 23 abr 2013];64(1):96-106. Disponível em: <http://www.rygforskning.dk/sites/default/files/files/articles/Kottner%20el%20al%20%202011.pdf>
10. Hora HRM, Monteiro GTR, Arica J. Confiabilidade em questionários para qualidade: um estudo com o Coeficiente Alfa de Cronbach. *Prod Prod.* [Internet]. 2010. [acesso 23 abr 2013];11(2):85-103. Disponível em: <http://seer.ufrgs.br/ProdutoProducao/article/view/9321/8252>
11. Trochim WMK. Research methods: knowledge bases. [Internet]. 2nd. ed. Cincinatti (OH): Atomic Dog Publishing; 2006. [acesso 23 abr 2013]. Disponível em: <http://socialresearchmethods.net/kb/>
12. Wuensch KL. The Intraclass Correlation Coefficient. Karl Wuensch's Statistics Lessons [Internet]. Greenville (USA): East Carolina University - Department of Psychology; [2010]; [atualizada em 20 abr 2013; acesso 23 abr 2013]. Disponível em: <http://core.ecu.edu/psyc/wuenschk/StatsLessons.htm>
13. Vituri DW, Matsuda LM. Content validation of quality indicators for nursing care evaluation. *Rev Esc Enferm USP.* [Internet]. 2009. [acesso 23 abr 2013];43(2):429-37. Disponível em: http://www.scielo.br/pdf/reeusp/v43n2/en_a24v43n2.pdf
14. Westmoreland D, Wesorick B, Hanson D, Wyngarden K. Consensual validation of clinical practice model practice guidelines. *J Nurs Care Qual.* 2000;14(4):16-27.
15. Crocker L, Llabre M, Miller MD. The generalizability of content validity ratings. *J Educ Measure.* 1988; 25(4):287-99.
16. Lilford RJ, Mohammed MA, Braunholtz D, Hofer TP. The measurement of active errors: methodological issues, *Qual Saf Health Care.* [Internet]. 2003. [acesso 30 abr 2012];12 Suppl 2:8-12. Disponível em: http://qualitysafety.bmj.com/content/12/suppl_2/ii8.full.pdf+html
17. Stemler SE. A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Prac Assess Res Eval.* [Internet]. 2004. [acesso 15 abr 2012];9(4). Disponível em: <http://pareonline.net/getvn.asp?v=9&n=4>
18. Haley SM, Osberg JS. Kappa Coefficient Calculation Using Multiple Ratings Per Subject: A Special Communication. *Phys Ther.* [Internet]. 1989. [acesso 15 abr 2012];69:970-4. Disponível em: <http://ptjournal.apta.org/content/69/11/970>
19. Chang A. Cohen's and Fleiss's Kappa program Concordance in ordinal data. In: Chang A. StatTools Home Page [Internet]. Queensland: Austrália; 2011. [acesso 20 abr 2012]; Disponível em: http://www.stattools.net/CohenKappa_Pgm.php#Fleiss%27s%20Kappa%20from%20rating%20scores
20. Chang A. Intraclass correlation for parametric data Introduction and explanation. In: Chang A. StatTools Home Page [Internet]. Queensland: Austrália; 2011 [acesso 15 abr 2012]. Disponível em: http://www.stattools.net/ICC_Exp.php
21. Wilson-Genderson M, Broder HL, Phillips C. Concordance between caregiver and child reports of children's oral health-related quality of life. *Community Dent Oral Epidemiol.* 2007;35 Suppl 1:32-40.
22. Perroca MG, Gaidzinki RR. Assessing the interrater reliability of an instrument for classifying patients - kappa quotient. *Rev Esc Enferm USP.* [Internet]. 2003. [acesso 23 abr 2013];37(1):72-80. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0080-62342003000100009
23. Zegers M, Bruijne MC, Wagner C, Groenewegen PP, Wal GVD, Vet HCW. The inter-rater agreement of retrospective assessments of adverse events does not improve with two reviewers per patient record. *J Clin Epidemiol.* 2010;63(1):94-112.
24. Fleiss JL. Reliability of measurement. In: Fleiss JL. *The Design and Analysis of Clinical Experiments.* New York (USA): John Wiley; 1999. p. 1-31.
25. Polit DF, Beck CT. Assessing data quality. In: Polit DF, Beck CT. *Nursing research: principles and methods.* 7th ed. Philadelphia (USA): Lippincott Williams & Wilkins; 2004. p. 413-48.

Recibido: 26.4.2013

Aceptado: 3.9.2013