



ORIGINAL

Desarrollo y validación de estándares para unidades de calidad de centros sanitarios



A. Aloy-Duch^{a,*}, M. Santiñà Vila^b, F. Ramos-D'Angelo^c, L. Alonso Calo^d,
M.E. Llaneza-Velasco^e, B. Fortuny-Organs^f y A. Apezetxea-Celaya^g

^a Director de Calidad y Planificación, Hospital General de Granollers, Granollers, Barcelona, España

^b Sociedad Española de Calidad Asistencial (Past President), Investigador del Institut de Recerca Biomèdica August Pi i Sunyer, Barcelona, España

^c Coordinador de Calidad, Hospital Royo Villanova, Zaragoza, España

^d Responsable de Calidad y Seguridad del Paciente, Hospital Universitario Central de Asturias, Área Sanitaria IV SESPA, Oviedo, Asturias, España

^e Servicio de Microbiología, Hospital Universitario Central de Asturias, Presidenta de la Asociación Calidad Asistencial del Principado de Asturias - PASQAL, Oviedo, Asturias, España

^f Unidad de Calidad, Hospital Marina Salud, Denia, Alicante, España

^g Unidad de Calidad, Organización Sanitaria Integrada Bilbao-Basurto, Osakidetza, Servicio Vasco de Salud, Bilbao, Vizcaya, España

Recibido el 21 de julio de 2023; aceptado el 28 de septiembre de 2023

PALABRAS CLAVE

Calidad asistencial;
Estándares;
Estudio de validación;
Técnica Delphi;
Unidades de calidad

Resumen

Objetivo: En España, las unidades de calidad asesoran a los centros sanitarios sobre la metodología de la calidad asistencial. Un sistema de estándares de buena práctica, podría facilitar el funcionamiento y mejorar los resultados de estas unidades. El objetivo fue describir el proceso de elaboración, consenso y validación de estándares, para servir de guía y orientar adecuadamente las funciones, los resultados y la mejora continua de las unidades de calidad en los centros sanitarios.

Material y métodos: Se realizó un estudio cualitativo/cuantitativo, prospectivo y transversal, aplicando el método Metaplan, la técnica e-Delphi y un estudio de simulación. Los participantes eran coordinadores de estas unidades, pertenecientes a 14 comunidades autónomas españolas, y distribuidos en 4 paneles de expertos. Estos, consensuaron los estándares y evaluaron los distintos tipos de validez.

* Autor para correspondencia.

Correo electrónico: aaloyduch@gmail.com (A. Aloy-Duch).

Resultados: Los 204 estándares propuestos por un comité científico se redujeron a 157 con el Metaplan, a 110 con el e-Delphi y a 96 con la revisión final del comité (consenso en el 87,3%, validez de contenido). La validez de constructo mostró un alfa de Cronbach > 0,7 ($p < 0,001$); se reafirmó la validez de contenido en el taller de simulación (80% se «entendían»; $p < 0,001$; y, existían «evidencias documentales» en el 84%; $p < 0,001$); se aceptó la validez de apariencia (75% de «relacionadas con dimensiones de la calidad»; $p < 0,001$); y se comprobó la validez de criterio con una sensibilidad del 84,2%, especificidad del 98,3% e índice kappa de 0,84.

Conclusiones: Se han desarrollado unos estándares válidos para unidades de calidad de centros sanitarios.

© 2023 FECA. Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

KEYWORDS

Health care quality;
Standards;
Validation study;
Delphi technique;
Quality units

Development and validation of standards for quality units of health centers

Abstract

Objective: In Spain, the Quality Units advise health centres, services and professionals on the methodology of continuous improvement of the quality of care. A system based on good practice standards could provide these units with a tool to improve their results and value their work. The objective was to develop, agree on and validate standards, to properly guide and orient the functions, results and continuous improvement of the Quality Units in health centers.

Material and methods: A qualitative-quantitative, prospective and cross-sectional study was carried out, applying the Metaplan method, the e-Delphi technique and a simulation study. The participants were coordinators of these units, belonging to 14 Spanish Autonomous Communities and distributed in four experts' panels. They agreed on the standards to be used and evaluated the different types of validity.

Results: The 204 standards proposed by the scientific committee were reduced to 157 with Metaplan, to 110 with e-Delphi, and to 96 with the committee's final review (87.3% consensus, content validity). The construct validity showed a Cronbach's alpha > 0.7 ($P < .001$); the validity of content was reaffirmed in the simulation workshop (80% "understood" each other, $P < .001$; and there was "documentary evidence" in 84%, $P < .001$); face validity was accepted (75% "related to quality dimensions", $P < .001$); and the validity of the criteria was verified with a sensitivity of 84.2%, a specificity of 98.3%, and a kappa index of 0.84.

Conclusions: Valid standards have been developed for Quality Units in health centers.

© 2023 FECA. Published by Elsevier España, S.L.U. All rights reserved.

Introducción

Avedis Donabedian, definió la calidad asistencial como «el nivel de utilización de los medios más adecuados para conseguir las mayores mejoras en la salud»¹. Las variadas acepciones de la definición de «calidad» convergen en las denominadas «dimensiones», sintetizadas en las siguientes: experiencia del paciente, seguridad clínica y eficiencia², las cuáles deberían formar parte del contenido temático de cualquier futuro estándar.

Un estándar se define como aquel criterio que sirve de modelo, con un grado de cumplimiento aceptable, que se debe exigir a un indicador para alcanzar el nivel de calidad deseado³. Coexiste, además, otra acepción de la palabra «estándar»⁴, referida a aquella recomendación que debería alcanzar un equipo para incrementar la calidad de la asistencia; y esta última es, precisamente, la definición empleada en este trabajo a modo de declaración de «criterios de buena práctica». «Recomendaciones» que deben categorizarse por contenidos y temáticas para establecer una lógica organizativa, y que otros autores y sociedades científicas ya han

publicado inspiradas en los 3 conceptos de estructura, proceso y resultado⁵ y en los «Principios de la calidad»⁶.

Los estándares deben ser válidos, entendiendo por validez el grado en que estos miden realmente lo que se pretende medir, asegurando su utilidad con los 4 tipos de validez, las de: «contenido», «criterio», «constructo» y «apariencia»⁷.

En España, se han constituido históricamente unidades de calidad (UCAL) en los distintos ámbitos asistenciales de atención primaria, hospitalaria, intermedia/sociosanitaria y salud mental⁸. Estos equipos abordan la calidad desde una perspectiva de gestión, con dependencia directa de sus direcciones ejecutivas, para asesorar a los profesionales sobre la metodología, los objetivos relacionados con las dimensiones de la calidad y las actividades de mejora.

En el sector salud, se han desarrollado estándares para orientar a los profesionales sobre la asistencia en diversas enfermedades, y para evaluar la gestión de la calidad de las organizaciones sanitarias⁹. Sin embargo, no se han desarrollado estándares que permitan a las UCAL mejorar sus propios procesos y resultados. La revisión exhaustiva

de diferentes bases de datos y plataformas científicas de habla hispana y anglosajona, no han identificado estándares dirigidos específicamente a las UCAL.

La Sociedad Española de Calidad Asistencial (SECA)¹⁰ ha trabajado desde su fundación en asesorar a los profesionales en todos aquellos aspectos de la calidad asistencial. La SECA constató la necesidad de elaborar unos estándares de calidad, por lo que en su Plan Estratégico 2016-2020¹¹ estableció una línea de acción en este sentido.

El objetivo de este trabajo fue describir el proceso de elaboración, consenso y validación de un conjunto de estándares, para servir de guía y orientación de las funciones, resultados y mejora continua a las UCAL en los centros sanitarios.

Material y métodos

Tipo de estudio, ámbitos y población

Se realizó, entre 2015 y 2017, un estudio prospectivo y transversal cualitativo/cuantitativo aplicando, por fases, el método Metaplan, la técnica e-Delphi y un estudio/taller de simulación. Participaron, voluntariamente, profesionales con responsabilidad en UCAL de centros sanitarios, representantes de 14 comunidades autónomas (CC. AA.) españolas, con más de 2 años de experiencia como coordinadores de calidad, actuando en algunos de los siguientes ámbitos de atención: primaria, hospitalaria, intermedia/sociosanitaria o salud mental. Los participantes se distribuyeron en 4 paneles de expertos para trabajar en las 4 fases del trabajo:

- Propuesta de estándares por el comité científico y resultados (usando el método Metaplan [panel de expertos 1]).
- Estándares resultantes de las dos rondas del estudio e-Delphi (panel de expertos 2).
- Revisión y categorización de los estándares (panel de expertos 3).
- Estándares resultantes del estudio/taller de simulación (panel de expertos 4).

Comité científico, paneles de expertos, actividades, tipos de validez evaluadas y análisis estadísticos

Un comité científico formado por 12 expertos, 9 profesionales médicos y 3 profesionales de enfermería, monitorizó el estudio y redactó, en 2015, una primera propuesta de estándares basados en su experiencia y en la revisión bibliográfica realizada sobre estándares aplicables a las UCAL, categorizados provisionalmente en 9 categorías; sus miembros, designados por la SECA, acreditaban más de 10 años de experiencia en calidad asistencial y en el liderazgo de las UCAL. Ese mismo año, se constituyó el primer panel de expertos empleando el método Metaplan¹² para discutir la propuesta inicial del comité. Con los estándares seleccionados en el Metaplan y revisados por el comité, se organizó el segundo panel de expertos, con el que se realizó la técnica cualitativa Delphi en 2 rondas¹³, con el envío de un cuestionario por ronda a través de una plataforma web

en línea (e-Delphi). Para el e-Delphi no se definió un tamaño muestral mínimo, sino que la selección se realizó bajo el criterio de oportunidad proponiendo una participación superior a 40 profesionales. Los participantes puntuaron los estándares mediante una escala de Likert de 1 a 9 puntos, con respuestas anónimas en las 2 rondas del e-Delphi. Para la evaluación de la opinión respecto a cada estándar y la validez de contenido¹⁴, se utilizó el porcentaje por valores de la escala Likert, y las siguientes pruebas con sus valores umbral óptimos entre paréntesis: el porcentaje ($\geq 70\%$), la media (≥ 7) con sus intervalos de confianza, la desviación estándar, la mediana (≥ 8), el rango o amplitud intercuartílica (RI) (≤ 2), el índice de validez de contenido (IVC) de Lawshe ($> 0,50$) y el coeficiente de validez de contenido de Aiken (V de Aiken) ($> 0,70$). Se consideró un estándar como consensuado (alto o medio), cuando cumplía con al menos 3 de las pruebas estadísticas anteriores. Los estándares del e-Delphi fueron revisados y categorizados, según la temática de cada estándar, por un tercer panel de expertos según los 3 ejes de estructura (categorías 1 a 4), proceso (categoría 5) y resultado (categorías 6 a 9)⁵; permitiendo comprobar la validez de constructo, a partir del cálculo del coeficiente alfa de Cronbach y del coeficiente de correlación intraclass¹⁵.

Finalmente, se organizó un estudio/taller de simulación para validar los estándares con un cuarto panel de expertos. En el taller se reevaluó la validez de contenido con pruebas binomiales, preguntando a los participantes si cada estándar se «entendía» y si se podría justificar mediante pruebas documentales; para la validez de apariencia, se preguntó si se podía relacionar cada estándar con alguna de las dimensiones de la calidad; y para evaluar la validez de criterio, en la simulación, el comité científico elaboró artificialmente un «patrón oro» o «semi-oro»^{16,17} propio, con la creación de un conjunto de indicadores positivos o negativos, a criterio del comité, en un hipotético centro sanitario virtual, para que los participantes evaluarán el cumplimiento de ese centro teórico con los estándares surgidos en el e-Delphi; a los resultados se les realizó un análisis de sensibilidad y especificidad, una Chi-cuadrado de Pearson y un índice kappa.

Se asumió una significación estadística cuando el p-valor era $\leq 0,05$, realizándose los cálculos con la versión 28^a del programa SPSS® Statistics de IBM.

Resultados

Comité científico, primer panel de expertos en el Metaplan y revisión

En la tabla 1, se describen las características demográficas de los participantes en los paneles de expertos. En la tabla 2, se detalla la categorización en cada fase del proceso de elaboración de los estándares.

Técnica e-Delphi. Segundo panel de expertos

Como resultado del promedio de las 2 rondas del e-Delphi (tabla 2), del total de los 110 estándares, se consensuaron 87 (79%); estos, junto a los 23 (21%) restantes no consensuados, se pusieron a disposición del comité para su revisión y decisión final. Se describieron 110 estándares con el resultado

Tabla 1 Características de los participantes en los 4 paneles de expertos y fases

Características	Método Metaplan	Técnica e-Delphi	Revisión de estándares	Estudio-taller de simulación
Año	2015	2015	2016	2017
Participantes (n.º)	99	53	40	63
Media de edad ± DE (años)	48,0 ± 7	48,4 ± 8	48,1 ± 8	48,7 ± 7
Mujeres (%)	64	62	65	63
CC. AA. participantes (n.º)	14	14	14	14
Categoría profesional ^a (%)	87	81	82	84
Ámbito de trabajo ^b (%)	62	61	63	62
Unidad de calidad estructurada como equipo (%)	80	91	90	85
Coordinadores o responsables (%) ^c	61	62	63	62
Jornada laboral parcial (%)	59	53	55	54

CC. AA.: comunidades autónomas; DE: desviación estándar.

^a Médicos, enfermeras o técnicos.^b Atención especializada aguda o atención intermedia/sociosanitaria. Resto coordinadores de calidad en atención primaria y en salud mental.^c Denominación de los responsables de la unidad de calidad.**Tabla 2** Categorización y distribución de los estándares en las distintas fases del estudio

Ejes de categorización	Categorías	Estándares propuestos por el comité científico (%)	Estándares resultantes del Metaplan (%)	Estándares resultantes del Delphi (2 rondas) (%)	Estándares resultantes de la revisión del comité científico y evaluados en el estudio de simulación (%)
Estructura	Directivo	15	12	18	9
	Estratégico	19	17	16	11
	Profesionales	8	7	5	9
	Recursos y alianzas	1	2	8	6
Proceso	Gestión por procesos	34	29	24	19
Resultado	Resultados en los pacientes	8	10	6	4
	Resultados en los profesionales	1	4	5	13
	Resultados sociales	8	11	11	6
	Resultados principales	6	8	7	22
Total de estándares (n.º)		204	157	110	96

n.º: número.

del consenso agrupados de las 2 rondas del e-Delphi ([tabla 3](#)); destacó un bajo nivel de acuerdo en la categoría 6 (Resultados en los pacientes), y medio en las categorías 4 (Recursos y alianzas) y 8 (Resultados sociales). Se evaluó la validez de constructo confirmado la consistencia de los 2 coeficientes, con valores superiores a 0,70 ([tabla 4](#)), excepto en el coeficiente de correlación intraclass en el apartado de «Resultados en los profesionales» con un valor de 0,66.

Revisión por el tercer panel de expertos

De los 110 estándares se obtuvo, con la revisión del tercer panel de expertos, una propuesta final de consenso de 96 (87,3%) de estándares. De los que 34 (35%) procedían de una fusión literal a partir de los resultados del e-Delphi y de las propuestas del segundo y tercer panel de expertos; y, 62 (65%) procedían directamente de las 2 rondas del e-Delphi.

Tabla 3 Resultado del consenso en los 110 estándares obtenidos en las 2 rondas agrupadas del e-Delphi

Categorías	Porcentaje de puntuación entre 7 y 9 (sobre 9 en la escala de Likert) ^a	ME	Intervalos de confianza inferiores 95%	Intervalos de confianza superiores 95%	DE	MN	RI	Lawshe	V de Aiken	Nivel de consenso ^b global
Categoría 1 (estándares del 1.1. al 1.20.)	71-91%	6,4-8,1	6,2-7,8	7,2-8,4	1,0-2,0	7,0-8,0	1,0-2,0	0,55-0,96	0,60-0,79	Alto
Categoría 2 (estándares del 2.1. al 2.18.)	74-96%	7,1-8,3	6,2-8,0	7,1-8,5	0,9-2,1	7,0-8,0	1,0-2,0	0,51-0,81	0,69-0,81	Alto
Categoría 3 (estándares del 3.1. al 3.5.)	78-94%	7,4-8,1	5,9-7,8	7,9-8,4	1,2-1,9	8,0	1,0-2,0	0,58-0,89	0,71-0,79	Alto
Categoría 4 (estándares del 4.1. al 4.9.)	67-96%	6,4-8,0	5,9-7,6	7,0-8,3	1,3-2,4	7,0-8,0	1,0-2,0	0,66-0,92	0,61-0,77	Medio
Categoría 5 (estándares del 5.1. al 5.26.)	80-100%	7,2-8,6	6,0-8,4	7,2-8,8	0,6-2,2	8,0-9,0	1,0-2,0	0,51-0,81	0,70-0,84	Alto
Categoría 6 (estándares del 6.1. al 6.7.) ^c	85%	7,7	5,6-7,3	6,7-8,1	1,6-2,3	8,0	2,0	0,74	0,74	Bajo
Categoría 7 (estándares del 7.1. al 7.5.)	76-94%	7,1-8,0	5,9-7,6	6,8-8,4	1,3-2,0	8,0	1,0-2,0	0,55-0,89	0,68-0,78	Alto
Categoría 8 (estándares del 8.1. al 8.12.)	71-80%	6,9-7,4	5,0-7,0	7,0-7,9	1,5-2,0	7,0-8,0	1,0-2,0	0,51-0,61	0,66-0,72	Medio
Categoría 9 (estándares del 9.1. al 9.8.)	79-96%	7,4-8,2	6,9-7,9	7,9-8,6	1,2-1,9	8,0	1,0-2,0	0,66-0,96	0,72-0,80	Alto

DE: desviación estándar; Lawshe: índice de validez de contenido (IVC); ME: media; MN: mediana; RI: rango o amplitud intercuartílica; V de Aiken: coeficiente de validez de contenido de Aiken.

Umbrales mínimos estadísticos para los valores de consenso: Likert $\geq 70\%$, ME ≥ 7 , MN ≥ 8 , RI ≤ 2 , Lawshe $\geq 0,5$; V de Aiken $\geq 0,70$.

Categorías y umbrales de los parámetros estadísticos: 1: Enfoque directivo (80% de estándares: alto o medio); 2: Enfoque estratégico (89% de estándares: alto o medio); 3: Profesionales (100% de estándares: alto o medio); 4: Recursos y alianzas (67% de estándares: alto o medio); 5: Gestión por procesos (88% de estándares: alto o medio); 6: Resultados en los pacientes (14% de estándares: alto o medio); 7: Resultados en los profesionales (80% de estándares: alto o medio); 8: Resultados sociales (67% de estándares: alto o medio); 9: Resultados principales de los equipos de calidad (100% de estándares: alto o medio).

^a Resultados sobre los estándares consensuados (valores mínimos y máximos).

^b Consenso: ALTO o MEDIO (≥ 3 de cumplimiento de los parámetros de los umbrales mínimos estadísticos); BAJO (≤ 2 de cumplimiento de los parámetros de los umbrales anteriores).

^c Solo un estándar consensuado.

Tabla 4 Validez de constructo. Coeficientes alfa de Cronbach e intraclass

Códigos de los estándares	Categorías (enfoques y resultados) de estándares	N.º de estándares	Coeficiente alfa de Cronbach (estándar > 0,70)	Coeficiente de correlación intraclass			
				Valor	Intervalo de confianza del 95%		Valor de p
					Límite inferior	Límite superior	
1.	Directivo	20	0,94	0,93	0,90	0,96	< 0,001
2.	Estratégico	18	0,94	0,93	0,90	0,96	< 0,001
3.	Profesionales	5	0,89	0,88	0,83	0,93	< 0,001
4.	Recursos y alianzas	9	0,82	0,80	0,70	0,87	< 0,001
5.	Gestión por procesos	26	0,95	0,95	0,93	0,97	< 0,001
6.	Resultados en los pacientes	7	0,82	0,81	0,72	0,88	< 0,001
7.	Resultados en los profesionales	5	0,71	0,66	0,48	0,78	< 0,001
8.	Resultados sociales	12	0,85	0,82	0,74	0,88	< 0,001
9.	Resultados principales	8	0,92	0,91	0,87	0,94	< 0,001
Total	Conjunto de estándares	110	0,99	0,98	0,97	0,99	< 0,001

p: significación estadística.

Además, el panel confirmó la categorización de los estándares en los 3 conceptos de estructura, proceso y resultado, y 9 sub-apartados o categorías ([tabla 2](#)), según la temática de cada estándar para las UCAL. En las 4 primeras categorías del eje «estructura» se establecieron distintos enfoques: el «directivo» con los planes y responsabilidades sobre la calidad; el «estratégico» con la estrategia institucional de calidad en indicadores y acciones de mejora; la participación de los «profesionales» en grupos de mejora y acciones formativas sobre calidad; y, los «recursos y alianzas» necesarias para las UCAL, respecto a presupuestos y relación con otras instituciones. En el eje y categoría de procesos, se incluyeron los estándares referentes al funcionamiento de la unidad, las guías de trabajo y las líneas estratégicas de calidad. En las 4 categorías del eje de «resultados», se incluyó la relacionada con los «pacientes», su participación y formación en calidad; con los «profesionales», formación, competencia y proyectos de investigación en calidad; la participación en la «sociedad» con la colaboración con otras instituciones; y, finalmente, los resultados «principales» de las UCAL, con indicadores evaluados, objetivos y acciones de mejora realizadas.

Algunos ejemplos a destacar por categorías serían: a) en el eje de «estructura» y «enfoque directivo», que «existe un equipo, estructural y funcionalmente a modo de equipo o UCAL, que gestiona, coordina y asesora metodológicamente a la organización en los planes y actividades de calidad» (estándar 1); b) en el eje y categoría de «proceso», que «los procesos de la UCAL tienen responsables/propietarios,

que asumen el desarrollo de la estrategia de la gestión de cada proceso de la unidad» (estándar 36); y c) en el eje de «resultados» y resultados «principales» de los equipos de calidad, que «se mide, periódica y anualmente, el grado de cumplimiento de los objetivos internos planificados por la UCAL respecto a los resultados, las tendencias, y la comparativa interna o externa con otros centros, si existiera» (estándar 76).

Estudio de simulación. Cuarto panel de expertos

Se constituyó y reunió un cuarto panel de expertos. Las características demográficas de los participantes en el estudio/taller de simulación fueron similares a las fases anteriores ([tabla 1](#)). En la simulación se aseguró la validez de contenido y de apariencia ($p < 0,001$), a partir de los resultados a tres preguntas dicotómicas ([tabla 5](#)). Además, se comprobó la validez de criterio con una sensibilidad (84,2%) y especificidad (98,3%) adecuadas, cuando se les preguntó a los participantes si el estándar se cumplía (Sí/No) para cada caso planteado en el centro virtual, respecto al «patrón oro» creado por el comité científico (índice de kappa de 0,84) ([tabla 6](#)).

Por último, se describen los 96 estándares resultantes de la revisión del comité sobre los obtenidos del e-Delphi, evaluados en el taller de simulación; en el [Anexo](#) se describen los estándares literales finales.

Tabla 5 Validez de contenido y apariencia en el estudio de simulación. Pruebas binomiales

	Pruebas binomiales								
	N.º	Media	Desviación estándar	Categoría	n.º	Porcentaje observado	Porcentaje de prueba	Valor de p	
1. ¿Se entiende el estándar? (Sí/No) (validez de contenido)	96	0,80	0,40	Sí No	77 19	80 20	50	< 0,001	
2. ¿Se puede justificar el estándar con evidencias documentales? (Sí/No) (validez de contenido)	96	0,84	0,37	Sí No	81 15	84 16	50	< 0,001	
3. ¿El estándar refleja alguna dimensión de la calidad asistencial? (Sí/No) (validez de apariencia)	96	0,73	0,45	Sí No	72 24	75 25	50	< 0,001	

N.º: estándares totales; n.º: número de respuestas; p: significación estadística exacta (bilateral).

Discusión

Limitaciones del estudio

Se basan, en primer lugar, en el tamaño de las distintas muestras, si bien se consideró que más de 50 profesionales por fase podrían ser suficientes, al ser un estudio principalmente cualitativo. En segundo lugar, por un posible sesgo de selección de los expertos por su participación voluntaria si bien, en contraposición, se destacaría el anonimato de las respuestas y la alta representación geográfica. Y, en tercer lugar, la escasez de estudios previos en esta área de investigación.

Estándares como «criterios de buena práctica»

Uno de los conceptos clave en la elaboración de los estándares ha sido considerarlos como «criterios de buena práctica», tal y como aplican muchas de las sociedades científicas para sus estándares¹⁸. La metodología empleada en este trabajo ha sido compleja: por el número de paneles de expertos utilizando estrategias de reunión y técnicas diversas participativas, tanto presenciales como «en línea»; por las fases secuenciales del Metaplan, del e-Delphi y del taller de simulación, para evaluar los tipos de validez; y, por compaginar el carácter cualitativo con las herramientas estadísticas cuantitativas, que dieran consistencia y significación a los resultados.

Paneles de expertos

A diferencia de otros estudios^{19,20}, fue relevante el número de coordinadores participantes, de diferentes UCAL de la mayoría de CC. AA. El perfil profesional «tipo» era el de una mujer, médica o enfermera hospitalaria, con una media de edad de 48 años; si bien destaca la presencia de otras profesiones como farmacéuticos, biólogos o economistas. Por otra parte, alrededor del 15% de los coordinadores no disponían de una UCAL estructurada como equipo, sino que su figura era única, con dependencia funcional de

alguna de las direcciones del centro. A pesar de la participación de UCAL pertenecientes a los cuatro ámbitos asistenciales, las 3/5 partes de ellas pertenecían a la atención especializada o intermedia/sociosanitaria. Sorprende que la mitad de sus responsables no tuvieran una dedicación completa de su jornada a las UCAL. Esta última característica se diferencia de otros sectores de servicios como, por ejemplo, en las empresas alimentarias, en las cuales existe una alta demanda de profesionales a jornada completa²¹. ¿Por qué en el sector salud, tan importante para los ciudadanos, no hay demanda de expertos en calidad asistencial, y la dedicación de sus profesionales no es mayoritariamente a jornada completa?

Consenso con la técnica Delphi

La técnica «en línea» e-Delphi, ampliamente aplicada en otros estudios, permitió el consenso de los estándares²², potenció la diversidad de los participantes, disminuyó el tiempo del proyecto, facilitó el número de rondas, y aumentó su participación simultánea desde diferentes áreas geográficas²³. Con el e-Delphi, se consensuaron la mayoría de los estándares con niveles superiores al 80%, si bien hubo menos acuerdo, en primer lugar, con los resultados en los pacientes, planteándose el siguiente reto futuro: ¿cuál debiera ser el impacto directo o indirecto de las UCAL en los pacientes?; en segundo lugar, hubo consensos moderados en los estándares de gestión económica de las mismas, por lo que las UCAL precisarían probablemente de una mayor profesionalización; y finalmente, destaca la necesidad de un mayor consenso en los estándares de ámbito social, como la participación y colaboración de los profesionales de las UCAL en otras instituciones externas, su presencia en los medios de comunicación, y el incremento de publicaciones sobre calidad.

A diferencia de otros estudios realizados²⁴, el trabajo aquí presentado destaca por la retroalimentación informativa a los panelistas sobre sus respuestas, el elevado número de expertos, y la rigurosidad de los estudios de validez aplicados. Además, Diamond et al.²⁵, en una revisión sistemática con 100 estudios Delphi propusieron, como así

Tabla 6 Validez de criterio. Análisis de la sensibilidad y especificidad, Chi-cuadrado de Pearson e índice kappa de concordancia

Validez de criterio. Análisis de sensibilidad y especificidad			«Patrón oro». Datos de referencia del comité científico		Total
	No	Recuento	No	Sí	
Cumplimiento del estándar	No	Recuento	57	6	63
	Sí	Porcentaje	90,5	9,5	100,0
		Recuento	1	32	33
		Porcentaje	3,0	97,0	100,0
		Recuento	58	38	96
Total		Porcentaje	60,4	39,6	100,0
		Intervalo de confianza del 95%			
		Límite inferior	Límite superior		
Sensibilidad	84,2%	69,6%	92,6%		
Especificidad	98,3%	90,9%	99,7%		
Valor predictivo positivo	97,0%	84,7%	99,5%		
Valor predictivo negativo	90,5%	80,7%	95,6%		
Proporción de falsos positivos	1,7%	0,3%	9,1%		
Proporción de falsos negativos	15,8%	7,4%	30,4%		
Exactitud	92,7%	85,7%	96,4%		
Odds ratio diagnóstica	304,00	35,03	2.638,29		
Probabilidad pre-prueba (prevención)	39,6%	—	—		
Medidas simétricas					
	Valor	Error estándar asintótico ^a	T. aproximada ^b	Valor de p	
Chi-cuadrado de Pearson	69,24			< 0,001	
Kappa	0,84	0,06	8,32	< 0,001	
N.º de casos válidos	96				

a: no se presupone la hipótesis nula; b: utilización del error estándar asintótico que presupone la hipótesis nula; p: significación aproximada.

aparecen demostrados en este trabajo, los siguientes indicadores: cuantificar el nivel de acuerdo, la voluntariedad de los participantes, los análisis de tendencia central y dispersión, y los umbrales con los valores para el consenso.

Validez de los estándares

Para el desarrollo de los estándares se consideró imprescindible asegurar la validez de los mismos en sus 4 tipos: de contenido, constructo, apariencia y criterio²⁶. Así, con el e-Delphi, la validez de contenido demostró que los estándares eran relevantes²⁷; y la validez de constructo probó la consistencia interna entre ellos, demostrando la existencia de factores y temáticas comunes en todos los estándares.

El estudio/taller de simulación posibilitó la experimentación con un modelo que imitaba unas condiciones similares a las reales²⁸, estudiándose 3 tipos de validez: contenido, apariencia y criterio. Se ratificó la de «contenido» al explicitar los panelistas la «comprensión» de los estándares ($p < 0,001$) y su posible justificación mediante «pruebas documentales» ($p < 0,001$); y la de «apariencia» al relacionar los estándares con alguna de las dimensiones de la calidad ($p < 0,001$). Por último, para la validez de criterio, al no detectarse otra medida referente publicada, la metodología aplicada en el taller con la creación de un patrón «oro/semi-oro» permitió calcular y obtener una sensibilidad y especificidad altas, y una Chi-cuadrado y un índice kappa significativos, demostrando la capacidad de los estándares para detectar y constatar problemas de funcionamiento en las UCAL²⁹.

Categorías y tipos de estándares

Las categorías que clasifican los estándares contienen, sintéticamente, los conceptos de gestión de la calidad más prevalentes en la actualidad³⁰. Así, en el listado final, las categorías de «estructura» definen los enfoques de planificación y facilitación, sobre liderazgo, planes de calidad, y recursos humanos. En el concepto de «proceso» se ubican los estándares relacionados con la puesta en marcha de las actividades planificadas, con especial atención en las líneas estratégicas sobre la seguridad de los pacientes. Y, en el concepto de «resultado» se detallan, entre otros, la experiencia de los pacientes, el desarrollo de los planes de calidad, las actividades formativas en calidad, y los principales resultados de las UCAL referentes al cumplimiento de objetivos y de acciones de mejora.

Una vez desarrollados los estándares sería clave una estrategia de difusión de los mismos desde la SECA a los centros sanitarios españoles y a sus UCAL. Además, los estándares deberían permitir, a corto y medio plazo, crear un sistema de acreditación de UCAL en España.

En conclusión, la implementación de herramientas cualitativas refrendadas cuantitativamente por distintas pruebas estadísticas, con la participación de profesionales expertos en la metodología de la calidad asistencial en cada uno de sus centros, ha permitido realizar un proyecto inédito de gran magnitud, desarrollando estándares orientados a la potenciar y mejorar el trabajo de las UCAL en centros sanitarios.

Aprobación ética y consentimiento para participar

El estudio fue aprobado por el Comité de Investigación del centro al que pertenecía el investigador principal, y por el Comité de Ética e Investigación de la Universidad Internacional de Cataluña con el código MED-2022-10, versión del protocolo 1.0. Para el estudio e-Delphi se obtuvo el consentimiento informado previo de los profesionales participantes, teniendo en cuenta todos los aspectos establecidos al respecto.

Financiación

En algunas de las actividades descritas en este trabajo, la Sociedad Española de Calidad Asistencial, ha colaborado en la organización de las reuniones anuales españolas de coordinadores de calidad, subvencionando los gastos logísticos de las mismas.

Conflicto de intereses

Los autores declaran que no existe ningún tipo de conflicto de intereses de carácter económico, profesional o personal en el proyecto presentado, que pudiera influir indebidamente y pretendiera prevalecer sobre la finalidad propia de la investigación.

Agradecimientos

A todos los compañeros de profesión, coordinadores de muchas de las UCAL de centros sanitarios españoles, que

participaron con motivación y proactividad en el comité científico y en los paneles de expertos para desarrollar los estándares; y, a los pacientes, usuarios y ciudadanos que son, en definitiva, la razón del trabajo diario estas unidades.

Anexo. Material adicional

Se puede consultar material adicional a este artículo en su versión electrónica disponible en [doi:10.1016/j.jhqr.2023.09.009](https://doi.org/10.1016/j.jhqr.2023.09.009).

Bibliografía

1. Donabedian A. *The Definition of Quality and Approaches to Its Management Vol 1. Explorations in Quality Assessment and Monitoring*. Ann Arbor: Health Administration Press; 1980.
2. Busse R, Panteli D, Quentin W. An introduction to healthcare quality: defining and explaining its role in health systems. En: Busse R, Klazinga N, Panteli D, et al, editores. *Improving healthcare quality in Europe: Characteristics, effectiveness and implementation of different strategies*. Copenhagen (Denmark): European Observatory on Health Systems and Policies; 2019. Health Policy Series, No. 53 [acceded 2 Nov 2021] Available from: <https://www.ncbi.nlm.nih.gov/books/NBK549277/>
3. Mira JJ, Gómez JM. Criterio, Indicador y Estándar. Elche: Universitas. Universidad Miguel Hernández de Elche [consultado 2 Nov 2021] Disponible en: <http://navarrof.orgfree.com/Docencia/Calidad/UT7/criterio.pdf>
4. López-Andújar Aguiriano L, Fernández-Valderrama Benavides J, Tejedor Fernández M, Millán Soria J, Preciado Barahona MJ, Benito Fernández J, et al. Criterios de acreditación de servicios de urgencias de hospitales. 4.^a ed. Madrid: Sociedad Española de Medicina de Urgencias y Emergencias (SEMES); 2019 [consultado 2 Nov 2021] Disponible en: <https://www.semes.org/wp-content/uploads/2019/11/CRITERIOS-ACREDITACION-SEMES-v6-2019.pdf>
5. Tossaint-Schoenmakers R, Versluis A, Chavannes N, Talboom-Kamp E, Kasteleyn M. The Challenge of Integrating eHealth Into Health Care: Systematic Literature Review of the Donabedian Model of Structure Process, and Outcome. *J Med Internet Res.* 2021;23(5):e27180 [acceded 2 Nov 2021] Available from: <https://doi.org/10.2196/27180>
6. Gregori D, Napolitano G, Scarinzi C, Semeraro A, Rosato R, Pagano E, et al. Knowledge, practice and faith on Total Quality Management principles among workers in the Health Care System: Evidence from an Italian investigation. *Journal of Evaluation in Clinical Practice.* 2009;15:69-75, <http://dx.doi.org/10.1111/j.1365-2753.2008.00956.x>.
7. Hernández Sampieri R. Metodología de la investigación. 6.^a edición México: Mc Graw Hill; 2014. p. 200-10 [consultado 2 Nov 2021] Disponible en: https://apiperiodico.jalisco.gob.mx/api/sites/periodicooficial.jalisco.gob.mx/files/metodologia_de_la_investigacion_-roberto.hernandez_sampieri.pdf
8. Sociedad Española de Calidad Asistencial. La cartera de servicios de la unidad de calidad de un hospital. Madrid: 2.^a Reunión de Coordinadores de Calidad de Hospitales; 2009 [consultado 2 Nov 2023] Disponible en: <http://docplayer.es/24724516-2a-reunion-de-coordinadores-de-calidad-de-hospitales-la-cartera-de-servicios-de-la-unidad-de-calidad-de-un-hospital.html>
9. Prieto-Velasco M, del Pino MD, Buades Fuster JM, Craver L, Pons Prades R, Ruiz San Millán JC, et al. Unidades de Enfermedad Renal Crónica Avanzada en España: una encuesta nacional sobre los estándares de estructura, recursos, resul-

- tados y seguridad del paciente. *Nefrología*. 2020;40:608–22, <http://dx.doi.org/10.1016/j.nefro.2020.06.006>.
10. Sociedad Española de Calidad Asistencial [consultado 2 Nov 2023] Disponible en: <http://calidadesistencial.es/>
 11. Sociedad Española de Calidad Asistencial. Líneas estratégicas 2016–2020. Sociedad Española de Calidad Asistencial [consultado 2 Nov 2023] Disponible en: <https://calidadesistencial.es/plan-estrategico-seca-antigua/>
 12. Veiga-Seijo S, Movilla-Fernández MJ, Rivas-Quarneti N. Scoping Review sobre el Metaplan como Metodología de Investigación Cualitativa en Ciencias de la Salud. *NTQR*. 2020;3:1000–7 [accessed 2 Nov 2021] Available from: <https://publi.ludomedia.org/index.php/ntqr/article/view/222>.
 13. Hong QN, Pluye P, Fàbregues S, Bartlett G, Boardman F, Cargo M, et al. Improving the content validity of the mixed methods appraisal tool: A modified e-Delphi study. *J Clin Epidemiol*. 2019;111:49–59 [accessed 2 Nov 2021] Available from: <https://pubmed.ncbi.nlm.nih.gov/30905698/>
 14. Almanasreh E, Moles R, Chen TF. Evaluation of methods used for estimating content validity. *Res Social Adm Pharm*. 2019;15:214–21 [accessed 2 Nov 2021] Available from: <https://doi.org/10.1016/j.sapharm.2018.03.066>
 15. Sürütü L, Maslakçı A. Validity and reliability in quantitative research. *Business & Management Studies: An International Journal*. 2020;8:2694–726 [accessed 2 Nov 2021] Available from: <https://www.bmij.org/index.php/1/article/view/1540>
 16. Trikalinos TA, Balion CM. Chapter 9: options for summarizing medical test performance in the absence of a "gold standard". *J Gen Intern Med*. 2012;27 Suppl. 1:S67–75, <http://dx.doi.org/10.1007/s11606-012-2031-7>.
 17. Power L, Mullally D, Gibney ER, de van der Schueren MAE, Corish CA. A review of the validity of malnutrition screening tools used in older adults in community and healthcare settings - A MaNuEL study. *Clin Nutr ESPEN*. 2018;24:1–13, <http://dx.doi.org/10.1016/j.clnesp.2018.02.005>.
 18. Tabrizi JS, Gharibi F. Primary healthcare accreditation standards: A systematic review. *Int J Health Care Qual Assur*. 2019;32:310–20, <http://dx.doi.org/10.1108/IJHCQA-02-2018-0052>.
 19. Gold R, Oreja-Guevara C. Advances in the management of multiple sclerosis spasticity: Multiple sclerosis spasticity guidelines. *Expert Rev Neurother*. 2013;13 Suppl 12:S5–9, <http://dx.doi.org/10.1586/14737175.2013.865880>.
 20. Escobar-Pérez J, Martínez A. Validez de contenido y juicio de expertos: una aproximación a su utilización. *Avances en Medición*. 2008;6:27–36 [consultado 2 Nov 2021] Disponible en: https://www.researchgate.net/publication/302438451_Validez_de_contenido_y_juicio_de_expertos.Una_aproximacion_a_su_utilizacion
 21. Totaljobs. London: Quality & Safety Health - Spain; 2022 [accessed 2 Apr 2022] Available from: <https://www.totaljobs.com/jobs/quality-and-safety-health/in-spain?radius=10&q=Quality+%26+Safety+Health>
 22. Calvet X, Panés J, Alfaro N, Hinojosa J, Sicilia B, Gallego M, et al. Delphi consensus statement: Quality Indicators for Inflammatory Bowel Disease Comprehensive Care Units. *J Crohns Colitis*. 2014;8:240–51 [accessed 2 Nov 2021] Available from: <https://doi.org/10.1016/j.crohns.2013.10.010>
 23. Meshkat B, Cowman S, Gethin G, Ryan K, Wiley M, Brick A, et al. Using an e-Delphi technique in achieving consensus across disciplines for developing best practice in day surgery in Ireland. *J Hosp Adm*. 2014;3:1–8 [accessed 2 Nov 2021] Available from: <https://core.ac.uk/download/pdf/60772805.pdf>
 24. Boulkedid R, Abdoul H, Loustau M, Sibony O, Alberti C. Using and reporting the Delphi method for selecting healthcare quality indicators: A systematic review. *PLoS One*. 2011;6:e20476, <http://dx.doi.org/10.1371/journal.pone.0020476>.
 25. Diamond IR, Grant RC, Feldman BM, Pencharz PB, Ling SC, Moore AM, et al. Defining consensus: A systematic review recommends methodologic criteria for reporting of Delphi studies. *J Clin Epidemiol*. 2014;67:401–9, <http://dx.doi.org/10.1016/j.jclinepi.2013.12.002>.
 26. Hasson F, Keeney S. Enhancing rigour in the Delphi technique research. *Technol Forecast Soc Change*. 2011;78:1695–704, <http://dx.doi.org/10.1016/j.techfore.2011.04.005>.
 27. Pedrosa I, Suárez-Álvarez J, García-Cueto E. Evidencias sobre la validez de contenido: avances teóricos y métodos para su estimación. *Acción psicol*. 2013;10:3–18, <http://dx.doi.org/10.5944/ap.10.2.11820>.
 28. Cook DA, Brydges R, Zendejas B, Hamstra SJ, Hatala R. Technology-enhanced simulation to assess health professionals: A systematic review of validity evidence, research methods, and reporting quality. *Acad Med*. 2013;88:872–83 [accessed 2 Nov 2021] Available from: https://journals.lww.com/academicmedicine/Fulltext/2013/06000/Technology_Enhanced_Simulation_to_Assess_Health.34.aspx
 29. Boateng GO, Neilands TB, Frongillo EA, Melgar-Quiñonez HR, Young SL. Best Practices for Developing and Validating Scales for Health Social, and Behavioral Research: A Primer. *Front Public Health*. 2018;6, <http://dx.doi.org/10.3389/fpubh.2018.00149>.
 30. Suárez Gutiérrez R, Díaz Vázquez CA. Calidad asistencial. Modelos de gestión de la calidad. *Form Act Pediatr Aten Prim*. 2010;3:25–30 [consultado 2 Nov 2021] Disponible en: <http://gestionclinica.sespa.es/pdf/jornadas2012/t2.lectura1.pdf>