

“VALIDACIÓN DE LA UTILIDAD CLÍNICA DE LA METODOLOGÍA DE SÍNTESIS TERCIARIA THE UMBRELLA COLLABORATION®, UNA HERRAMIENTA INFORMÁTICA SEMIAUTOMÁTICA DE INTELIGENCIA ARTIFICIAL PARA LA SÍNTESIS DE LA EVIDENCIA CIENTÍFICA”

Autor

Beltrán Carrillo Manrique

Directores

Dr. Francisco Javier Gómez Pavón | Dr. Juan José Baztán Cortés

Tutor

Dr. Fernando Canillas del Rey

**“VALIDACIÓN DE LA UTILIDAD CLÍNICA DE LA
METODOLOGÍA DE SÍNTESIS TERCIARIA THE UMBRELLA
COLLABORATION®, UNA HERRAMIENTA INFORMÁTICA
SEMIAUTOMÁTICA DE INTELIGENCIA ARTIFICIAL PARA
LA SÍNTESIS DE LA EVIDENCIA CIENTÍFICA”**

Autor

Beltrán Carrillo Manrique

Directores

Dr. Francisco Javier Gómez Pavón | Dr. Juan José Baztán Cortés

Tutor

Dr. Fernando Canillas del Rey

Villanueva de la Cañada, 2025

_____ **RESUMEN TESIS / ABSTRACT THESIS** _____

TÍTULO: Validación de la utilidad clínica de la metodología de síntesis terciaria The Umbrella Collaboration®, una herramienta informática semiautomática de inteligencia artificial para la síntesis de la evidencia científica.

INTRODUCCIÓN: La síntesis terciaria de evidencia representa un nivel avanzado de integración de hallazgos, diseñado para consolidar información de revisiones sistemáticas (RS) previas y responder a preguntas complejas de investigación. The Umbrella Collaboration® (TU®) propone una metodología semiautomatizada asistida por inteligencia artificial (IA) que busca superar las limitaciones estructurales de las revisiones en paraguas tradicionales (URT: *umbrella review* tradicional), como el tiempo de ejecución prolongado, la actualización insuficiente, la complejidad en la interpretación de resultados y la no inclusión de todas las fuentes disponibles, lo que no solo limita su alcance, sino que también puede restringir los resultados encontrados al omitir potencialmente información relevante de RS publicadas.

OBJETIVO: Evaluar la metodología TU® en comparación con URT en términos de identificación de resultados de interés (RdI), magnitud del efecto (tamaño de efecto: TE), dirección, significación estadística, certeza del efecto y tiempo de ejecución. Además, analizar la percepción de utilidad de TU® por parte de profesionales sanitarios de servicios de geriatría en hospitales universitarios de la Comunidad de Madrid.

MÉTODOS: Se llevaron a cabo dos estudios para abordar los objetivos planteados. El Estudio I examinó la concordancia entre TU® y URT en la identificación de RdI y en sus propiedades clave, como TE, dirección, significación estadística y certeza. Adicionalmente se midió el tiempo de ejecución de los proyectos. El Estudio II evaluó la percepción de utilidad de TU® mediante encuestas realizadas a profesionales sanitarios de geriatría en hospitales universitarios de la Comunidad de Madrid. Los análisis estadísticos permitieron identificar concordancias y diferencias entre ambas metodologías.

RESULTADOS: En el Estudio I, TU® identificó 410 RdI frente a 86 identificados por los URT, con un factor de incremento global de 4.77 (1.87–18.63). TU® coincidió con los URT en 73 RdI (84.88%) e identificó 337 RdI adicionales no identificados por los URT. La concordancia categórica del TE mostró coincidencias consistentes en el 77.2%, 89.2% y 93.8% en los diferentes niveles de análisis O, A y B, respectivamente. Los coeficientes de correlación de Spearman para los TE cuantitativos fueron de 0.637, 0.512 y 0.399 ($p < 0.001$ para O y A; $p = 0.005$ para B). La concordancia en la dirección y significación estadística del efecto fue del 82.19% (datos originales) y del 87.67% (datos revisados), con χ^2 de 13.19 ($p < 0.001$) y 25.28 ($p < 0.001$), V de Cramer de 0.425 (moderado) y 0.588 (grande), e índices Kappa de 0.357 (débil) y 0.514 (moderado). La correlación de certeza entre TU® y GRADE tuvo un coeficiente de Spearman de 0.446 ($p = 0.025$). El tiempo medio de ejecución con TU® fue de 4:46 horas ($\pm 2:30$), con una mediana de 4:36 horas y un rango entre 1:34 a 10:08 horas.

En el Estudio II, la encuesta, cumplimentada por 121 profesionales sanitarios, mostró una puntuación media global de 4.35 ± 0.86 en una escala de 1 a 5. Los médicos (N=104) obtuvieron una media de 4.36 ± 0.88 y los enfermeros (N=17) de 4.26 ± 0.76 , sin diferencias significativas. Las mujeres (N=96) obtuvieron una media de 4.31 ± 0.83 y los hombres

(N=25) de 4.46 ± 0.96 , sin diferencias significativas. Tampoco se encontraron diferencias significativas entre los grupos etarios (<33 años, 33-43 años, 44-54 años, y 55-65 años). En el subanálisis por edad, los mayores (≥ 44 años, N=27) obtuvieron una media de 4.60 ± 0.57 , frente a 4.26 ± 0.91 en los jóvenes (<44 años, N=94; con diferencias significativas en la mayoría de los ítems evaluados).

CONCLUSIÓN: TU® ha demostrado ser una metodología válida para la síntesis terciaria de evidencia, con resultados comparables a las URT en términos de identificación, TE, dirección, significación estadística y certeza de los RdI. Además, TU® reduce significativamente el tiempo de ejecución, completando los proyectos en pocas horas frente a los meses requeridos por las URT. La valoración global de los profesionales sanitarios fue positiva, con una aceptación mayor entre los profesionales mayores. Estos resultados sugieren que TU® es una herramienta eficiente y útil, aunque se requiere su validación en otros contextos para consolidar su aplicabilidad.

PALABRAS CLAVE: Síntesis de evidencia terciaria, The Umbrella Collaboration, revisiones umbrella, metodología de investigación en salud, síntesis asistida por inteligencia artificial, toma de decisiones basada en evidencia.

TITLE: Validation of the Clinical Utility of the Tertiary Synthesis Methodology The Umbrella Collaboration®: A Semi-Automated Artificial Intelligence Tool for Synthesizing Scientific Evidence.

INTRODUCTION: Tertiary evidence synthesis represents an advanced level of integrating findings, designed to consolidate information from prior systematic reviews (SR) and address complex research questions. The Umbrella Collaboration® (TU®) introduces a semi-automated methodology supported by artificial intelligence (AI) designed to overcome the structural limitations of traditional umbrella reviews (URT). These limitations include prolonged execution times, insufficient updates, complexity in result interpretation, and the exclusion of available sources, which not only constrains the scope of analysis but also risks missing relevant findings from published SR.

OBJECTIVE: To assess the performance of the TU® methodology in comparison to URT regarding the identification of results of interest (RdI), effect magnitude (effect size: TE), direction, statistical significance, and certainty, and project execution time. Furthermore, to explore the perceived utility of TU® among healthcare professionals working in geriatric services at university hospitals in the Community of Madrid.

METHODS: Two studies were conducted to address the stated objectives. Study I assessed the concordance between TU® and URT in identifying RdI and their key properties, including TE, direction, statistical significance, and certainty. Additionally, project execution times were measured. Study II evaluated the perceived utility of TU® through surveys administered to geriatric healthcare professionals at university hospitals in the Community of Madrid. Statistical analyses were performed to identify concordances and differences between the two methodologies.

RESULTS: In Study I, TU® identified 410 RdI compared to 86 identified by URT, achieving an overall identification increase factor of 4.77 (range: 1.87–18.63). TU® matched URT in 73 RdI (84.88%) while identifying an additional 337 RdI not captured by URT. Consistent agreement in TE categories was observed in 77.2%, 89.2%, and 93.8% of cases across the O, A, and B levels of analysis, respectively. Spearman correlation coefficients for quantitative TE measures were 0.637, 0.512, and 0.399 ($p < 0.001$ for levels O and A; $p = 0.005$ for level B). Agreement in the direction and statistical significance of the effect ranged from 82.19% (original data) to 87.67% (revised data), with χ^2 values of 13.19 ($p < 0.001$) and 25.28 ($p < 0.001$), corresponding Cramér's V values of 0.425 (moderate) and 0.588 (strong), and Cohen's Kappa coefficients of 0.357 (weak) and 0.514 (moderate). Certainty levels between TU® and GRADE showed a positive correlation, with a Spearman coefficient of 0.446 ($p = 0.025$). Execution times for TU® projects averaged 4:46 hours ($\pm 2:30$), with a median time of 4:36 hours and a range from 1:34 to 10:08 hours.

In Study II, a survey completed by 121 healthcare professionals from geriatric services across university hospitals in Madrid yielded a global mean score of 4.35 ± 0.86 on a 1-to-5 scale. Physicians (N=104) reported a mean score of 4.36 ± 0.88 , while nurses (N=17) reported 4.26 ± 0.76 , with no statistically significant differences. Female respondents (N=96) had a mean score of 4.31 ± 0.83 , and male respondents (N=25) scored 4.46 ± 0.96 ,

also with no significant differences. No significant differences were observed among age groups (<33 years, 33–43 years, 44–54 years, and 55–65 years). However, in a subgroup analysis, older participants (≥ 44 years, N=27) reported a significantly higher mean score of 4.60 ± 0.57 compared to younger participants (<44 years, N=94), who scored 4.26 ± 0.91 , with significant differences observed across most survey items.

CONCLUSIONS: TU® has proven to be a valid methodology for tertiary evidence synthesis, delivering results comparable to URT in terms of identifying RdI, TE, direction, statistical significance, and certainty. Additionally, TU® significantly reduces execution time, completing projects in a few hours compared to the months typically required by URT. Healthcare professionals provided positive feedback on the tool, with higher acceptance observed among older professionals. These findings indicate that TU® is an efficient and practical tool, though further validation in diverse contexts is needed to establish its broader applicability.

KEY WORDS: Tertiary Evidence Synthesis, The Umbrella Collaboration, Umbrella Reviews, Health Research Methodology, AI-assisted Synthesis, Evidence-based Decision Making.

ABREVIATURAS

E-I: Estudio-I

E-II: Estudio-II

ECA: ensayo clínico aleatorizado

EE.UU.: Estados Unidos de América

GRADE: Grading of Recommendations Assessment, Development and Evaluation

HR: hazard ratio

HU: Hospital Universitario

IA: inteligencia artificial

JBI: Joanna Briggs Institute

LLM: Large Language Models (Modelos de Lenguaje a Gran Escala)

LSR: Living Systematic Review (Revisión Sistemática Viva)

MD: diferencia de medias

OMS: Organización Mundial de la Salud

OR: odds ratio

PLS: Plain Language Summaries (Resúmenes en Lenguaje Sencillo)

PMID: PubMed Identifier

Rdl: resultados de interés

RR: riesgo relativo

RRp: revisión Rápida

RS: revisión sistemática

RS/MA: revisión sistemática con o sin metaanálisis

SMD: diferencia de medias estándar

TE: tamaño de efecto

TU®: The Umbrella Collaboration®

URT: umbrella review tradicional

WMD: diferencia de medias ponderada

ÍNDICE

1. INTRODUCCIÓN Y JUSTIFICACIÓN	
1.1 LA SÍNTESIS DE LA EVIDENCIA: UN PILAR PARA LA SALUD PÚBLICA Y LA PRÁCTICA CLÍNICA.....	2
1.2 SINTESIS SECUNDARIA DE LA EVIDENCIA.....	4
1.3 LIMITACIONES ESTRUCTURALES DE LAS RS/MA	
1.3.1 Tiempo excesivo para completar una RS/MA.....	6
1.3.2 Actualización de RS/MA.....	7
1.3.3 Accesibilidad y comprensión de resultados de síntesis de evidencia.....	7
1.3.4 Diferencias de resultados entre RS/MA ante la misma pregunta de investigación.....	9
1.4 SOLUCIONES A LAS LIMITACIONES ESTRUCTURALES DE RS/MA	
1.4.1 Revisiones rápidas.....	9
1.4.2 Revisiones sistemáticas vivas.....	10
1.4.3 Automatización del proceso de síntesis secundaria.....	10
1.4.4 Mejorar el acceso y comprensión de los resultados de síntesis de la evidencia en ciencias de salud.....	12
1.4.5 Síntesis terciaria de la evidencia.....	13
1.4.5.1 Definición de síntesis terciaria de la evidencia.....	13
1.4.5.2 Metodología de la síntesis terciaria.....	15
1.5 INNOVACIÓN EN LA SINTESIS DE LA EVIDENCIA: EL PAPEL DE THE UMBRELLA COLLABORATION®.....	17
2 HIPÓTESIS Y OBJETIVOS	
2.1 HIPÓTESIS.....	23
2.2 OBJETIVOS	
2.2.1 Objetivo principal.....	23
2.2.2 Objetivos específicos.....	23
3 MATERIAL Y MÉTODOS	
3.1 INTRODUCCION.....	26
3.2 DETALLES DE LOS MÉTODOS ESTUDIO I (E-I): ANÁLISIS COMPARATIVO DE METODOLOGIAS	
3.2.1 Diseño del Estudio I: análisis de resultados de metodologías comparativas.....	26
3.2.2 Variables del Estudio I: análisis de resultados de metodologías comparativas.....	26
3.2.2.1 Definición de Rdl y comparación de identificación de Rdl entre metodologías.....	27
3.2.2.2 Comparación y Análisis de Tamaño de Efecto de Rdl.....	27
3.2.2.3 Metodología de la transformación de las métricas a la <i>d</i> de Cohen.....	28
3.2.2.4 Ajustes metodológicos realizados ante la detección de resultados no ajustados en la comparación de TE entre metodologías comparativas.....	28
3.2.2.5 Justificación Niveles de Concordancia TE categorizados.....	29
3.2.3 Definición de la dirección del efecto de Rdl y comparación entre metodologías.....	30
3.2.4 Definición de la significación estadística de Rdl y comparación entre metodologías.....	31
3.2.5 Definición de certeza de la evidencia de los Rdl y comparación de metodologías.....	31
3.2.6 Definición de Tiempo de Ejecución de Proyectos y Comparación de Metodologías.....	32
3.3 DETALLES DE LOS MÉTODOS E-II: COMPRENSIÓN Y UTILIDAD DE THE UMBRELLA COLLABORATION®.....	
3.3.1 Diseño del E-II: comprensión y utilidad de The Umbrella Collaboration®.....	32
3.3.2 Variables del E-II.....	36

3.4 PROCEDIMIENTO EXPERIMENTAL	38
3.4.1 Procedimiento Experimental E-I	
3.4.1.1 Selección y análisis de URT en el área de geriatría.....	38
3.4.1.2 Configuración de Proyectos en Paralelo con la metodología experimental The Umbrella Collaboration®.....	39
3.4.1.3 Proceso de Búsqueda y Selección por The Umbrella®.....	40
3.4.1.4 Verificación de Datos por el Revisor Humano.....	40
3.4.1.5 Normalización y Representación de Datos.....	40
3.4.1.6 Integración y Análisis Continuo (nuevas búsquedas automáticas cada 24h).....	41
3.4.2 Procedimiento Experimental E-II	41
3.5 RECOPIACIÓN DE DATOS	46
3.5.1 Recopilación de datos E-I	46
3.5.2 Recopilación de datos E-II	47
3.6 ANÁLISIS ESTADÍSTICO: ANÁLISIS DE DATOS Y MÉTODOS ESTADÍSTICOS	48
3.6.1 Análisis Estadístico: Análisis de Datos y Métodos Estadísticos E-I	
3.6.1.1 Análisis Cualitativo de Resultados de Interés: identificación de Rdl, tamaño de efecto categorizado, la dirección del efecto de Rdl y la significación estadística del efecto de Rdl.....	48
3.6.1.2 Análisis Cuantitativo de Resultados de Interés: tamaño de efecto y certeza.....	50
3.6.1.3 Procesamiento y análisis de datos.....	51
3.6.2 Análisis Estadístico Estudio II (E-II)	51
3.7 CONSIDERACIONES ÉTICAS	52
4 RESULTADOS	
4.1 RESULTADOS DE LA COMPARACIÓN ENTRE METODOLOGÍAS COMPARATIVAS: TU® FRENTE A URT (ESTUDIO-I)	56
4.1.1 Identificación de Umbrella Reviews Tradicionales en el área de geriatras para utilizar como modelo de referencia	56
4.1.2 Identificación de Resultados de Interés	
4.1.2.1 Resultados globales de Rdl identificados por TU® y URT.....	60
4.1.2.2 Análisis de concordancia en la identificación de Rdl entre TU® y URT.....	61
4.1.2.3 Factor de incremento en la identificación de Rdl por TU®.....	62
4.1.3 Tiempos de Ejecución	
4.1.3.1 Análisis de la relación entre el número de referencias por TU® y el tiempo de ejecución.....	63
4.1.4 Evaluación de la Concordancia y Correlación de los Tamaños de Efecto derivados de Metodologías Comparativas	
4.1.4.1 Evaluación de la concordancia y correlación derivadas de metodologías comparativas. Nivel O.....	66
4.1.4.1.1 Evaluación de la concordancia de los TE derivados de metodologías comparativas. Nivel O.....	66
4.1.4.1.2 Evaluación de la correlación de los TE derivados de metodologías comparativas. Nivel O.....	68
4.1.4.1.3 Conclusión Nivel O.....	70
4.1.4.1.4 Análisis de los 15 valores con discrepancia de más de un nivel detectado en el Nivel O.....	70
4.1.4.2 Evaluación de la concordancia y correlación derivadas de metodologías comparativas.Nivel A.....	71
4.1.4.2.1 Evaluación de la concordancia derivada de metodologías comparativas. Nivel A.....	72
4.1.4.2.2 Evaluación de la correlación derivada de metodologías comparativas. Nivel A.....	74
4.1.4.2.3 Conclusión Nivel A.....	75

4.1.4.3 Evaluación de la concordancia y correlación derivadas de metodologías comparativas Nivel B.....	76
4.1.4.3.1 Evaluación de la concordancia derivada de metodologías comparativas. Nivel B.....	76
4.1.4.3.2 Evaluación de la correlación derivada de metodologías comparativas. Nivel B.....	80
4.1.4.3.3 Conclusión Nivel B.....	81
4.1.4.4 Consecuencias de las modificaciones en los niveles de análisis: concordancia, intersección, pendiente y coeficiente de determinación.....	82
4.1.5 Dirección y Significación Estadística de los Resultados de Interés	
4.1.5.1 Análisis de la dirección y significación estadística de los Rdl con datos “originales”.....	86
4.1.5.2 Análisis estadístico de la dirección de efecto de Rdl de metodologías comparativas.....	87
4.1.5.3 Análisis estadístico de la significación estadística del efecto de Rdl de metodologías comparativas.....	87
4.1.5.4 Análisis de la dirección y significación estadística de los Rdl con datos “revisados”.....	88
4.1.5.5 Análisis estadístico de la dirección de efecto de Rdl de metodologías comparativas.....	89
4.1.5.6 Análisis estadístico de la significación estadística del efecto de Rdl de metodologías comparativas.....	90
4.1.6 Resultados de Concordancia de Niveles de Certeza de Rdl entre metodologías	91
4.2 RESULTADOS ENCUESTA DE SATISFACCIÓN Y UTILIDAD CLÍNICA DE LA METODOLOGÍA Y APLICACIÓN The Umbrella Collaboration® (ESTUDIO-II)	93
4.2.1 Datos descriptivos del total de la muestra	94
4.2.2 Datos de la valoración global de la utilidad clínica de la metodología y del formato de presentación de resultados de TU®	
4.2.2.1 Análisis de las puntuaciones medias, desviaciones estándar por ítem y promedio de la encuesta TU®.....	96
4.2.2.2 Análisis descriptivo de frecuencias de puntuación por ítem de la encuesta TU® según la muestra total de participantes.....	97
4.2.3 Análisis por categoría profesional: médicos y enfermeros	98
4.2.3.1 Análisis de las puntuaciones medias, DE y significación de la encuesta de comprensión y utilidad TU® entre categorías profesionales (médicos y enfermeros).....	98
4.2.3.2 Análisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU® según las dos categorías profesionales participantes (médico / enfermero).....	100
4.2.4 Análisis por Género	
4.2.4.1 Análisis de las puntuaciones medias, DE y significación de la encuesta de utilidad TU® según género (mujer / hombre).....	100
4.2.4.2 Análisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU® según género.....	101
4.2.5 Análisis por grupos de edad	
4.2.5.1 Análisis de las puntuaciones medias, DE y significación de la encuesta de utilidad TU® por grupos de edad.....	101
4.2.5.2 Análisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU® según edad.....	102
4.2.5.3 Subanálisis las puntuaciones medias, DE y significación de la encuesta de utilidad TU® por categorías participantes jóvenes (< 44 años) y participantes mayores (≥ 44 años).....	103
4.2.5.4 Subanálisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU® por categorías participantes jóvenes (< 44 años) y participantes mayores (≥ 44 años).....	105
4.2.6 Análisis por Hospitales	
4.2.6.1 Análisis de las puntuaciones medias, DE y valor p de la encuesta de utilidad TU® por Hospitales.....	105
4.2.6.2 Análisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU® por Hospitales.....	107
4.2.7 Análisis de médicos especialistas en geriatría y resto de especialidades	
4.2.7.1 Diferencias poblacionales entre Geriatras y no Geriatras.....	107

4.2.7.2 Puntuaciones medias, DE y valor p de la encuesta de utilidad TU® entre Geriatras y no Geriatras.....	108
4.2.7.3 Análisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU® entre Geriatras y no Geriatras.....	109
4.2.8 Análisis de médicos especialista en geriatría	
4.2.8.1 Análisis de las puntuaciones medias y DE de la encuesta de utilidad TU® entre Geriatras.....	109
4.2.8.2 Análisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU® en el grupo de Geriatras.....	110
4.2.8.3 Análisis de las puntuaciones medias, DE y diferencia estadística de la encuesta de utilidad TU® entre Geriatras por grupos de edad.....	110
4.2.8.4 Análisis descriptivo de la frecuencia de valoración de valoración por ítem en la encuesta de utilidad de TU® por grupos de edad de los Geriatras.....	112
4.2.8.5 Análisis de las puntuaciones medias, DE y significación estadística de la encuesta de utilidad TU® entre Geriatras jóvenes (< 44 años) y Geriatras mayores (≥ 44 años).....	112
4.2.8.6 Subanálisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU® por categorías geriatras jóvenes (< 44 años) y geriatras mayores (≥ 44 años).....	114
4.2.8.7 Análisis de las puntuaciones medias, DE y significación estadística de la encuesta de utilidad TU® entre Geriatras por hospitales.....	115
4.2.8.8 Análisis descriptivo de la frecuencia de valoración de valoración por ítem en la encuesta de utilidad de TU® entre Geriatras por hospitales.....	116
5 DISCUSIÓN	
5.1 DISCUSIÓN GLOBAL.....	120
5.2 DISCUSIÓN DE ESTUDIO I.....	120
5.2.1 Identificación de URT en el área de la geriatría para la comparación metodológica.....	121
5.2.2 Identificación de Resultados de Interés (Rdl).....	122
5.2.3 Discusión Tiempos de Ejecución Proyectos The Umbrella Collaboration®.....	124
5.2.4 Comparativa de las propiedades de los Rdl entre ambas metodologías	
5.2.4.1 Discusión sobre la Concordancia y Correlación de los Tamaños de Efecto entre las Metodologías Comparativas URT y TU.....	127
5.2.4.2 Dirección y Significación Estadística de los Resultados de Interés (Rdl).....	130
5.2.4.3 Correlación entre las Certezas de los Rdl de las Metodologías.....	131
5.2.5 Contextualización de resultados con estudios previos.....	132
5.3 DISCUSIÓN DEL ESTUDIO II.....	133
6 LIMITACIONES, FORTALEZAS Y LÍNEAS FUTURAS DE INVESTIGACIÓN	
6.1 LIMITACIONES DE LA INVESTIGACIÓN.....	141
6.1.1 Limitaciones del Estudio Comparativo entre TU® y URT.....	141
6.1.2 Limitaciones de la Encuesta sobre la Percepción de TU®.....	143
6.2 FORTALEZAS DEL ESTUDIO.....	144
6.3 LÍNEAS FUTURAS DE INVESTIGACIÓN PARA LA VALIDACIÓN DE TU®.....	145
6.3.1 Ampliación del tamaño de las muestras.....	145
6.3.2 Comparaciones con diseño ciego.....	146
6.3.3 Evaluación con diversos perfiles de revisores.....	146
6.3.4 Ampliación del ámbito de aplicación.....	146
6.3.5 Encuestas de percepción a grupos diversos de las partes interesadas.....	146
6.3.6 Validación de capacidades no exploradas.....	147

7 CONCLUSIONES.....	149
8 ANEXOS	
8.1 ANEXO 1. Metodología de Umbrella Reviews.....	154
8.2 ANEXO 2. Metodologías de la Transformación de las Métricas a la <i>d</i> de Cohen.....	160
8.3 ANEXO 3. Referencias Excluidas y Motivos de Exclusión en la Búsqueda Inicial.....	164
8.4 ANEXO 4. Resultado por proyectos de resultados de interés identificados por las dos metodologías.....	168
8.5 ANEXO 5. Análisis de los Rdl con TE discordantes de mas de un nivel en el Nivel O (N=15)...	196
8.6 ANEXO 6. Análisis de la dirección y Significación Estadística del efecto de Rdl por proyectos de forma individualizada, re-análisis URT con Rdl discordantes.....	206
8.7 ANEXO 7. Análisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU® según las dos categorías profesionales participantes (médico / enfermero).....	224
8.8 ANEXO 8. Análisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU® según género (mujer / hombre).....	226
8.9 ANEXO 9. Análisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU® por grupos de edad.....	228
8.10 ANEXO 10. Subanálisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU® por categorías participantes jóvenes (< 44 años) y participantes mayores (≥ 44 años).....	234
8.11 ANEXO 11. Análisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU® por Hospitales.....	238
8.12 ANEXO 12. Análisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU® entre Geriatras y no Geriatras.....	244
8.13 ANEXO 13. Análisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU® en el grupo de Geriatras.....	246
8.14 ANEXO 14. Análisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU® por grupos de edad de los Geriatras.....	248
8.15 ANEXO 15. Subanálisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU® por categorías geriatras jóvenes (< 44 años) y geriatras mayores (≥ 44 años).....	254
8.16 ANEXO 16. Análisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU® entre Geriatras por hospitales.....	258
8.17 ANEXO 17. Certificado Beca Fundación Alfonso X El Sabio.....	264
8.18 ANEXO 18. Certificado presentación 64 Congreso de las Sociedad Española de Geriatría y Gerontología, Málaga 2024.....	265
8.19 ANEXO 19. Certificado presentación 20 Congreso de la European Geriatric Medicine Society, Valencia 2024.....	266
9 PUBLICACIÓN.....	269
10 BIBLIOGRAFÍA.....	301

ÍNDICE DE TABLAS

Tabla 1	Pasos Metodológicos de URT según distintos autores.....	16
Tabla 2	Códigos, desarrollo y descripción de los ítems de la encuesta online sobre utilidad de The Umbrella Collaboration®.....	37
Tabla 3	URT incluidos búsqueda inicial / URT incluidos para la comparación de resultados.....	57
Tabla 4	Tabla global de Rdl identificados por URT y TU®: totales, coincidencia y no coincidencia.....	60
Tabla 5	Tabla de contingencia. Número de Rdl identificados por URT frente número de Rdl identificados por TU®.....	61
Tabla 6	Rdl identificados por TU® y URT, y Rdl adicionales y factor de identificación de TU®.....	62
Tabla 7	Número de proyecto, fecha, tiempo y números de términos de búsqueda y coincidencia referencias entre metodologías.....	63
Tabla 8	Medidas de Asociación y Significación de Correlación entre Número de Referencias y Tiempo para proyectos TU®.....	64
Tabla 9	Cálculo normalidad de variables relacionados con tiempo de ejecución TU®.....	65
Tabla 10	Tabla de contingencia. Variables TE términos cualitativos: TE URT vs TE TU® Nivel O.....	66
Tabla 11	Distribución de Concordancia y Discrepancia entre Metodologías según Nivel de Acuerdo. Nivel O.....	68
Tabla 12	Medidas de Asociación y Significación de Correlación de TE entre Metodologías. Nivel O.....	68
Tabla 13	Cálculo normalidad variables cuantitativos Rdl Nivel O.....	69
Tabla 14	Evaluación y Decisiones sobre Valores Discordantes (N=15) en el TE entre URT y TU® en el Nivel O.....	71
Tabla 15	Tabla de contingencia. Variables TE términos cualitativos: TE URT vs TE TU®. Nivel A.....	72
Tabla 16	Distribución de Concordancia y Discrepancia entre Metodologías según Nivel de Acuerdo. Nivel A.....	73
Tabla 17	Medidas de Asociación y Significación de Correlación de TE entre Metodologías. Nivel A.....	74
Tabla 18	Cálculo normalidad variables cuantitativos Rdl Nivel A.....	74
Tabla 19	Tabla de contingencia. Variables TE términos cualitativos: TE URT vs TE TU®. Nivel A.....	77
Tabla 20	Distribución de Concordancia y Discrepancia entre Metodologías según Nivel de Acuerdo. Nivel B.....	78
Tabla 21	Medidas de Asociación y Significación de Correlación de TE entre Metodologías. Nivel B.....	80
Tabla 22	Cálculo normalidad varibales cuantitativos Rdl Nivel B.....	80

Tabla 23	Concordancia de dirección y significación estadística del efecto de Rdl entre metodologías (The Umbrella frente a URT original).....	86
Tabla 24	Tabla de contingencia. Variables: dirección efecto Rdl URT frente TU®.....	87
Tabla 25	Tabla de contingencia. Variables: significación estadística del efecto de Rdl URT frente TU®.....	88
Tabla 26	Concordancia de dirección y significación estadística del efecto de Rdl entre metodologías (The Umbrella frente a URT revisado).....	89
Tabla 27	Tabla de contingencia. Variables: dirección efecto Rdl URT (revisados) frente TU®.....	90
Tabla 28	Tabla de contingencia. Variables: significación estadística del efecto de Rdl URT (revisados) frente TU®.....	91
Tabla 29	Listado de valores de certeza transformados de los Rdl de ambas metodologías.....	91
Tabla 30	Análisis de correlación. Variables: valores de certeza Rdl transformados TU® frente URT.....	93
Tabla 31	Cálculo normalidad variables certeza transformados Rdl.....	93
Tabla 32	Características de la muestra.....	94
Tabla 33	Especialidad de los médicos participantes.....	95
Tabla 34	Valoración media y DE de cada ítem de la encuesta sobre la utilidad de TU®.....	96
Tabla 35	Frecuencia global de las valoraciones de cada ítem de la encuesta.....	98
Tabla 36	Valoración media, DE y significación a cada ítem de por categoría profesional.....	99
Tabla 37	Valoración media, DE y significación de cada ítem por género.....	100
Tabla 38	Puntuación media, DE y valor p de cada ítem de la encuesta de utilidad de TU® por grupos de edad.....	101
Tabla 39	Puntuaciones media, DE y valor-p por ítem de la encuesta por grupo de edad jóvenes (< 44 años) y mayores (> 44 años).....	104
Tabla 40	Puntuación media, DE y valor-p de cada ítem por Hospitales.....	105
Tabla 41	Características Demográficas de los Participantes según Especialidad (Geriatras vs. No Geriatras).....	107
Tabla 42	Puntuación media y DS de cada ítem entre Geriatras y no Geriatras.....	108
Tabla 43	Puntuación media y DS de cada ítem entre Geriatras (N=93).....	109
Tabla 44	Puntuación media, DE y valor-p de cada ítem por grupos de edad de geriatras.....	111
Tabla 45	Puntuación media y DE por ítem por grupo de geriatras de edad jóvenes (< 44 años) y mayores (> 44 años).....	113
Tabla 46	Puntuación Media, DE y significación (ANOVA) por geriatras distintos hospitales.....	115

Tabla 47	Proyecto 1. Rdl identificados por URT y TU®: total, coincidentes y no coincidentes.....	168
Tabla 48	Proyecto 1. Rdl identificados por URT y emparejamiento por TU®.....	169
Tabla 49	Proyecto 1. Rdl identificados por TU® no identificados por URT.....	169
Tabla 50	Proyecto 2. Rdl identificados por URT y TU®: total, coincidentes y no coincidentes.....	170
Tabla 51	Proyecto 2. Rdl identificados por URT y emparejamiento por TU®.....	171
Tabla 52	Proyecto 2. Rdl identificados por TU® no identificados por URT.....	172
Tabla 53	Proyecto 3. Rdl identificados por URT y TU®: total, coincidentes y no coincidentes.....	174
Tabla 54	Proyecto 3. Rdl por URT y emparejamiento por TU®.....	175
Tabla 55	Proyecto 3. Rdl identificados por TU® no identificados por URT.....	175
Tabla 56	Proyecto 4. Rdl identificados por URT y TU®: total, coincidentes y no coincidentes.....	177
Tabla 57	Proyecto 4. Rdl identificados por URT y emparejamiento por TU®.....	178
Tabla 58	Proyecto 4. Rdl identificados por TU® no identificados por URT.....	178
Tabla 59	Proyecto 5. Rdl identificados por URT y TU®: total, coincidentes y no coincidentes.....	179
Tabla 60	Proyecto 5. Rdl identificado por URT y emparejamiento por TU®.....	180
Tabla 61	Proyecto 5. Rdl identificado por TU® no identificado por URT.....	180
Tabla 62	Proyecto 6. Rdl identificados por URT y TU®: total, coincidentes y no coincidentes.....	181
Tabla 63	Proyecto 6. Rdl identificado por URT y emparejamiento por TU®.....	183
Tabla 64	Proyecto 6. Rdl identificado por TU® no identificado por URT.....	183
Tabla 65	Proyecto 7. Rdl identificados por URT y TU®: total, coincidentes y no coincidentes.....	188
Tabla 66	Proyecto 7. Rdl identificados por URT y emparejamiento por TU®.....	190
Tabla 67	Proyecto 7. Rdl identificados por TU® no identificados por URT.....	190
Tabla 68	Proyecto 8. Rdl identificados por URT y TU®: total, coincidentes y no coincidentes.....	192
Tabla 69	Proyecto 8. Rdl identificados por URT y emparejamiento por TU®.....	193
Tabla 70	Proyecto 8. Rdl identificados por TU® no identificados por URT.....	194
Tabla 71	P1. (Veronese, 2023/(KAO). Concordancia de dirección y significación estadística del efecto del Rdl entre TU® y URT (N=5).....	206
Tabla 72	P2. (Marx, 2021/(DII). Concordancia de la dirección y significación estadística del efecto de Rdl entre TU® y URT (N=30).....	209
Tabla 73	P3. (Shen, 2022/(EXS). Coincidencia de sentido y significación de efecto entre TU® y URT en Rdl identificados por ambos métodos (N=5).....	213

Tabla 74	P4. (Conneely, 2022/EDI). Coincidencia de sentido y significación de efecto entre TU® y URT en Rdl identificados por ambos métodos (N=7).....	215
Tabla 75	P5. (Gazzaniga, 2023/DGM). Coincidencia de sentido y significación de efecto entre TU® y URT en Rdl identificados por ambos métodos (N=4).....	218
Tabla 76	P6. (Musazadeh, 2023/VDD). Coincidencia de sentido y significación de efecto entre TU® y URT en Rdl identificados por ambos métodos (N=8).....	219
Tabla 77	P7. (Veronese, 2021(1) /LRI). Coincidencia de sentido y significación de efecto entre TU® y URT en Rdl identificados por ambos métodos (N=6).....	220
Tabla 78	P8. (Veronese, 2021(2) /HOA). Coincidencia de sentido y significación de efecto entre TU® y URT en Rdl identificados por ambos métodos (N=8).....	222
Tabla 79	Frecuencia de valoración a cada ítem de la encuesta de profesionales Médicos (N=104).....	224
Tabla 80	Frecuencia de valoración a cada ítem de la encuesta de profesionales Enfermeros (N=17).....	225
Tabla 81	Frecuencia de valores a cada pregunta de la encuesta en mujeres (N=97).....	226
Tabla 82	Frecuencia de valores a cada pregunta de la encuesta en hombres (N=24).....	227
Tabla 83	Frecuencia de valores a cada ítem de la encuesta de utilidad de TU® en el grupo de edad < 33 años (N=68).....	228
Tabla 84	Frecuencia de valores a cada ítem de la encuesta de utilidad de TU® en el grupo de edad 33-43 años (N=26).....	229
Tabla 85	Frecuencia de valores a cada ítem de la encuesta de utilidad de TU® en el grupo de edad 43 a 54 años (N=11).....	230
Tabla 86	Frecuencia de valores a cada pregunta de la encuesta de utilidad de TU® en el grupo de edad 54 a 65 años (N=16).....	231
Tabla 87	Frecuencia de valores a cada pregunta de la encuesta en el grupo de edad MAYORES (N=27).....	235
Tabla 88	Frecuencia de valores a cada pregunta de la encuesta en el grupo de edad MENORES (N=94).....	236
Tabla 89	Frecuencia de valores a cada ítem de la encuesta en HU Puerta de Hierro-Majadahonda (N=13).....	238
Tabla 90	Frecuencia de valores a cada pregunta de la encuesta en HU Clínico San Carlos (N=25).....	239
Tabla 91	Frecuencia de valores a cada ítem de la encuesta en HU La Paz (N=21).....	240

Tabla 92	Frecuencia de valores a cada ítem de la encuesta en HU Ramón y Cajal (N=22).....	241
Tabla 93	Frecuencia de valores a cada ítem de la encuesta en HU Central de la Cruz Roja (N=40).....	242
Tabla 94	Análisis de las diferencias en las frecuencias de valoración de cada ítem entre Geriatras (N=93) y no Geriatras (N=11).....	244
Tabla 95	Frecuencia de valores a cada pregunta de la encuesta en los Geriatras (N=93).....	246
Tabla 96	Frecuencia de valoración de cada ítem en grupo de edad menor de 33 años (N=57).....	248
Tabla 97	Frecuencia de valoración de cada ítem en el grupo de edad 33 a 44 años (N=22).....	249
Tabla 98	Frecuencia de valoración de cada ítem en el grupo geriatras de edad entre 45 a 54 años (N=7).....	250
Tabla 99	Frecuencias valoración de cada ítem en grupo de geriatras de edad entre 55 a 65 años (N=7).....	251
Tabla 100	Frecuencia de valores a cada pregunta de la encuesta en el grupo de edad Geriatras MAYORES (N=14).....	255
Tabla 101	Frecuencia de valores a cada pregunta de la encuesta en el grupo de edad Geriatras MENORES (N=79).....	256
Tabla 102	Frecuencias de valoración de cada ítem en geriatras del HU Puerta de Hierro Majadahonda (N=7; 7.5%).....	258
Tabla 103	Frecuencias de valoración de cada ítem en geriatras del HU Clínico San Carlos (N=20; 21.5%).....	259
Tabla 104	Frecuencias de valoración de cada ítem en geriatras del HU La Paz (N=21; 22.6%).....	260
Tabla 105	Frecuencias de valoración de cada ítem en geriatras del HU Ramón y Cajal (N=16; 17.2%).....	261
Tabla 106	Frecuencias de valoración de cada ítem en geriatras del HU Cruz Roja (N=29; 31.2%).....	262

ÍNDICE DE FIGURAS

Figura 1	Enlace QR de acceso a la encuesta.....	41
Figura 2	Página inicio The Umbrella Collaboration®.....	42
Figura 3	Listado de especialidades página web The Umbrella Collaboration®.....	42
Figura 4	Listado de proyectos dentro del área Test The Umbrella Collaboration®.....	43
Figura 5	Bubble Plot principal Knee osteoarthritis and health outcomes.....	44
Figura 6	Despliegue ventana información de resultado de interés KOA. Subclinical atherosclerosis risk en Knee osteoarthritis and health outcomes.....	45
Figura 7	Forest Plot de resultado de interés KOA. Subclinical atherosclerosis risk en Knee osteoarthritis and health outcomes.....	45
Figura 8	Bubble plot secundario y listado de RS/MA de resultado de interés KOA. Subclinical atherosclerosis risk en Knee osteoarthritis and health outcomes.....	46
Figura 9	Flujograma búsqueda y selección URT geriatría.....	67
Figura 10	RdI identificados por URT concordancia TU®.....	68
Figura 11	RdI identificados por TU® concordancia URT.....	61
Figura 12	Gráfico dispersión de la relación entre el Número de Referencias incluidas en los proyecto TU® y el Tiempo de ejecución (h) de los proyectos con regresión lineal.....	64
Figura 13	Categorías de TE cualitativo según metodologías. Nivel O (N=66).....	67
Figura 14	Distribución de Concordancia TE entre Metodologías Nivel O (N=66).....	68
Figura 15	Gráfico dispersión de la relación entre el TE de URT (<i>d</i> de Cohen) y The Umbrella (<i>r</i>) con regresión lineal. Nivel O.....	69
Figura 16	Categorías de TE cualitativo según metodologías. Nivel A (N=65).....	73
Figura 17	Distribución de Concordancia TE entre Metodologías. Nivel A (N=65).....	74
Figura 18	Gráfico dispersión de la relación entre el TE de URT (<i>d</i> de Cohen) y The Umbrella (<i>r</i>) con regresión lineal. Nivel A.....	75
Figura 19	Categorías de TE cualitativo según metodologías. Nivel B (N=48).....	77
Figura 20	Distribución de Concordancia TE entre Metodologías. Nivel B (N=48).....	78
Figura 21	Gráfico dispersión de la relación entre el TE de URT (<i>d</i> de Cohen) y The Umbrella (<i>r</i>) con regresión lineal. Nivel B.....	80
Figura 22	Grados de concordancia de TE (%) según nivel de análisis (O, A, B).....	82
Figura 23	Valor intersección en los distintos niveles de análisis (O, A y B).....	83
Figura 24	Valor de pendiente de recta de ajuste en los distintos niveles de análisis (Nivel O, A y B).....	84
Figura 25	Valores de coeficiente determinación en los distintos niveles de análisis (O, A y B).....	85
Figura 26	Gráfica dispersión valores transformados certeza RdI (N=25).....	92
Figura 27	Distribución grupos de edades (N=121).....	95

Figura 28	Participantes por hospitales (N; %)	96
Figura 29	Proyecto 1. Rdl identificados por URT (N=5) concordancia TU®	168
Figura 30	Proyecto 1. Rdl identificados por TU® (N=12) concordancia URT	169
Figura 31	Proyecto 2. Rdl identificados por URT (N=38) concordancia TU®	170
Figura 32	Proyecto 2. Rdl identificados por TU® (N=71) concordancia URT	171
Figura 33	Proyecto 3. Rdl identificados por URT (N=7) concordancia TU®	174
Figura 34	Proyecto 3. Rdl identificados por TU® (N=34) concordancia URT	175
Figura 35	Proyecto 4. Rdl identificados por URT (N=10) concordancia TU®	177
Figura 36	Proyecto 4. Rdl identificados por TU® (N=22) concordancia URT	177
Figura 37	Proyecto 5. Rdl identificados por URT (N=4) concordancia TU®	179
Figura 38	Proyecto 5. Rdl identificados por TU® (N=25) concordancia URT	180
Figura 39	Proyecto 6. Rdl identificados por URT (N=8) concordancia TU®	182
Figura 40	Proyecto 6. Rdl identificados por TU® (N=149) concordancia URT	182
Figura 41	Proyecto 7. Rdl identificados por URT (N=6) concordancia TU®	189
Figura 42	Proyecto 7. Rdl identificados por TU® (N=49) concordancia URT	189
Figura 43	Proyecto 8. Rdl identificados por URT (N=8) concordancia TU®	192
Figura 44	Proyecto 8. Rdl identificados por TU® (N=48) concordancia URT	193
Figura 45	Frecuencia de los valores desglosados de cada ítem de la encuesta de profesionales Médicos (N=104)	225
Figura 46	Frecuencia de los valores desglosados de cada ítem de la encuesta de profesionales Enfermeros (N=17)	225
Figura 47	Frecuencia de los valores desglosados de cada ítem de la encuesta en mujeres (N=96)	227
Figura 48	Frecuencia de los valores desglosados de cada ítem de la encuesta en hombres (N=24)	227
Figura 49	Frecuencia de los valores a cada ítem de la encuesta de utilidad de TU® en el grupo de edad < 33 años (N=68)	229
Figura 50	Frecuencia de los valores a cada ítem de la encuesta de utilidad de TU® en el grupo de edad 33 a 43 años (N=26)	230
Figura 51	Frecuencia de los valores a cada ítem de la encuesta de utilidad de TU® en el grupo de edad 43 a 54 años (N=11)	231
Figura 52	Frecuencia de los valores a cada ítem de la encuesta de utilidad de TU® en el grupo de edad 55 a 65 años (N=11)	232
Figura 53	Frecuencia de valores a cada pregunta de la encuesta en el grupo de edad MAYORES (N=27)	236

Figura 54	Frecuencia de valores a cada pregunta de la encuesta en el grupo de edad MENORES (N=94)	237
Figura 55	Frecuencia de valores en cada ítem en el HU Puerta de Hierro-Majadahonda	239
Figura 56	Frecuencia de valores en cada ítem en el HU Clínico San Carlos (N=25)	240
Figura 57	Frecuencia de valores en cada ítem en el HU La Paz (N=21)	241
Figura 58	Frecuencia de valores en cada ítem en el HU Ramón y Cajal (N=22)	242
Figura 59	Frecuencia de valores en cada ítem en el HU Cruz Roja (N=40)	243
Figura 60	Frecuencia de valores a cada ítem de la encuesta en los Geriatras (N=93)	246
Figura 61	Frecuencias de valoración por ítem en geriatras < 33 años (N=57)	249
Figura 62	Frecuencias de la valoración por ítem geriatras entre 33 - 44 años (N=22)	250
Figura 63	Frecuencias de valoración por ítem en geriatras entre 45 - 54 años (N=11)	251
Figura 64	Frecuencia de valoración por ítem en geriatras entre 55 - 65 años (N=7)	252
Figura 65	Valoraciones medias y DE por ítem geriatras jóvenes (< 44 años) y mayores (> 44 años)	255
Figura 66	Frecuencia de valores a cada pregunta de la encuesta en el grupo de edad Geriatras MAYORES (N=14)	256
Figura 67	Frecuencia de valores a cada pregunta de la encuesta en el grupo de edad Geriatras MENORES (N=79)	257
Figura 68	Frecuencias de valoración por ítem en geriatras del HU Puerta de Hierro Majadahonda (N=7; 7,5%)	259
Figura 69	Frecuencias de valoración por ítem en geriatras del HU Clínico San Carlos (N=20; 21,5%)	260
Figura 70	Frecuencias de valoración por ítem en geriatras del HU La Paz (N=21; 22,6%)	261
Figura 71	Frecuencias de valoración por ítem en geriatras del HU Ramón y Cajal (N=16; 17,2%)	262
Figura 72	Frecuencias de valoración por ítem en geriatras del HU Cruz Roja (N=29; 31,2%)	263

ÍNDICE DE ANEXOS

ANEXO 1	Metodología de Umbrella Reviews	6
ANEXO 2	METODOLOGÍA DE LA TRANSFORMACION DE LAS MÉTRICAS A LA <i>d</i> DE COHEN.....	7
ANEXO 3	Referencias Excluidas y Motivos de Exclusión en la Búsqueda Inicial.....	8
ANEXO 4	RESULTADOS POR PROYECTOS DE RESULTADOS DE INTERÉS IDENTIFICADOS POR LAS DOS METODOLOGÍAS.....	8
ANEXO 5	ANÁLISIS DE R _{di} CON TE DISCORDANTES DE MAS DE UN NIVEL EN EL NIVEL O (N=15).....	11
ANEXO 6	ANÁLISIS DE DIRECCIÓN Y SIGNIFICACIÓN ESTADISTICA DEL EFECTO DE R _{di} POR PROYECTOS DE FORMA INDIVIDUALIZADA, RE-ANÁLISIS URT CON R _{di} DISCORDANTES.....	13
ANEXO 7	Análisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU [®] según las dos categorías profesionales participantes (médico / enfermero).....	15
ANEXO 8	Análisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU [®] según género (mujer / hombre).....	16
ANEXO 9	Análisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU [®] por grupos de edad.....	17
ANEXO 10	Subanálisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU [®] por categorías participantes jóvenes (< 44 años) y participantes mayores (≥ 44 años).....	17
ANEXO 11	Análisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU [®] por Hospitales.....	17
ANEXO 12	Análisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU [®] entre Geriatras y no Geriatras.....	18
ANEXO 13	Análisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU [®] en el grupo de Geriatras.....	18
ANEXO 14	Análisis descriptivo de la frecuencia de valoración de valoración por ítem en la encuesta de utilidad de TU [®] por grupos de edad de los Geriatras.....	21
ANEXO 15	Análisis de las puntuaciones medias, DE y significación estadística de la encuesta de utilidad TU [®] entre Geriatras jóvenes (< 44 años) y Geriatras mayores ≥ 44 años).....	22
ANEXO 16	Análisis descriptivo de la frecuencia de valoración de valoración por ítem en la encuesta de utilidad de TU [®] entre Geriatras por hospitales.....	22
ANEXO 17	CERTIFICADO BECA FUNDACION ALFONSO X EL SABIO.....	22
ANEXO 18	CERTIFICADO PRESENTACIÓN 64 CONGRESO DE LA SOCIEDAD ESPAÑOLA DE GERIATRÍA Y GERONTOLOGÍA, MALAGA 2024.....	22
ANEXO 19	CERTIFICADO PRESENTACIÓN 20 CONGRESO DE LA EUROPEAN GERIATRIC MEDICINE SOCIETY, VALENCIA 2024.....	22

1

INTRODUCCIÓN Y JUSTIFICACIÓN

1.1 LA SÍNTESIS DE LA EVIDENCIA: UN PILAR PARA LA SALUD PÚBLICA Y LA PRÁCTICA CLÍNICA

En el ámbito de la salud, el conocimiento ha emergido como una piedra angular para la toma de decisiones informadas para todas las partes interesadas. Los editores de PLoS Medicine (1), enfatizan el derecho inalienable al acceso a información de salud relevante, fiable y actualizada. Este principio es fundamental en cada avance médico, intervención y decisión clínica, destacando la importancia crítica del conocimiento en el ámbito de la salud. A pesar de esto, en muchas circunstancias el acceso a información de salud sigue siendo un desafío significativo. El Consejo Europeo (2) reconoce la importancia del acceso a información de salud creíble, y señala que comunicar esta información de una manera clara desde fuentes fiables facilita el acceso a información válida. En este contexto, el acceso a información de salud se considera un determinante clave del derecho humano a alcanzar el más alto estándar de salud, implicando responsabilidades legales para los gobiernos bajo el derecho internacional de derechos humanos para proveer acceso a la información sanitaria (3).

La Organización Mundial de la Salud (OMS) a través de su programa Red de Políticas Informadas por Evidencia (4), enfatiza el uso sistemático de los resultados de la investigación y de la evidencia para el desarrollo de políticas de salud. La OMS también resalta la importancia de la traducción del conocimiento en salud (*Knowledge Translation*) como un puente entre la investigación y la práctica clínica, que busca facilitar la integración, el intercambio y la aplicación eficaz del conocimiento procedente de la investigación en salud entre todas las partes interesadas con el objetivo de facilitar la toma de decisiones salud (5) (6). Por otro lado, el acceso a la evidencia sanitaria al más alto nivel, y su comprensión, es fundamental para cerrar la brecha entre los resultados de la investigación y la práctica si se quiere tener un impacto favorable en la salud (7). Todo esfuerzo por agregar los resultados de la investigación en el nivel más alto de evidencia posible y en un formato comprensible, actualizado y accesible contribuye a cerrar la distancia entre el conocimiento existente y su aplicación práctica en la toma de decisiones en salud pública e individual por parte de todas las partes interesadas (8).

Las partes interesadas en salud juegan todas un papel fundamental en la toma de decisiones informadas sobre políticas, programas o prácticas relacionadas con la salud. Cottrell et al. (9) y el Canadian Institutes of Health Research (10) identifican a los pacientes (incluidos los cuidadores, asociaciones de pacientes), proveedores de atención médica, y formuladores y gestores de políticas sanitarias como partes interesadas clave del conocimiento en salud y en la toma de decisiones en el ámbito sanitario. Estos grupos aportan perspectivas únicas que enriquecen la comprensión y aplicación de la evidencia científica en la práctica, subrayando la importancia de la colaboración multidisciplinaria en la investigación y la implementación en la práctica de conocimientos que derivan en mejoras concretas en la salud y el bienestar a nivel comunitario e individual. La importancia de los usuarios del conocimiento radica en su capacidad de acceder, comprender y evaluar adecuadamente el nivel más alto de

evidencia de información en salud, para poder aplicar el conocimiento adquirido en la toma de decisiones en salud. Su rol es esencial para asegurar la traducción del conocimiento y las evidencias científicas a la práctica, motivo por el cual la información en salud debe ser del mayor nivel posible, pero a la vez ser accesible y comprensible para todos los implicados, permitiendo así una participación fundamentada en las decisiones de salud.

La definición de “evidencia” no es sencilla, y es objeto de un considerable debate, como se refleja en una revisión de alcance realizada por Yu et al. (11), que identifica hasta 54 definiciones diferentes de “evidencia” en las ciencias sociales y de la salud. Es sabido que el conocimiento derivado de la investigación en salud no tiene todo el mismo nivel de evidencia, es por ello importante conocer la existencia de distintos niveles de evidencia, conocido como la jerarquía de la evidencia, para la toma de decisiones fundamentadas en la evidencia en salud. La jerarquía de la evidencia es una parte importante de la medicina basada en la evidencia donde la calidad y relevancia de la información científica se distingue y muestra gráficamente en la pirámide de evidencia (12). Esta jerarquía clasifica los estudios desde los estudios primarios (estudios transversales, casos-controles, estudios de cohorte y ensayos clínicos aleatorizados (ECA)) en la base hasta los estudios de síntesis de la evidencia en la cúspide (revisiones sistemáticas con o sin metaanálisis (RS/MA) y las guías de práctica clínica), representando el nivel más alto de evidencia (13). La pirámide de la evidencia, aunque tradicionalmente ha sido una guía fundamental en la investigación clínica, está en constante evolución para reflejar los avances en la generación de conocimiento, especialmente con la incorporación de datos del mundo real y su conversión en evidencia del mundo real, lo que podría reconfigurar su estructura para integrar de manera más efectiva las pruebas obtenidas en entornos clínicos y cotidianos (14).

La síntesis de evidencia constituye un proceso esencial en el ámbito de la investigación, en la formulación de políticas y en la toma de decisiones en salud para todas las partes, incluidos los pacientes, al integrar información y conocimientos de múltiples fuentes y disciplinas en unos resultados únicos. Este proceso en salud tiene como objetivo proporcionar información sobre temas específicos y complejos que rara vez pueden ser respondidos por estudios individuales. La elaboración de un trabajo de síntesis de evidencia en salud que incluya todo lo publicado en el campo de la pregunta de investigación, concisa, transparente y accesible facilita el acceso a la mejor evidencia disponible y a la toma de decisiones informadas por la evidencia en salud pública e individual. Hay múltiples métodos de síntesis que convierten grandes cantidades de datos en información útil y fácilmente comprensible, enriqueciendo a quienes buscan conocimiento, fomentando debates y facilitando la toma de decisiones informadas en salud. Los principios clave para lograr una síntesis efectiva incluyen la inclusión de toda la información, la transparencia, la precisión y la facilidad de comprensión de los resultados, exigiendo la colaboración entre un amplio espectro de participantes, que abarca desde quienes crean políticas hasta especialistas independientes y consumidores finales del conocimiento (15).

Un primer nivel de síntesis de la evidencia es la reconocida como síntesis secundaria de la evidencia. Donde esencialmente se analizan los resultados que ya se han encontrado en estudios primarios anteriores y se agregan para obtener una imagen más completa. Es un método bien reconocido y ampliamente aceptado que ayuda a entender mejor un área de conocimiento, facilitando decisiones más informadas. En esencia, es como armar un gran rompecabezas con piezas que vienen de diferentes cajas, pero que juntas nos dan una imagen completa. Sin embargo, más allá de la síntesis secundaria, existe un nivel menos estudiado, pero igualmente importante para la toma de decisiones en salud: la síntesis terciaria. Este nivel ha recibido diversos nombres, que incluye revisiones de revisiones, *overviews*, *umbrella reviews*, *meta-reviews*, meta-metaanálisis y estudios meta-epidemiológicos. En esta disertación hemos optado por utilizar el término “Umbrella Review” añadiendo el término de “tradicional”, por los que lo abreviamos con el acrónimo “URT”, para especificar el método utilizado hasta ahora para realizar síntesis terciaria. La diferencia entre los análisis secundarios y terciarios se sitúa en el nivel y el tipo de estudios que cada uno utiliza como unidad de estudio. Mientras que los estudios de síntesis secundarios se centran en combinar los hallazgos de estudios primarios, como los ECAs, los terciarios avanzan un paso más allá al sintetizar y agregar los resultados de estudios secundarios, como las RS/MA. La síntesis terciaria no siempre está presente en representaciones tradicionales de la pirámide de la evidencia, pero recientes adaptaciones sugieren su inclusión en la cúspide (16,17), reconociendo su valor al proporcionar una visión aún más consolidada, inclusiva y transversal de la evidencia disponible.

En este contexto, la síntesis de la evidencia de los hallazgos de estudios ya realizados, tanto en su formato secundario como terciario, emergen como herramientas fundamentales al clarificar lo que se conoce de lo que no se conoce sobre una pregunta de investigación (18). El paso siguiente en nuestra discusión profundizará en la naturaleza de estos estudios de síntesis y cómo, al sintetizar de forma concisa datos agregados de estudios ya realizados pueden ser utilizados por diferentes partes interesadas con un nivel adecuado de alfabetización en salud, para transformarse en acciones clínicas concretas y políticas de salud fundamentadas.

1.2. SINTESIS SECUNDARIA DE LA EVIDENCIA

Entrando en detalle dentro de las diferentes técnicas de síntesis secundaria de evidencia, nos enfocaremos en explorar la técnica más valorada en la jerarquía de la evidencia, que es la RS/MA, destacando su importancia y aplicación en el ámbito de la investigación y la toma de decisiones en salud, pero a la vez reconociendo sus limitaciones. Una RS/MA, es una metodología de investigación diseñada para recopilar, evaluar críticamente y sintetizar toda la evidencia disponible de los resultados de estudios primarios ya publicados, y que cumplan con criterios de elegibilidad preespecificados, para responder a una pregunta de investigación específica. Este proceso se caracteriza por el uso de métodos explícitos y sistemáticos, seleccionados con el objetivo de minimizar sesgos, para influir mejor en la

toma de decisiones tanto en la práctica médica como en el desarrollo de políticas de salud. Las RS/MA se distinguen por su rigurosidad metodológica, que incluye la definición de una pregunta de investigación clara, la implementación de una estrategia de búsqueda reproducible en varias bases de datos y motores de búsqueda, la especificación de criterios de inclusión y exclusión, la selección y evaluación crítica de la calidad y el riesgo de sesgo de los estudios incluidos así como la certeza en los resultados, y concluyen con la presentación de los resultados de síntesis en formato narrativo o de agregación cuantitativa con metaanálisis (19).

Las RS/MA han ganado un impulso significativo en las últimas décadas, convirtiéndose en una parte importante de la cadena de información científica y herramientas clave para la toma de decisiones en salud y la práctica médica (20,21). El amplio alcance de las RS/MA permite incluir una gama más amplia de condiciones que un solo estudio primario de forma aislada, fortaleciendo o debilitando la generalización de los resultados. A través de un enfoque riguroso y reproducible, las RS/MA cierran la brecha entre la generación acelerada de nuevos resultados de la investigación primaria y la necesidad de resumir esta información de una manera consistente y accesible, y tienen como objetivo proporcionar una base sólida para la toma de decisiones informadas en el ámbito de la salud (22–26).

La implementación de RS/MA tiene limitaciones estructurales inherentes a su metodología rigurosa, que pueden condicionar su aplicabilidad y utilidad en situaciones específicas. Estas preocupaciones son reconocidas tanto por los expertos en metodología como por las partes interesadas que confían en estas revisiones para tomar decisiones clínicas y de salud pública. Nos centramos en cuatro aspectos críticos. En primer lugar, generalmente se estima que el tiempo necesario para completar una RS/MA es de 1 a 2 años, lo que puede considerarse demasiado largo para los usuarios de conocimiento que requieren el mayor nivel de evidencia para la toma de decisiones en salud. Una segunda limitación surge del rápido desarrollo del conocimiento científico. Esto plantea desafíos importantes para mantener la actualización de las RS/MA, aumentando el riesgo de que la evidencia sintetizada quede obsoleta poco después e incluso antes de su publicación. La tercera limitación destacada es la complejidad de los resultados presentados por las RS/MA, lo que afecta su interpretación por las partes interesadas que no tengan el nivel adecuado de alfabetización. Finalmente, existe la limitación de que la RS/MA no incluyan toda la información relevante, limitando así la fiabilidad, relevancia y la certeza en los resultados, pudiéndose llegar al caso de que diferentes RS/MA ante la misma pregunta de investigación presenten hallazgos encontrados entre ellas.

Ante las limitaciones estructurales anteriormente identificadas en la metodología de las RS/MA, todas las partes interesadas están activamente buscando estrategias que permitan superar estos obstáculos. Este esfuerzo colectivo tiene como objetivo encontrar un equilibrio entre la rigurosidad metodológica y la necesidad de proporcionar evidencia accesible, actualizada, relevante y certera en tiempos más acordes con las demandas de los tomadores de decisiones en el ámbito de la salud. En este contexto, Cochrane

Collaboration (27), Joanna Briggs Institute (28)(JBI) y Campbell Collaboration (29) emergen como actores clave, liderando la exploración de nuevas metodologías para facilitar la realización y actualización de revisiones sistemáticas de manera más eficiente. Este enfoque colaborativo y multidisciplinario refleja un compromiso compartido con la mejora continua de la calidad y la utilidad de la síntesis de evidencia en la toma de decisiones clínicas y de políticas de salud pública.

1.3. LIMITACIONES ESTRUCTURALES DE LAS RS/MA

1.3.1. Tiempo excesivo para completar una RS/MA

El proceso de elaboración de una RS/MA es inherentemente riguroso y laborioso, lo que resulta en tiempos de desarrollo necesariamente prolongados. En promedio, completar una RS/MA puede llevar entre 1 a 2 años, con un tiempo medio de RS/MA registradas en PROSPERO de 67.3 semanas (30). Este extenso período limita la capacidad de las RS/MA para satisfacer demandas urgentes de información, un desafío particularmente evidente durante eventos críticos como la pandemia de SARS-CoV-19 (31) o la crisis del virus del Ébola en 2013 (32). Los resultados de estudios al respecto indican que no ha habido una reducción significativa en ese tiempo de elaboración de las RS/MA a lo largo de los años, como lo demuestra la comparación de los datos entre Borah et al. (30) y Sampson et al. (33) que también describió el tiempo medio de una RS/MA entre final de última búsqueda y publicación de 61 semanas.

Existen excepciones llamativas a estos tiempos excesivos de realización de una RS/MA, si se emplea un equipo pequeño y experimentado de expertos en metodología, utilizando herramientas de automatización, que logran completar RS/MAs de tamaño moderado en tan solo dos semanas (34). Lo que ha llevado al desarrollo de una metodología específica conocida como el método de revisión sistemática de 2 semanas (*2weekSR method*), que ha demostrado ser capaz de realizar una RS/MA completa en un promedio de 11 días laborables. Que en función del tamaño y complejidad de la revisión, ofrece un ahorro de tiempo considerable sin recurrir a las estrategias metodológicas asociados con las revisiones rápidas (35,36).

Otro factor que contribuye a la prolongación en la elaboración de las RS/MA es la necesidad de contar con equipos de trabajo relativamente grandes. Si bien la experiencia combinada de todos los participantes enriquece el proceso, también incrementa la complejidad de la coordinación y por ello el tiempo de elaboración. Es importante destacar que la influencia del equipo en la conducción de una revisión sistemática es un aspecto crítico tanto en positivo como en negativo. La composición del equipo, sobre todo por la existencia de conflictos de interés, puede comprometer la integridad de la revisión sistemática y sus conclusiones. Esto subraya la importancia de prestar atención a quién está realizando estos proyectos de síntesis secundaria, más allá de confiar únicamente en los métodos y conclusiones de las RS/MA para validar su calidad (22,30,37).

Ante el desafío del prolongado tiempo de elaboración de las RS/MA, dos soluciones innovadoras han cobrado especial relevancia: el desarrollo de Revisiones Rápidas (RRp) y la incorporación de sistemas de automatización en la metodología de las RS/MA. Estas estrategias representan esfuerzos significativos para acelerar la producción de síntesis secundaria de evidencia intentando no comprometer la calidad y rigurosidad inherentes a estas investigaciones.

1.3.2. Actualización de RS/MA

La actualización de RS/MA constituye un pilar fundamental para asegurar la relevancia y certeza de los resultados de la síntesis de la evidencia científica publicada. A pesar de los criterios internacionalmente aceptados para una adecuada actualización, que incluyen la revisión de la literatura cada dos años o cuando surjan datos nuevos relevantes en el campo de la pregunta de investigación de la RS/MA, y la verificación de nuevos estudios primarios que cumplen los criterios de inclusión en los seis meses previos a la publicación, la práctica habitual no siempre cumple con estas expectativas. El no adherirse a los estándares establecidos puede resultar en la difusión de evidencia obsoleta o la diseminación de datos incorrectos (38,39). Es crucial actualizar las RS/MA de manera oportuna, ya que se ha identificado que una proporción significativa de estas RS/MA no actualizadas podrían omitir información relevante que alteraría las conclusiones dos años después de su publicación. La situación se agrava ya que en no pocas ocasiones las RS/MA ni siquiera están actualizadas en el momento de su publicación (40).

La discrepancia entre las expectativas y la realidad se evidencia al examinar la frecuencia de actualización de las revisiones sistemáticas Cochrane y de las revisiones sistemáticas no Cochrane. Las revisiones Cochrane se actualizan aproximadamente a los dos años en el 50% de los casos, mientras que las revisiones no Cochrane llegan con dificultad a la cifra del 5%, lo que contribuye a la presencia, y a disposición de los usuarios de conocimiento, de resultados de síntesis de información incompleta y potencialmente imprecisa (41). Esta limitación en el "periodo de supervivencia" de las RS/MA, especialmente en áreas con altas tasas de publicación, subraya la necesidad de adoptar métodos de actualización más rigurosos y sistemáticos en las RS/MA, esencial para mantener la precisión y confiabilidad de la evidencia científica, fundamentales en la toma de decisiones de salud (42). Para mitigar esta limitación estructural se han desarrollado las Revisiones Sistemáticas Vivas (Living Systematic Review, LSR).

1.3.3. Accesibilidad y comprensión de resultados de síntesis de evidencia

Navegar eficazmente por el continuo de salud demanda el dominio de cuatro competencias fundamentales, cada una reflejando una dimensión esencial de la alfabetización en salud y estadística. La primera competencia, acceso, implica la habilidad para buscar, encontrar y obtener información de salud pertinente. La segunda, comprensión, se centra en la capacidad para entender la información de salud una vez que ha sido accedida. La tercera, evaluación, hace referencia a la habilidad para interpretar, filtrar, juzgar y

evaluar críticamente la información de salud obtenida. Finalmente, la cuarta competencia, aplicación, concierne a la capacidad para comunicar y utilizar eficazmente la información en la toma de decisiones que promuevan y mejoren la salud (43). Estas competencias son indispensables para una gestión efectiva de la información proveniente de la evidencia en salud, para la toma de decisiones en la salud personal y comunitaria.

La comprensión y la evaluación adecuada de los resultados de los estudios de síntesis de la evidencia en salud presentan desafíos significativos debido a la complejidad estadística y la diversidad de las métricas utilizadas en la expresión de los resultados. Esta situación se agrava cuando consideramos la amplia gama de usuarios destinatarios de esta información y su diferente nivel de alfabetización en salud y estadística. Sin embargo, muchos de los cuales no poseen el nivel de alfabetización en salud y estadística necesario para acceder, entender, evaluar y aplicar eficazmente estos hallazgos en la toma de decisiones.

Investigaciones han evidenciado una comprensión deficiente de la bioestadística y la interpretación de resultados de investigación incluso entre profesionales sanitarios, subrayando una brecha significativa en el conocimiento esencial para una práctica basada en la evidencia. Los estudios realizados (44–48) muestran consistentemente bajos niveles de conocimiento en bioestadística entre médicos en formación especializada. Los médicos especialistas muestran también, un nivel de alfabetización estadística que oscila entre bajo y medio. Nivel que se ha mantenido a lo largo del tiempo según se observa entre los resultados de estudios que han analizado este asunto a lo largo de los años (49–53). A pesar de una precisión en sus respuestas sobre valores estadísticos generalmente baja, los médicos suelen presentar una alta confianza en sus conocimientos, lo que sugiere una sobreestimación peligrosa de su entendimiento en estadística.

Es fácil entender que la situación no mejora cuando se amplía la mirada al público en general y pacientes. Estudios revelan que un alto porcentaje de la población adulta en EE.UU. y Europa presenta niveles de alfabetización en salud y estadística insuficientes para gestionar eficazmente su bienestar, su salud y navegar por el sistema de atención sanitaria (54,55), sobre todo en el complejo mundo de la síntesis de la evidencia.

Esta situación resalta la importancia de avanzar hacia una mayor democratización y divulgación de los conocimientos de salud, enfatizando la participación activa de los pacientes en las decisiones que afectan a su atención médica. El tránsito de un modelo de atención paternalista a uno más colaborativo, donde los pacientes y los profesionales sanitarios trabajan conjuntamente, es esencial para mejorar los resultados de salud, incrementar la satisfacción con los servicios recibidos, reducir costos y optimizar la experiencia de atención. Para lograr este objetivo, es imprescindible garantizar que todos los tomadores de decisiones en salud accedan a información clara y entendible, liberándolos de la dependencia exclusiva de los conocimientos proporcionados exclusivamente por los proveedores de cuidados de salud, quienes pueden estar influidos por sus propios sesgos y limitaciones. Este enfoque subraya la necesidad de implementar estrategias efectivas que mejoren la alfabetización en salud y el acceso y comprensión a la información relevante (56–58).

1.3.4. Diferencias de resultados entre RS/MA ante la misma pregunta de investigación

La existencia de discrepancias en los resultados entre distintas RS/MA que abordan la misma pregunta de investigación es un fenómeno documentado y que introduce retos para todas las partes interesadas (59). Los resultados de las RS/MA no es un proceso exento de subjetividad, lo que puede conducir a conclusiones divergentes, independientemente de la metodología aplicada para recopilar o analizar los datos (60). La variabilidad en los resultados puede deberse no solo a diferencias metodológicas sino también a la presencia de sesgos, incluidos aquellos introducidos por los autores con intereses personales o financieros en los resultados (22). La divergencia en los hallazgos entre distintas RS/MA puede atribuirse a numerosas decisiones y juicios, a veces arbitrarios, tomados durante las distintas fases del proceso de investigación (61,62).

1.4. SOLUCIONES A LAS LIMITACIONES ESTRUCTURALES DE RS/MA

1.4.1. Revisiones rápidas

Las revisiones rápidas (RRp) representan una respuesta metodológica al desafío de proporcionar una síntesis oportuna de evidencia científica en situaciones donde se requiere una toma de decisiones rápida y/o los recursos son limitados. Esta modalidad se diferencia de las RS/MA en que se pueden lograr resultados en un corto plazo de tiempo, típicamente entre 5 a 12 semanas, simplificando o modificando el proceso u omitiendo directamente ciertos pasos metodológicos de las RS/MA (63). El objetivo final de las RRp es mantener un equilibrio entre la rapidez y la fiabilidad de la evidencia generada. Las RRp no reemplazan la profundidad y exhaustividad de las RS/MA tradicionales, más bien, proporciona una alternativa viable para satisfacer las necesidades inmediatas de información, y equilibra eficazmente el tiempo de producción y la integridad metodológica.

La necesidad de desarrollar un marco para las RRp fue reconocida por la OMS en 2017, cuando publicó su primera guía práctica destinada a apoyar la implementación de políticas sanitarias y sistemas de salud basados en evidencia (64). Además, la Cochrane Collaboration adaptó sus procesos durante la pandemia de SARS-CoV-19 para establecer estándares mínimos que garantizaran la calidad y rapidez en la producción de RRp (31). De la misma manera la JBI reconoce el valor de las RRp como parte del ecosistema de la síntesis de evidencia, recomendando mantener el rigor y transparencia exigida para todas sus revisiones (65). La continua evolución y refinamiento metodológico de las RRp por parte de grupos como el Cochrane Rapid Review Methods Group (66–70) subraya la importancia de las RRp en el panorama actual de la síntesis de evidencia.

A pesar de su valor práctico, especialmente para gestores sanitarios y elaboradores de guías clínicas que operan bajo restricciones de tiempo y presupuesto (71), existen preocupaciones respecto a su validez frente a las RS/MA, tradicionalmente vistas como

el patrón oro en la síntesis de evidencia (72). Las RRp pueden conducir a resultados divergentes en comparación con las RS/MA convencionales (73). Sin embargo, desde una perspectiva pragmática, las RRp son consideradas beneficiosas cuando la ausencia de información constituye un riesgo mayor que la posible imprecisión derivada de su metodología adelgazada y acelerada. Este enfoque se justifica especialmente en áreas donde las decisiones no pueden esperar a la realización de RS/MA exhaustivas.

1.4.2.Revisiones sistemáticas vivas

Una revisión sistemática viva (Living Systematic Review: LSR) es una metodología de síntesis secundaria diseñada para mantener a las RS/MA actualizadas mediante la incorporación de nuevas evidencias relevantes tan pronto como estén disponibles (74). La LSR se refiere a la forma y periodicidad en que se actualizan las RS/MA, no es una metodología nueva para llevarlas a cabo. Este enfoque responde a la necesidad de que los resultados de la síntesis secundaria de evidencia se mantengan actualizados. A diferencia de las revisiones sistemáticas convencionales, las LSR están caracterizadas por una monitorización activa y continua de las bases de datos incluidas, realizando búsquedas periódicas y se realizarán las actualizaciones de forma inmediata, integrando los nuevos datos relevantes que cumplen con los criterios de inclusión establecidos a los resultados obtenidos previamente. Las LSR conservan la precisión metodológica de las RS/MA, pero al actualizarse constantemente con los resultados recientes se asegura que se mantenga la relevancia y confianza en la certeza de los resultados (75). Las LSR resultan particularmente valiosas en campos donde existe alta tasa de publicaciones, la incertidumbre es prevalente y el nuevo conocimiento puede influir significativamente en las decisiones de política o práctica (76). La implementación de las LSR es facilitada por la automatización y la participación de grupos amplios de investigadores, lo que reduce la carga de trabajo y mejora la eficiencia en el proceso de revisión (77). La frecuencia de actualización de una LSR suele ser flexible, pero según recomienda la Cochrane Collaboration, generalmente implica realizar al menos búsquedas mensuales, ajustándose a la velocidad con la que se genera nueva evidencia en el campo estudiado (74).

1.4.3.Automatización del proceso de síntesis secundaria

El proceso de automatización en una revisión sistemática puede entenderse como la integración de herramientas tecnológicas que optimizan varias etapas del proceso de revisión. Estos procesos han estado presentes como una necesidad desde hace tiempo (78). El objetivo principal de la automatización en la investigación de síntesis de evidencia es mejorar la velocidad, eficiencia y precisión, reduciendo la carga manual, minimizando los errores, y aumentando la calidad, reproducibilidad y actualidad de los estudios (79,80). Esto no solo facilita la realización de estudios más robustos, sino que también democratiza el acceso a metodologías rigurosas para equipos con recursos limitados. Este proceso puede implicar el uso de software de soporte, conjunto de herramientas diseñadas para automatizar y facilitar tareas específicas, como la de duplicación, selección de estudios, extracción de datos y gestión de referencias, mejorando la eficiencia del proceso sin reemplazar la

supervisión y el juicio humano o herramientas basadas en modelos de lenguaje como los *large language models* (LLM). Estas son herramientas basadas en inteligencia artificial (IA) que utilizan procesamiento de lenguaje natural para comprender, analizar y generar texto, permitiendo automatizar tareas complejas como definir las estrategias de búsqueda, clasificación de estudios, la síntesis de información y la generación de resúmenes con un mayor nivel de autonomía y comprensión contextual (81). Con el objetivo de facilitar el uso de las herramientas de automatización se generan las plataformas digitales colaborativas. En el contexto de revisiones sistemáticas es un entorno integral que combina software de soporte para automatizar tareas específicas, junto con herramientas avanzadas basadas en LLM para buscar, analizar, clasificar y sintetizar información, facilitando la colaboración en tiempo real entre los miembros del equipo de revisión y automatizando todo el proceso. La efectividad de diversas herramientas de automatización continúa bajo evaluación constante para determinar su utilidad (70,72,79,82–88). Sin embargo, hasta la fecha está siendo lenta la adopción de herramientas de automatización en el proceso de elaboración de estudios de síntesis de evidencia. Las razones principales de este hecho son el desconocimiento de su existencia, la falta de confianza en la automatización, la complejidad de la programación y los desafíos de implementación (89,90).

En el año 2014 se creó el “Systematic Review Toolbox” (91), como catálogo online de diferentes herramientas y guías para prestar asistencia en el proceso de elaboración de revisiones sistemáticas en diferentes áreas. En su última actualización, mayo 2022, incluye 235 herramientas de software y 112 documentos de guía, algunas de acceso libre, otras de pago, útiles para la asistencia de diferentes pasos de diferentes formas de revisión sistemática. Numerosas herramientas de IA son presentadas cada semana (92). Dado el gran número de herramientas de IA disponibles para las diferentes etapas del proceso de síntesis de evidencia, es importante seleccionar la herramienta más adecuada para la etapa del proceso de síntesis en el que se precise asistencia (93,94). Es importante señalar que hasta la fecha las herramientas de automatización asisten a diferentes partes de la elaboración de la síntesis secundaria, pocas herramientas abordan el proceso completo, pocas herramientas abordan el asunto de la actualización o presentación de resultados “en vivo”, y ninguna está diseñada de forma específica para la síntesis terciaria de la evidencia.

La supervisión humana es crucial en el uso de herramientas de automatización (95) especialmente en el contexto de síntesis de evidencia (96). Si bien las tecnologías como el software de soporte y los LLM ofrecen enormes beneficios al acelerar procesos y mejorar la eficiencia, su implementación sin control adecuado puede generar problemas que comprometan la calidad y confianza de los resultados. El control y supervisión por parte de un humano, conocido como funcionamiento semiautomático (68), permite la corrección de errores, el control de sesgos, la evaluación de la relevancia contextual, el manejo de la complejidad y toma de decisiones críticas, así como garantía de transparencia (84). Aunque estas herramientas pueden mejorar significativamente la eficiencia en ciertas etapas del proceso, como la búsqueda de literatura, el cribado de títulos y resúmenes, y la extracción de datos, su capacidad para manejar de manera autónoma y completa todo el proceso de

revisión sin supervisión humana está todavía lejos de ser una realidad. La combinación de un revisor humano con programas automáticos para el cribado inicial de estudios ha demostrado ser eficaz, al menos en revisiones de estudios de intervención (97–100).

1.4.4. Mejorar el acceso y comprensión de los resultados de síntesis de la evidencia en ciencias de salud

Para abordar el reto de la falta de alfabetización en salud y estadística, existen dos enfoques principales: el primero busca elevar directamente el nivel de alfabetización entre las partes interesadas, mientras que el segundo se enfoca en simplificar la presentación de la información científica para hacerla más accesible.

La mejora de la alfabetización en salud se considera un elemento crucial para todos los usuarios del conocimiento de salud, especialmente para pacientes, cuidadores y la población en general. Optimizando la alfabetización en salud no solo capacita a las personas para tomar decisiones seguras e informadas sobre su salud, lo que conduce a mejores resultados sanitarios, adherencias a tratamientos y cuidados efectivos y a una menor carga económica, sino que también reduce las desigualdades sanitarias (101).

Incrementar la alfabetización implica desarrollar programas educativos integrales en varios niveles, destinados a fortalecer las habilidades necesarias, de todas las partes interesadas, para acceder, entender y aplicar información en salud. A pesar de la nobleza de estos programas y su necesidad indiscutible, las intervenciones diseñadas para aumentar el nivel de alfabetización han tenido un impacto variable en los resultados de salud, especialmente entre los grupos socioeconómicamente desfavorecidos. Con solo un 62% de estas intervenciones mostrando efectos positivos, y a menudo solo parcialmente (102). Esta situación sugiere que, aunque mejorar la alfabetización del público en general es un objetivo loable y necesario, alcanzar un alto nivel de alfabetización universalmente puede ser un proceso prolongado, costoso y complicado aún más por la creciente complejidad del conocimiento en salud.

El segundo enfoque se presenta como una solución más práctica y fácil de aplicar. El objetivo es simplificar la forma en que se presentan los resultados de la investigación, haciéndolos más fáciles de entender independientemente del nivel de alfabetización en salud y estadística. Este método no pretende reemplazar la educación en alfabetización, sino complementarla, proporcionando a las personas herramientas inmediatas para acceder y comprender la información de salud relevante sin requerir una amplia formación académica. La dificultad en interpretar estadísticas complejas como las correlaciones, riesgos relativos, diferencias medias estandarizadas u otras (103) destaca la necesidad de desarrollar estrategias de comunicación efectivas, como el uso de resúmenes en lenguaje sencillo (Plain Language Summaries, PLS), y la presentación de datos a través de medios visuales intuitivos, que permita a una audiencia más amplia tomar decisiones informadas basadas en la mejor evidencia disponible (104). Esto incluye el uso de términos cualitativos para describir la magnitud de los efectos y la integración de estas prácticas en la presentación de resultados estadísticos, en todas las publicaciones científicas. Este enfoque

pragmático para la democratización y divulgación de la información en salud representa un paso adelante para involucrar a todas las partes interesadas en el proceso de toma de decisiones, independientemente de su nivel previo de alfabetización en salud o estadística. Con este objetivo, surgen iniciativas internacionales que promueven la simplificación del mensaje científico, como los PLS, exigidos por instituciones como Institutos Nacionales de Salud de EE. UU. (105), el Instituto Nacional de Investigación en Salud y Atención del Reino Unido (106), la Agencia de Investigación y Calidad de la Atención Médica de EE. UU., en colaboración con sus Centros de Práctica Basada en Evidencia (107), y Cochrane Collaboration que exige PLS para todas las revisiones publicadas en la Biblioteca Cochrane (108). En el mismo sentido, en las guías de práctica clínica se resalta la importancia de que las recomendaciones sean presentadas en un formato claro y comprensible, facilitando que tanto los profesionales de la salud como los pacientes comprendan el sentido de la recomendación y como se llegó a ella (109).

1.4.5. Síntesis terciaria de la evidencia

1.4.5.1. Definición de síntesis terciaria de la evidencia

Es una realidad que diferentes RS/MA sobre la misma pregunta de investigación pueden arrojar conclusiones contradictorias, por ello, surge como intento de solución la síntesis terciaria de evidencia, que supone el análisis y agregación de los resultados de RS/MA con similares preguntas de investigación, como medio para acercarse a una comprensión más completa de la "verdad" científica. La confianza exclusiva en los hallazgos de una única RS/MA puede ser arriesgada, dado que la suma de múltiples análisis puede ofrecer una perspectiva más ajustada y menos errónea sobre un campo de conocimiento específico (110). Esto se debe, en parte, a que un único metaanálisis de una comparación de tratamientos para un resultado específico proporciona una visión restringida, especialmente cuando existen múltiples tratamientos o varios resultados importantes que considerar (111).

Frente a la diversidad de resultados obtenidos de diferentes RS/MA sobre idénticas cuestiones de investigación, y el desafío adicional que representa la sobreproducción de estas RS/MA, a veces en mayor cantidad que los propios estudios primarios (112), que complica la gestión de la evidencia (113,114) y desborda la capacidad de los profesionales de salud para mantenerse actualizados con la evidencia más reciente (115), el desarrollo de nuevas metodologías de síntesis se ha mostrado como una necesidad fundamental. En este sentido se desarrolla la síntesis terciaria, proponiéndose como una solución metodológica para abordar tanto la variabilidad en los resultados de las RS/MA existentes como la sobreabundancia de producción, facilitando la incorporación y aplicación efectiva de esta información en la práctica clínica y la toma de decisiones médica (116).

La síntesis terciaria de la evidencia, (emplearemos el acrónimo URT para representar este nivel de estudio) representa una metodología avanzada para compilar sistemáticamente y evaluar críticamente múltiples RS/MA publicadas sobre las mismas preguntas de investigación y que cumplan criterios de elegibilidad predefinidos (117–120) a través de un proceso meticuloso, riguroso y transparente (121,122). Estos estudios están diseñados

para abordar la creciente saturación de publicaciones científicas y la diversidad en los resultados que se presentan entre ellas, facilitando así una visión más clara y organizada de la evidencia disponible en áreas con una alta producción de publicaciones (17). Los URT se enfocan principalmente en resumir la calidad de los estudios, identificar resultados de interés (RdI), magnitudes de efecto, incertidumbre, heterogeneidad y posibles sesgos en un campo de investigación amplio, evaluando principalmente el efecto de intervenciones sobre diferentes condiciones o las asociaciones epidemiológicas de una exposición a un factor de riesgo. El resultado final de un URT, desde una perspectiva clínica, es una valoración resumida, estandarizada y sistemática de múltiples estudios de intervención o exposiciones epidemiológicas basados en RS/MA publicados. Tienen sentido realizarse y dan información cuando para la pregunta de investigación en estudio existe un número suficiente de RS/MA publicados, en el caso de que el número de RS/MA sea limitado o nulo, no tiene sentido realizar un trabajo de síntesis terciaria, y sería más adecuado realizar una RS/MA (123). Además de sintetizar la evidencia existente, los URT desempeñan un papel de mapear la evidencia disponible y de identificar lagunas de investigación en el ámbito de la salud, facilitando así la dirección de futuras investigaciones hacia estas áreas insuficientemente exploradas (124,125). Este enfoque metodológico se aplica cada vez con mayor frecuencia, reflejando un patrón similar al observado con el aumento de estudios de investigación primaria que impulsó el desarrollo de las RS/MA como herramientas de síntesis secundaria (126). El ritmo del crecimiento del número de publicaciones de RS/MA ha crecido de forma exponencial desde 1995, manteniendo una relación con el número de publicaciones de ensayos controlados aleatorizados (ECA) por debajo de 1 hasta el 2017 (0.87) siendo un dato llamativo considerando que en el año 1995 dicha relación era menor de 0.1 (0.045) (127). Este ritmo de crecimiento de publicaciones no ha dejado de crecer, llegando a superar en el 2023 el número de entradas en Pubmed de RS/MA con respecto a ECAs, con una relación por encima de 1 (1.28). Esto ha llevado a que la publicación de síntesis terciaria haya experimentado un crecimiento exponencial en las dos últimas décadas, evidenciando un incremento octuplicado entre 2009 y 2020, con la notable cifra de un trabajo de síntesis terciaria publicado diariamente en el año 2020 (128). Las URT están diseñadas para servir a un amplio espectro de audiencias involucradas en el ámbito de la salud, destacando su rol fundamental en informar y orientar la toma de decisiones basada en evidencia. El público objetivo principal de las URT son los responsables de la toma de decisiones y gestores políticos en atención sanitaria, pero también resultan de interés para proveedores de atención médica, así como pacientes y población general (126,129). La síntesis terciaria pretende ocupar un lugar destacado en la jerarquía de la evidencia, y servir como un pilar esencial en la medicina basada en la evidencia y marcar un hito en el esfuerzo continuo por mejorar la calidad, accesibilidad y aplicabilidad de la información científica en el ámbito de la salud al ofrecer una visión completa, resumida y sintetizada de la evidencia de las RS/MA relevantes publicadas.

1.4.5.2. Metodología de la síntesis terciaria

La metodología empleada en los estudios de síntesis terciaria se deriva directamente de las prácticas establecidas en las revisiones sistemáticas, las cuales se rigen por estándares de conducta definidos para asegurar la rigurosidad, validez, transparencia y fiabilidad de sus resultados. No obstante, dada la naturaleza diferenciada de la síntesis terciaria, donde la RS/MA misma constituye la unidad de análisis, los enfoques metodológicos para su realización divergen de aquellos aplicados en la síntesis de estudios primarios.

La metodología para llevar a cabo las URT ha sido objeto de formalización por parte de diferentes autores a lo largo de los años, intentando establecer criterios específicos que reflejen las particularidades de este nivel de síntesis. Desde su concepción hace más de cuatro décadas (130), entonces denominada “*meta-metaanalysis*”, por incluir exclusivamente metaanálisis como unidad de estudio, la metodología ha evolucionado significativamente, ampliando su enfoque para incluir tanto las revisiones con metaanálisis como aquellas sin él, lo que ha llevado a cambios progresivos en la metodología, que permite a los autores realizar diferentes tipos de análisis en función de los objetivos, así como presentar de diversas maneras los resultados finales (131–133).

La presencia de particularidades propias de la síntesis terciaria se ha puesto de relieve en la falta de rigor metodológico que frecuentemente se observa en algunas URT publicadas, así como inconsistencias entre los protocolos y las publicaciones finales, evidenciando la necesidad de un desarrollo metodológico más robusto (134,135).

Históricamente, la metodología de las URT ha presentado lagunas, especialmente en fases avanzadas y sensibles del proceso de síntesis (136–138). La escasez de guías metodológicas específicas para estas revisiones, tanto por parte de agencias de Evaluación de Tecnologías Sanitarias como en la literatura académica, ha sido un vacío crítico para los autores interesados en acometer investigaciones de síntesis terciarias (139).

La estructuración de la metodología para la realización de síntesis terciaria ha sido influenciada significativamente por las recomendaciones de la Colaboración Cochrane, que desde su primera mención en el Handbook de Cochrane en la versión del 2008, han proporcionado un marco de referencia para la ejecución de estas revisiones (126,140).

En el esfuerzo por sistematizar y mejorar la práctica de síntesis terciaria de la evidencia, se ha realizado un análisis exhaustivo de la literatura existente para intentar presentar una guía de los pasos metodológicos detallada. Esta guía metodológica de la síntesis terciaria, articulado en 17 pasos clave, es el resultado de un proceso de revisión y síntesis de las directrices propuestas por la Colaboración Cochrane (126), el JBI (141) y otros estudios pioneros en el campo. El propósito de este enfoque integrador es proporcionar, en lo posible, un marco metodológico cohesivo, amplio y riguroso.

La guía propuesta se concibe como un recurso integral, diseñado para orientar a investigadores y profesionales en la realización de URT, desde la concepción inicial hasta la diseminación de los hallazgos. Al adherirse a este marco metodológico, se espera mejorar

la calidad y la consistencia de la síntesis terciaria, facilitando así la interpretación y la aplicación de sus resultados en la mejora de prácticas, políticas y resultados en salud. A continuación, se exponen las etapas metodológicas recomendadas por diferentes autores (**Tabla 1**). Se hace una descripción detallada de los pasos globales recomendados para esta metodología en el **ANEXO 1**.

Tabla 1. Pasos Metodológicos de URT según distintos autores

		Thompson 2010 (129)	Cooper, 2012 (143)	Pollock, 2016 (118)	Lunny 2017-18 (137,142)	Aromataris, 2020 (117)	Cant 2022 (114)	Belbasis 2022 (123)	Pollock 2023 (126)	Choi 2023 (17)	Recomen- dación Global	The Umbrella Collabora- tion®
1	Título e Información del Autor			•		•	•			•	•	•
2	Introducción			•		•	•			•	•	
3	Fuentes informativas (bases de datos)					•				•	•	•
4	Definición Clara de la Pregunta de Investigación, Objetivos y alcance	•	•	•	•	•	•	•	•	•	•	
5	Registro y Protocolo Predefinido					•	•				•	
6	Desarrollo de Criterios de Inclusión	•		•	•	•	•	•	•	•	•	
7	Estrategia de Búsqueda Exhaustiva	•	•	•	•	•	•	•	•	•	•	•
8	Selección Rigurosa de Estudios			•		•	•	•	•	•	•	•
9	Características de los estudios incluidos			•		•	•		•	•	•	
10	Recolección de Datos de RS incluidos			•	•	•		•	•	•	•	•
11	Evaluación del Solapamiento (Overlap)		•			•			•	•	•	

		Thompson 2010 (129)	Cooper, 2012 (143)	Pollock, 2016 (118)	Lunny 2017-18 (137,142)	Aromataris, 2020 (117)	Cant 2022 (114)	Belbasis 2022 (123)	Pollock 2023 (126)	Choi 2023 (17)	Recomen- dación Global	The Umbrella Collabora- tion®
12	Evaluación de la Calidad y Riesgo de Sesgos RS/ECA incluidos en S3+E		•	•	•	•	•		•	•	•	
13	Análisis y presentación de los Resultados	•	•	•	•	•	•	•	•	•	•	•
14	Evaluación Certeza Resultados		•	•	•			•	•	•	•	•
15	Limitaciones de la síntesis terciaria						•			•	•	
16	Discusión implicación de resultados					•	•	•		•	•	
17	Conclusiones			•		•	•		•		•	

1.5. INNOVACIÓN EN LA SINTESIS DE LA EVIDENCIA: EL PAPEL DE THE UMBRELLA COLLABORATION®

En el contexto actual del ámbito sanitario, se reconoce la necesidad de que todos los actores implicados en la toma de decisiones, desde los profesionales de la salud hasta los pacientes pasando por gestores de políticas sanitaria e investigadores, dispongan de acceso a información de alta calidad, que sea tanto cierta y completa, como que este pertinentemente actualizada, fundamentada en el nivel más elevado de evidencia científica existente y que sea comprensible. Este imperativo emerge de la necesidad crítica de fundamentar las decisiones sanitarias en datos fiables y obtenidos de manera transparente y reproducible, que no solo maximicen los beneficios frente a los riesgos de las intervenciones, sino que también promuevan la equidad y eficiencia en el acceso y los costes de la atención sanitaria.

Las URT presentan en sí mismas muchas limitaciones asociadas a su metodología (143,144), y abordan de manera efectiva una de las principales limitaciones estructurales de las RS/MA, ya que ofrecen una solución al problema de los resultados contradictorios que suelen presentar varias RS/MA sobre la misma pregunta de investigación. Al sintetizar estos estudios, las URT proporcionan una visión consolidada que minimiza estas discrepancias. Sin embargo, las URT comparten muchas de las limitaciones estructurales de las RS/MA descritas anteriormente, en cuanto a su tiempo de elaboración alargado, su facilidad para quedar desactualizadas y la complejidad para la comprensión de los resultados por parte de los usuarios.

The Umbrella Collaboration® (TU®) se presenta como una metodología innovadora en la síntesis terciaria de evidencia que busca abordar de manera integral las limitaciones estructurales inherentes a las RS/MA, extensivas a la síntesis terciaria. Estas limitaciones, incluyendo tiempos de ejecución alargados, actualización insuficiente, dificultad en la comprensión de resultados y falta de concordancia entre estudios, se abordan mediante una herramienta informática semiautomatizada que requiere y facilita la participación humana en el proceso. La síntesis terciaria en TU® permite analizar y agrupar los resultados de estudios de síntesis secundaria, utilizando como unidad de análisis RS/MA exclusivamente.

Este enfoque metodológico no solo simplifica el proceso de síntesis, sino que se inspira en metodologías reconocidas y ya implementadas para mitigar las principales limitaciones estructurales de la síntesis de evidencia. TU® emplea una estrategia similar a las RRP, ya que simplifica el proceso de síntesis a los pasos esenciales, permitiendo así acortar significativamente los tiempos de ejecución. Además, incorpora el concepto de LSR, mejorándolo de manera sustancial, al actualizar la evidencia de manera frecuente, incluso cada 24 horas o bajo demanda, lo cual es un avance importante en términos de mantener la relevancia de la información en campos de rápida evolución. Por otro lado, TU® presenta los resultados de forma visual e intuitiva, facilitando su comprensión y accesibilidad para usuarios con diversos niveles de formación estadística, lo que se alinea con los objetivos de democratización y transparencia en el acceso a la información.

Para cumplir estos objetivos, TU® utiliza un programa informático asistido en diferentes fases por IA, que facilita la generación de términos de búsqueda, la búsqueda y deduplicación de estudios, la agregación de resultados, la visualización final de datos en una plataforma de acceso público y la actualización permanente del proyecto. Este diseño reduce, aunque sigue requiriendo, la supervisión humana en ciertas etapas críticas. TU® se presenta como una novedad, al ser la primera herramienta de automatización diseñada específicamente para síntesis terciaria, y que asiste en todo el proceso de revisión terciaria.

No obstante, si bien la automatización, como propone TU®, promete un futuro más eficiente y ágil para la síntesis de la evidencia, también presenta retos y limitaciones inherentes al uso de tecnologías automatizadas en un campo tan complejo y crítico como el de la salud. La automatización permite acelerar considerablemente el proceso, limitar el trabajo y los errores humanos y reducir los costes a largo plazo, ya que optimiza el análisis de grandes volúmenes de datos y facilita la actualización continua de la evidencia. Estos beneficios son particularmente relevantes en un contexto donde la evidencia debe mantenerse permanentemente actualizada para responder a las demandas de un entorno sanitario en constante evolución. Además, la automatización tiene el potencial de minimizar los sesgos humanos, ofrecer consistencia, transparencia y reproducibilidad, y democratizar el acceso a herramientas avanzadas y resultados de alta calidad.

Sin embargo, este enfoque no está exento de desafíos, especialmente en lo que respecta a la interpretación contextual de los datos y a la necesidad de juicio crítico humano. Aunque la automatización agiliza el proceso de síntesis, puede carecer de la capacidad para evaluar factores contextuales específicos de cada estudio, lo que podría limitar la comprensión

de variaciones sutiles pero importantes. El riesgo de sesgos algorítmicos también es una preocupación, pues los sistemas automatizados están sujetos a las limitaciones de sus datos de entrenamiento, que pueden reproducir o amplificar sesgos preexistentes si no se manejan con precaución. La falta de transparencia en algunos algoritmos de IA puede convertirlos en una “caja negra” difícil de auditar, lo cual afecta la confianza y la verificación de los resultados sintetizados.

Además, la automatización enfrenta dificultades en el manejo de datos cualitativos, un componente esencial en muchas investigaciones de salud. Mientras que los sistemas automatizados funcionan bien con datos estructurados y cuantitativos, tienden a perder matices al analizar datos cualitativos, lo que podría llevar a interpretaciones incompletas. Para superar estas limitaciones, la automatización, de momento, debe integrarse con controles humanos rigurosos y con una supervisión experta que permita garantizar la integridad y relevancia de la evidencia.

En definitiva, la automatización de la síntesis de la evidencia se perfila como el futuro del campo de la síntesis de evidencia, representando una revolución en la manera en que se genera, analiza y difunde el conocimiento. Al preparar a los investigadores y profesionales de la salud para esta transición, se abre la posibilidad de una síntesis de evidencia más rápida, precisa y accesible, capaz de mantener su relevancia en un mundo donde la información se actualiza a una velocidad sin precedentes. Si bien aún se requiere una implementación cuidadosa y una validación exhaustiva de estos sistemas, la combinación de la eficiencia de la tecnología con el juicio crítico humano es el camino para maximizar los beneficios y mitigar los riesgos, asegurando que la automatización cumpla su promesa de mejorar la práctica basada en evidencia de manera ética y efectiva.

TU® representa una metodología conceptual que requiere validación empírica en comparación con los métodos tradicionales, como las URT, para determinar su efectividad en la práctica. Este estudio busca precisamente evaluar esta metodología, cuantificando su precisión, utilidad, y comparabilidad con los estándares actuales, y así contribuir al conocimiento y comprensión del papel de la automatización en la síntesis terciaria de evidencia. En última instancia, el objetivo es mejorar el acceso a la evidencia científica, optimizar su aplicación en la práctica clínica y facilitar la toma de decisiones informadas.

Al llevar a cabo esta investigación, se aspira a contribuir significativamente al cuerpo de conocimiento en el ámbito de la síntesis de la evidencia, ofreciendo perspectivas sobre la eficacia de una metodología emergente en comparación con el enfoque más tradicional. Esta tesis doctoral no solo busca validar una metodología experimental, sino también reflexionar sobre el estado actual y el futuro de la síntesis de evidencia en salud, explorando caminos hacia la optimización de la recopilación, análisis y presentación de la información científica para mejorar la toma de decisiones en salud.

2

HIPÓTESIS Y OBJETIVOS

2.1. HIPÓTESIS

La hipótesis de esta tesis doctoral se centra en la validación de TU®, una herramienta semiautomática diseñada para realizar síntesis terciaria de la evidencia científica en salud. Se plantea que TU® puede representar un avance metodológico en el campo de la síntesis de la evidencia, al ofrecer una herramienta integrada en una plataforma digital web que permita integrar y actualizar de forma continua los resultados de RS/MA publicados en Medline, reduciendo la dependencia del trabajo humano y acelerando el proceso de síntesis. Asimismo, se considera que esta metodología podría mejorar la accesibilidad y la comprensión de los resultados en comparación con las URT, facilitando su uso en la práctica clínica, la investigación y la docencia.

2.2 OBJETIVOS

2.2.1. Objetivo principal

Evaluar si la metodología TU® es capaz de obtener resultados de síntesis terciaria comparables a los obtenidos por las URT en el ámbito de la salud.

2.2.2. Objetivos específicos

1. Evaluar la concordancia de los resultados obtenidos mediante la metodología TU® frente a las URT en el área de geriatría, validando esta concordancia a través de la capacidad del método experimental para identificar resultados de interés (RdI), incluyendo su tamaño de efecto (TE), dirección, significación estadística y certeza, en comparación con los resultados de las URT en estos mismos aspectos.
2. Determinar el tiempo necesario para elaborar proyectos de síntesis terciaria con la metodología TU® y evaluar si este tiempo supone un valor añadido que compensa posibles diferencias en la precisión de los resultados en comparación con las URT.
3. Analizar si el formato de presentación de los resultados con la metodología TU® facilita la comprensión de la información en comparación con los formatos de presentación de las URT, con el objetivo de mejorar la accesibilidad y utilidad de la evidencia para los usuarios finales.

Esta tesis se propone evaluar esta hipótesis mediante un estudio comparativo, analizando los resultados generados por TU® como su alineación con los estándares establecidos por las URT. Se espera que los resultados de este estudio proporcionen evidencia empírica sobre la validez, fiabilidad y utilidad de TU® como herramienta de síntesis terciaria de la evidencia, estableciendo un nuevo estándar en la divulgación y análisis de la investigación en salud.

3

MATERIAL Y MÉTODOS

3.1 INTRODUCCION

En esta sección de "Materiales y Métodos", describimos el diseño y los procedimientos empleados para validar la metodología TU® y de su herramienta informática asistida por IA frente al patrón oro de síntesis terciaria de evidencia que son las URT. Para ello se diseñaron dos estudios, Estudio-I (E-I) y Estudio-II (E-II), para evaluar en primer lugar la comparación de la obtención de resultados entre metodologías (E-I) y la percepción de utilidad de TU® (E-II).

El E-I, realiza una comparación de los resultados obtenidos por TU® y las URT ante la misma pregunta de investigación. Este estudio comparativo se centra en la identificación, tamaño de efecto (TE), dirección del efecto, significación estadística y certeza de los resultados de interés (Rdl) hallados por ambas metodologías, y del tiempo de ejecución de los proyectos. El E-II, evalúa la utilidad percibida de TU® a través de un estudio transversal descriptivo mediante encuestas a profesionales sanitarios de servicios de geriatría de hospitales universitarios de la Comunidad de Madrid.

Este enfoque permite una comparación entre los resultados y la utilidad percibida de la metodología experimental TU®, frente a las URTs. Se describen los recursos empleados para recopilar los datos necesarios, incluyendo bases de datos, software estadístico y otros recursos digitales, y los procedimientos experimentales seguidos, para ambos estudios.

Este enfoque asegura la reproducibilidad de la investigación y facilita una evaluación detallada de las ventajas y limitaciones de TU®, contribuyendo al cuerpo de conocimiento en el campo de la síntesis de evidencia en salud.

3.2 DETALLES DE LOS MÉTODOS ESTUDIO I (E-I): ANÁLISIS COMPARATIVO DE METODOLOGÍAS

3.2.1 Diseño del Estudio I: análisis de resultados de metodologías comparativas

En el E-I, se emplea un diseño comparativo cuantitativo y cualitativo, para evaluar la concordancia entre las dos metodologías de síntesis terciaria a estudio: TU® y las URT. Se seleccionó una muestra de URT de la base de datos Medline (vía Pubmed) en geriatría, replicando el proceso de síntesis con TU®. El objetivo principal es determinar si TU® produce conclusiones comparables a las URT, y cuantificar el tiempo necesario para obtener los resultados.

3.2.2 Variables del Estudio I: análisis de resultados de metodologías comparativas

Se evaluaron las siguientes variables para comparar las dos metodologías de síntesis terciaria: identificación de Rdl, con sus propiedades asociadas como medida de la magnitud del TE, dirección del efecto, significación estadística y la certeza de la evidencia. Adicionalmente, se analizó el tiempo de ejecución de los proyectos de síntesis con la metodología experimental TU® para valorar su eficiencia operativa.

3.2.2.1 Definición de Rdl y comparación de identificación de Rdl entre metodologías

Un "Rdl" en una síntesis terciaria es un dato o hallazgo específico que ha sido previamente identificado y evaluado por una o más RS/MA que han abordado la misma pregunta de investigación. Los Rdl individuales de las RS/MA se agregan para conformar el Rdl resultante de la síntesis terciaria. En el E-I, se analizaron dos aspectos clave de la identificación de Rdl. Primero, se cuantificó el número total de Rdl obtenidos por cada metodología, lo que permitió realizar una comparación descriptiva y la congruencia estadística entre ambas. Segundo, se analizó el factor de incremento de identificación de TU® respecto a URT. Este factor multiplicador se define como el número de veces que TU® identifica más Rdl en relación con URT, proporcionando una medida directa de incremento en la identificación por parte de TU® con respecto a URT.

3.2.2.2 Comparación y Análisis de Tamaño de Efecto de Rdl

El Rdl se cuantifica con una magnitud o medida del efecto (tamaño de efecto o fuerza de una asociación) de una intervención o exposición. En esta disertación utilizaremos el término "Tamaño de Efecto (TE)" para referirnos al valor de la medida de efecto tanto si es para una medida de diferencias que comparan los valores medios entre dos grupos (estrictamente llamado "tamaño de efecto"), como si es una medida de razón que comparan las posibilidades de un evento entre grupos (estrictamente llamados "fuerzas de asociación"). En los estudios de síntesis secundaria y terciaria tradicionales se utilizan diversas métricas como la Diferencia de Medias Estándar (SMD), la Diferencia de Medias (MD), la Diferencia de Medias Ponderada (WMD), el Riesgo Relativo (RR), la Odds Ratio (OR) y la Hazard Ratio (HR). En el marco del proyecto TU®, se identificó como un desafío fundamental la interpretación de las métricas estadísticas utilizadas tradicionalmente. Estas métricas a menudo resultan complejas y poco accesibles para un amplio espectro de lectores. Ante esta problemática, TU® desarrolló fórmulas que normalizan estas métricas y las traducen a términos categóricos coloquiales más intuitivos y comprensibles. Este enfoque se inspira en las clasificaciones propuestas por Jacob Cohen (145), y otros expertos, quienes han sugerido categorías cualitativas como "trivial", "pequeño", "moderado" y "grande" para describir los tamaños de efecto. Cohen asignó a su medida d , un valor de $d = 0.2$ a un efecto pequeño, $d = 0.5$ a un efecto mediano y $d = 0.8$ a un efecto grande, proporcionando un marco intuitivo para interpretar la magnitud de las diferencias estadísticas. En TU®, se extendió este enfoque para incluir no solo los tamaños de efecto de las medias, sino también medidas de fuerza de asociación como el OR, el RR y el HR. Se siguió el abordaje inicial de Hopkins (104) al desafío de interpretación de los valores de diferentes magnitudes de efecto, pero ampliando el enfoque a otras métricas y rangos de valores no analizadas por Hopkins. Se generaron fórmulas específicas, incorporadas al software TU®, que describen el traslado de las medidas estadísticas convencionales utilizadas en RS/MA, a la medida propia de TU® de TE, denominada R_{TU} con sus equivalencias cualitativas en términos categóricos coloquiales: trivial ($0 \geq R_{TU} < 0.1$), small ($0.1 \geq R_{TU} < 0.3$), moderate ($0.3 \geq R_{TU} < 0.5$), large ($0.5 \geq R_{TU} < 0.9$) y very large ($0.9 \geq R_{TU} < 1$).

3.2.2.3 Metodología de la transformación de las métricas a la d de Cohen.

En la comparación de las magnitudes de los TE de los Rdl identificados mediante TU® y URT, nos enfrentamos con el desafío de que los TE empleados en las URT se expresan en diferentes métricas estadísticas. Las URT utilizan una variedad de métricas como la SMD, MD, WMD, RR, OR o HR, mientras que la metodología TU® emplea una métrica propia denominada R_{TU} . Para realizar una comparación directa entre estas metodologías, y poder valorar el comportamiento de la métrica de TU®, R_{TU} , frente a métricas consolidadas fue esencial transformar todas estas métricas utilizadas en las URT a un TE común. Decidimos utilizar la d de Cohen como métrica estandarizada, permitiendo así una interpretación uniforme y clara de los resultados. Ver **ANEXO 2** para justificación detallada de la metodología de la transformación de las métricas de TE a la d de Cohen.

La transformación de diferentes métricas de magnitud de efecto utilizadas en las URT a la d de Cohen permitió una comparación directa y coherente entre los resultados obtenidos por la metodología TU® y las URT. Pero precisó de ajustes metodológicos en distintos niveles de análisis al detectar resultados no ajustados a lo referido en las URT y RS/MA de procedencia.

3.2.2.4 Ajustes metodológicos realizados ante la detección de resultados no ajustados en la comparación de TE entre metodologías comparativas.

La evaluación comparativa entre TE de los Rdl identificados por ambas metodologías se realizó en 3 niveles, con variación de datos según se avanzó del primer conjunto (Nivel O), al segundo (Nivel A) hasta el tercero (Nivel B).

El Nivel O se corresponde con los 66 TE de Rdl identificados por ambas metodologías, con todas las métricas de URT transformados a la d de Cohen, y comparadas con la métrica R_{TU} de TU®.

El Nivel A se mantuvieron las métricas de Rdl de los URT transformados a la d de Cohen, excepto los que se presentaban como métrica de SMD que fueron utilizados directamente como TE sin transformar en la d de Cohen, y un valor fue eliminado por no quedar adecuadamente justificado su procedencia en el URT (corresponde con el Rdl identificado con el código FO15 en la base de datos del proyecto), por lo que se incluyeron 65 datos. Se corresponde con ajustes realizados tras el análisis de los TE discordantes en 2 o más niveles observados en el Análisis Nivel O. En el análisis de los TE discordantes se objetivó que las d de Cohen discordantes que provenían de SMD, el valor en la d de Cohen no se alineaba con los resultados y conclusiones sobre el TE de las RS/MA incluidas en las URT. Esto llevó como primer ajuste no transformar los SMD a d de Cohen; utilizándolos como TE directamente (**ANEXO 5**).

El Nivel B fueron eliminados los TE de URT que originalmente estuviesen en MD o WMD. Lo que redujo el número de datos a 48 en la evaluación Nivel B. La falta de los tamaños muestrales no permitió la estandarización o ajuste de las MD o WMD a SMD y posterior transformación a la d de Cohen, además, se observó que algunos de los casos en los TE

difirieron en más de un nivel de categoría en el Análisis Nivel O de TU® y URT fueron casos donde la extracción fue a partir de la diferencia de medias (MD) o la diferencia de medias ponderadas (WMD).

3.2.2.5 Justificación Niveles de Concordancia TE categorizados

La métrica R_{TU} se corresponde directamente con valores categóricos, trivial, pequeño, moderado, grande y muy grande, pero para poderse comparar directamente con los valores de las URT se unificaron las categorías grande y muy grande en una categoría única grande. Esto se justifica por no darse ningún valor “muy grande” entre los Rdl analizados. A su vez una vez transformados las magnitudes de los TE de las URT a la d de Cohen, estas se clasificaron en cuatro categorías para permitir una interpretación más directa de las diferencias en magnitud del efecto categorizados entre las dos metodologías. Las categorías fueron definidas como:

- Trivial: valores entre 0 y 0.19.
- Pequeño: valores entre 0.2 y 0.49.
- Mediano: valores entre 0.5 y 0.79.
- Grande: valores superiores a 0.8.

La comparación entre las categorías de TE de TU® y URT se realizó en base a la concordancia en estas categorías, dividiendo los resultados en cuatro niveles de concordancia:

1. Concordancia plena: ambas metodologías clasificaron el mismo resultado dentro de la misma categoría (ej. trivial-trivial, pequeño-pequeño...).
2. Concordancia Parcial: cuando la clasificación de categorías entre metodologías difirió en una sola categoría (ej. trivial-pequeño, pequeño-mediano...). Se consideraron que estas diferencias eran asumibles dado que no implicaron una discrepancia significativa en la interpretación práctica de los resultados.
3. Discordancia: cuando una metodología clasificó un resultado en una categoría que difiere en más de un rango respecto a la otra (ej. trivial-mediano, pequeño-grande...). Estas diferencias se consideraron sustanciales y reflejaron una discordancia significativa entre las metodologías.
4. Concordancia consistente: se agruparon los casos de concordancia plena con los de concordancia parcial para proporcionar una visión más amplia de la concordancia entre las metodologías. Este concepto permitió valorar la proximidad en la clasificación de los TE, incluso cuando hubo pequeñas diferencias. Dado que las discrepancias en una categoría no cambiaron significativamente la interpretación práctica de los resultados.

Justificación teórica para la “concordancia parcial”

La definición de “concordancia parcial” se justificó por varios factores. En primer lugar, las categorías definidas (trivial, pequeño, moderado, grande) son construcciones artificiales

diseñadas para agrupar una serie continua de valores en rangos discretos. En cualquier proceso de categorización de variables continuas, se produce una pérdida de precisión, ya que los valores adyacentes pueden caer en diferentes categorías, aunque su diferencia sea muy pequeña (por ejemplo, un valor de trivial de 0.19 y uno pequeño de 0.21). Este es un problema inherente a la categorización de variables continuas. No obstante, dado que el cambio entre categorías es discreto, tiene sentido considerar como "concordancia parcial" cualquier cambio que no exceda una categoría.

En las comparaciones entre metodologías, siempre existe un margen de tolerancia para la variabilidad. Esta variabilidad puede deberse a varios factores, como diferencias en la metodología estadística, la selección de datos o el enfoque interpretativo. En este contexto, dado que las metodologías comparadas evaluaron los mismos resultados, es razonable suponer que, si difieren en la clasificación, lo hacen en pequeños márgenes. Esto se debe a la cercanía de los límites de las categorías definidas. Por lo tanto, cuando TU® clasifica un resultado en una categoría mayor o menor en relación a URT, se puede considerar dentro de una "concordancia parcial". Este concepto está respaldado por el hecho de que la diferencia es menor que el salto del valor de un rango completo de categoría, lo que sugiere que ambas metodologías están muy próximas en su evaluación.

Desde una perspectiva práctica o clínica, la diferencia de una categoría podría no ser significativa. Es decir, una diferencia entre "trivial" y "pequeño", o entre "mediano" y "grande", podría no tener un impacto real en la interpretación de los resultados. La "concordancia parcial" puede justificarse bajo la idea de que la diferencia entre estas categorías es lo suficientemente pequeña como para no alterar las decisiones prácticas o interpretaciones basadas en los resultados.

En muchos análisis estadísticos, se acepta un rango de variabilidad como parte del proceso de inferencia. Por ejemplo, en análisis de intervalos de confianza o pruebas de hipótesis, siempre existe un margen de error debido a la naturaleza probabilística del análisis. De forma análoga, la "concordancia parcial" puede justificarse como una aproximación aceptable, dada la imprecisión inherente a la categorización y las diferencias metodológicas entre URT y TU®.

Cuando la diferencia excede una categoría, se entra en un territorio donde los resultados difieren de manera sustancial, y ya no se puede considerar una "concordancia parcial". A medida que la diferencia entre las clasificaciones crece más allá de una categoría, la discrepancia se vuelve estadística y clínicamente más relevante, lo que justifica que clasifiquemos estas diferencias como "discordancia".

3.2.3 Definición de la dirección del efecto de Rdl y comparación entre metodologías

La dirección del efecto de los Rdl indica si los resultados están asociados de forma favorable o desfavorable con una intervención (como en el caso de intervenciones médicas aplicadas a condiciones de salud) o si reflejan un aumento o disminución del riesgo (en el contexto de

exposiciones a factores de riesgo). En este estudio, se evaluó la dirección de los efectos observados por ambas metodologías. Los resultados se clasificaron en tres categorías: favorable, desfavorable o desconocido para intervenciones, y aumento, disminución o desconocido del riesgo para exposiciones. La comparación de la dirección del efecto entre los Rdl identificados y coincidentes de ambas metodologías permitió analizar la coherencia en la dirección de los efectos identificados.

3.2.4 Definición de la significación estadística de Rdl y comparación entre metodologías

En ambas metodologías, se evaluó la significación estadística de los Rdl para determinar si las diferencias observadas son probablemente genuinas y no producto del azar. Un resultado es estadísticamente significativo cuando la probabilidad de que el efecto observado sea debido al azar es muy baja. Para los Rdl identificados y coincidentes entre metodologías, se analizaron los valores p y los intervalos de confianza de las métricas utilizadas, comparando la significación estadística entre los Rdl coincidentes de TU® y URT.

3.2.5 Definición de certeza de la evidencia de los Rdl y comparación de metodologías

Existen múltiples enfoques para evaluar la calidad de los resultados de síntesis de la evidencia. Esto puede incluir la calidad estadística, que puede ser evaluada mediante escalas como la definida por Papatheodorou (115), la calidad metodológica a través de herramientas como AMSTAR-2 (110), y la confianza o certeza en los resultados de la evidencia sintetizada utilizando el sistema Grading of Recommendations Assessment, Development and Evaluation (GRADE) (146,147). Inicialmente, el sistema GRADE hacía referencia a la "calidad de la evidencia", que describía el grado de confianza en que los resultados de los estudios eran válidos para apoyar decisiones clínicas. Con el tiempo, este término evolucionó a "confianza en las estimaciones", subrayando la precisión con la que se podían realizar recomendaciones específicas basadas en la evidencia disponible (148). Actualmente, GRADE ha adoptado el término "certeza de la evidencia", que ha sido reconocido como una mejor representación de este concepto, ya que engloba tanto la confianza en la magnitud del efecto estimado como en su relevancia para la toma de decisiones (149). Debido a que la evaluación de la certeza por el sistema GRADE es el más reconocido y recomendable (150), y es el sentido de la valoración de calidad que más se aproxima al concepto de certeza utilizado en TU®, que busca evaluar cual es la confianza que se puede tener en los resultados de síntesis obtenidos, optamos por esta evaluación para el estudio de concordancia de la calidad entre las dos metodologías, empleando por ello solo a las URT seleccionadas que evaluaban la certeza de la evidencia con el sistema GRADE frente al nivel de certeza de TU® basado en una clasificación de certeza utilizando el análisis de sentimiento propio del procesamiento natural del lenguaje (151) y LLM (152). Ambos métodos se refieren al mismo concepto de confianza en los resultados, pero emplean escalas diferentes. El sistema GRADE utiliza clasificaciones nominales (*very low*, *low*, *moderate*, *high*), mientras que TU® emplea un sistema de puntuación basado en el análisis

de sentimiento comúnmente utilizado en el procesamiento natural del lenguaje y LLM, y que asigna un nivel de certeza basado en una escala de polaridad de negativo a positivo, que abarca de -1 a +1, de los resúmenes de las RS/MA. Ambas escalas fueron reclasificadas a un rango cuantitativo similar (0 - 1) para facilitar la comparación. Para armonizar estos sistemas dispares y permitir comparaciones cuantitativas, se reclasificaron la escala bipolar de TU® a la escala de razón de 0 a 1 con la fórmula: $X = (A1 + 1) / 2$ (siendo A1 el valor de certeza cuantitativo otorgado por TU® en la escala original). Igualmente se reclasificaron los valores nominales de GRADE a valores cuantitativos, aplicando la siguiente regla de reclasificación: Very Low=0, Low=0.33, Moderate=0.66 y High=1. En los casos en que hubiese varias valoraciones GRADE para el mismo Rdl se aplicaba una media aritmética. Este ajuste matemático permite que cada puntuación en la escala original se corresponda directamente con un punto específico en la nueva escala. Este método asegura que los valores extremos en la escala original se traduzcan en los correspondientes extremos de la nueva escala, manteniendo la relación proporcional entre las puntuaciones intermedias. El propósito de esta reclasificación es doble: primero, facilita la interpretación directa y la comparación estadística de los resultados entre dos sistemas diferentes de calificación de la certeza en resultados, permitiendo que los análisis estadísticos, como las correlaciones, sean posibles. Segundo, alinea las evaluaciones de certeza de TU® con un marco reconocido y ampliamente utilizado como GRADE, lo que aumenta la relevancia clínica y la aceptación de los resultados de TU®. Esta normalización es crucial para la validación de TU® como una herramienta fiable y comparable a las metodologías tradicionales en la síntesis de evidencia, proporcionando una base sólida para futuras evaluaciones y adopciones de esta tecnología en entornos de investigación y clínicos.

3.2.6 Definición de Tiempo de Ejecución de Proyectos y Comparación de Metodologías

Se evaluó el tiempo de ejecución de los proyectos en la metodología experimental, midiendo el tiempo exacto en TU® mediante el contador de tiempo de ejecución instalado en el área de revisión de la aplicación. Para este dato se limitó el análisis a la descripción de los resultados de ejecución de los proyectos TU®, no pudiendo realizar una comparación directa con los tiempos de ejecución de los URT, al no ser un dato registrado en este tipo de estudio, ni tampoco existir referencias publicadas al respecto.

3.3. DETALLES DE LOS MÉTODOS E-II: COMPRENSIÓN Y UTILIDAD DE THE UMBRELLA COLLABORATION®

3.3.1 Diseño del E-II: comprensión y utilidad de The Umbrella Collaboration®

Para evaluar la comprensión y utilidad de la metodología TU® por parte de profesionales sanitarios se realizó un estudio transversal descriptivo, E-II, mediante una encuesta como instrumento de recolección de datos, diseñada para evaluar la percepción de los participantes

profesionales sanitarios de servicios de geriatría de hospitales universitarios y acreditados en formación MIR de la Comunidad de Madrid sobre la metodología experimental frente a la metodología tradicional. Las preguntas compararon directamente ambos métodos, utilizando una escala Likert para cuantificar los resultados. Esta escala permite una evaluación gradiente de las respuestas, facilitando así una distinción clara y matizada de las opiniones de los encuestados respecto a diversos aspectos de la metodología TU®. La encuesta fue diseñada utilizando Google Surveys por su accesibilidad y eficiencia en la recolección y análisis de datos.

En el desarrollo teórico de la metodología empleada para la creación de la encuesta de valoración, se integraron dos enfoques fundamentales: el modelo SERVQUAL (153) y los procedimientos estándares en la realización de encuestas científicas del Centro de Investigaciones Sociológicas (154) y otros autores (155).

En la fase inicial de desarrollo del cuestionario para evaluar la metodología experimental TU® frente a los URT, se recurrió al modelo SERVQUAL como marco teórico de referencia para incorporar aspectos esenciales de calidad de servicio percibida por los potenciales usuarios futuros de la metodología experimental. Aunque SERVQUAL se diseñó originariamente para valorar servicios en el ámbito comercial, sus dimensiones fundamentales ofrecen una estructura útil para conceptualizar aspectos relevantes en la evaluación de cualquier tipo de servicio, incluido el ámbito de la presentación de resultados científicos. El modelo SERVQUAL, identifica cinco dimensiones clave de la calidad del servicio, que fueron utilizados de forma adaptativa para desarrollar la estructura y el contenido del cuestionario:

1. Fiabilidad: inspiró preguntas relacionadas con la consistencia y precisión de la metodología para presentar resultados fiables.
2. Capacidad de Respuesta: orientó la inclusión de ítems que evalúan la rapidez y eficacia de la metodología para facilitar el análisis e interpretación de datos.
3. Seguridad: fue fundamental para formular preguntas sobre la confianza y credibilidad que los usuarios perciben en la metodología.
4. Empatía: influenció la creación de preguntas sobre la experiencia del análisis y la satisfacción general que ofrece la metodología a sus usuarios.
5. Elementos Tangibles: aunque menos relevante en un contexto digital, se consideró para evaluar la apariencia visual y la facilidad de navegación de las interfaces utilizadas.

Siguiendo las directrices establecidas por el Centro de Investigaciones Sociológicas (154) se estructuró el diseño y ejecución de la encuesta de la siguiente manera:

1. Definición del Tema y Objetivo: la encuesta tiene como fin principal evaluar la facilidad de interpretación y empleabilidad de los resultados presentados por TU® frente a los URT, destacando la importancia de la visualización gráfica para la comprensión del usuario final.

2. Preparación y Documentación Introductoria: se proporcionaron materiales, incluyendo explicaciones sobre el propósito del estudio, garantías de anonimato y confidencialidad, conferencia presentación sobre la síntesis de la evidencia específicamente orientada a la síntesis terciaria y a la metodología TU®, documentación explicativa de los métodos tradicionales de exposición de resultados y acceso directo al interfaz interactivo de TU®.
3. Selección de la Muestra: se identificó un universo compuesto por profesionales sanitarios de servicios de geriatría de hospitales universitarios de la Comunidad de Madrid.
4. Redacción del Cuestionario: se incluyeron preguntas cerradas puntuadas en una escala Likert para permitir un análisis cuantitativo, además de recoger datos sociodemográficos relevantes. La escala Likert para la valoración de las percepciones de los usuarios es ampliamente reconocida por su eficacia en la medición de actitudes y opiniones en estudios sociales y de comportamiento. La escala utilizada en nuestra encuesta consta de cinco puntos, donde el valor 1 indica "Totalmente en desacuerdo" y el valor 5 señala "Totalmente de acuerdo". Esta escala permite una evaluación gradiente de las respuestas, facilitando así una distinción clara y matizada de las opiniones de los encuestados respecto a diversos aspectos de las metodologías evaluadas.
5. Formato y Distribución de la Encuesta: se optó por una modalidad en línea utilizando Google Survey por su eficiencia en costos, tiempo y facilidad de difusión.

La encuesta se redactó empleando una IA generativa, ChatGPT 4.o, OPENAI (156), con las siguientes instrucciones, transcripción directa:

“estamos validando una metodología experimental (TU) basada en programación informática e inteligencia artificial para realizar síntesis terciaria de evidencia de forma semiautomática comparándolo con la metodología tradicional de umbrella reviews (URT). Considerando las dimensiones de calidad del modelo SERVQUAL y los pasos habituales para la realización de una encuesta científica tenemos que diseñar una encuesta que mida las diferencias que tienen las metodologías es la forma de presentar los resultados; mientras que los URT los suelen presentar en tablas donde exponen los outcomes con distintas métricas de magnitudes de efecto con sus intervalos de confianza y otros valores estadístico que definen el valor del efecto de la intervención o exposición a riesgo; TU presenta los resultados en una interfaz interactiva, representando los outcomes sobre un bubble plot dispuestos en función del número de revisiones que aportan para ese outcome (eje Y) y en el eje X según el grado de certeza, luego una figura (triangulo si es significativo / circulo si el efecto no es significativo), los triángulos orientados en función si son favor o no de la intervención, o si aumentan o disminuyen el riesgo; luego cada figura tiene un color en función del valor de la magnitud de efecto (trivial, small, moderase, large y very large). La interfaz es interactiva, y al pulsar sobre la figura se despliega una ventana sobre datos de interés sobre la figura (N.º de revisiones que participan en ese outcome, N.º de estudio primarios, valor de la magnitud

en la escala $r(TU)$, y nivel de certeza según el valor otorgado por el programa); a su vez se despliegan otras dos ventanas: una es un forest plot donde se representan los valores normalizados a la escala $r(TU)$ con sus intervalos de confianza de todas las revisiones que analizaron el mismo outcome y el valor final agregado de todas ellas; y luego un bubble plot secundario donde se representa cada una de las revisiones que participan en ese outcome en función del número de estudios primarios, nº de participantes, significación según la forma de la figura, color según el tamaño del efecto, y sobre el eje de las X el valor de sentimiento asignado. A la vez expone el listado de revisiones que participan de ese outcome, pudiendo acceder al abstracto completo según viene publicado en Medline. De alguna forma TU pretende ser más fácil de entender al ser gráfica y visual, y no necesita mucho conocimiento de estadística para interpretar los resultados y por ello se supone que debe ser útil en la práctica. Mientras que las URT presentan los datos de una forma completa y estandarizada, precisa de cierto nivel de conocimiento de estadística para comprender los resultados. Tengo que hacer una encuesta a médicos sobre la percepción de la utilidad, facilidad de uso, facilidad de comprensión, aplicabilidad en la práctica diaria, y posiblemente más aspectos. Las preguntas deben ser 10 y las respuestas a cada una de ellas en una escala Likert de 1 a 5.”

La encuesta propuesta fue:

1. ¿Considera que la interfaz interactiva de la metodología TU® facilita la comprensión de los resultados en comparación con las tablas estáticas de los URT?
2. ¿Cree que la visualización de resultados mediante *bubble plots* en TU® ayuda a identificar rápidamente los *outcomes* más relevantes?
3. ¿La representación gráfica de los efectos (triángulos y círculos para significancia) en TU® mejora su capacidad para evaluar la relevancia clínica de los resultados?
4. ¿Encuentra que el acceso a datos detallados al hacer clic en las figuras en la interfaz de TU® mejora su experiencia de análisis de la evidencia?
5. ¿Piensa que la metodología TU® permite una interpretación más rápida de los datos en comparación con los URT?
6. ¿La información adicional proporcionada por los *forest plots* y *bubble plots* secundarios en TU® es útil para su práctica clínica?
7. ¿La facilidad de uso de la interfaz de TU® facilita una mayor exploración de los datos en comparación con los métodos tradicionales?
8. ¿Considera que la metodología TU® es aplicable en su práctica diaria para la toma de decisiones basadas en evidencia?
9. La metodología TU® requiere menos conocimientos estadísticos para interpretar los resultados en comparación con los URT?
10. En general, ¿qué tan satisfecho está con la metodología TU® como herramienta para la síntesis terciaria de evidencia?

3.3.2 Variables del E-II

Las variables principales del E-II fueron diseñadas para evaluar la percepción de profesionales sanitarios respecto a varios aspectos de la metodología experimental TU® comparada con la metodología tradicional. Estas variables derivaban directamente de las preguntas incluidas en la encuesta diseñada exprofeso. Las variables principales a estudiar incluyeron:

- 1. Facilidad de Comprensión: evalúa si la interfaz interactiva de la metodología TU® facilita la comprensión de los resultados en comparación con las tablas estáticas de los URT. Código P. 7.
- 2. Utilidad de la representación gráfica en diagramas de burbujas (*bubble plots*): mide la percepción sobre si la visualización de resultados mediante diagramas de burbujas en TU® ayuda a identificar rápidamente los Rdl. Código P. 8.
- 3. Evaluación de la Relevancia Clínica: considera si la representación gráfica de los efectos en TU® (triángulos y círculos para significancia) mejora la capacidad del usuario para evaluar la relevancia clínica de los resultados. Código P. 9.
- 4. Mejora en la Experiencia de Análisis: determina si el acceso a datos detallados al interactuar con las figuras en la interfaz de TU® mejora la experiencia de análisis de la evidencia. Código P. 10.
- 5. Rapidez de Interpretación: examina si la metodología TU® permite una interpretación más rápida de los datos en comparación con los URT. Código P. 11.
- 6. Aplicabilidad en la Práctica Diaria: evalúa si la metodología TU® es aplicable en la práctica diaria para la toma de decisiones basadas en evidencia. Código P. 12.
- 7. Facilidad de Uso: analiza si la facilidad de uso de la interfaz de TU® facilita una mayor exploración de los datos en comparación con los métodos tradicionales. Código P. 13.
- 8. Interés de la exposición histórica de los resultados: evalúa si la exposición histórica de los resultados que proporciona TU® es útil. Código P. 14.
- 9. Requerimientos de Conocimientos Estadísticos: mide si la metodología TU® requiere menos conocimientos estadísticos para interpretar los resultados en comparación con los URT. Código P. 15.
- 10. Satisfacción General: determina el nivel de satisfacción general con la metodología TU® como herramienta para la síntesis terciaria de evidencia. Código P. 16.

Se adjunta la **Tabla 2**, donde se detalla los códigos, desarrollo y descripción de cada uno de los 10 ítems de la encuesta de comprensión y utilidad de la metodología TU®.

Tabla 2. Códigos, desarrollo y descripción de los ítems de la encuesta online sobre utilidad de The Umbrella Collaboration®.

Código	Desarrollo del ítem en encuesta	Descripción
P. 7	¿Considera que la interfaz interactiva de la metodología "The Umbrella" (TU®) facilita la comprensión de los resultados en comparación con las tablas estáticas de los Umbrella Reviews Tradicionales (URT)?	Facilita comprensión
P. 8	¿Cree que la visualización de resultados mediante bubble plots en "The Umbrella" (TU®) ayuda a identificar rápidamente los resultados de interés más relevantes?	Bubble plots
P. 9	¿La representación gráfica de los resultados de interés (ilustrada por figuras, colores y tamaños) en "The Umbrella" (TU®) mejora su capacidad para evaluar la relevancia clínica de los resultados?	Representación gráfica
P. 10	¿Encuentra que el acceso a datos detallados al hacer clic en las figuras de los resultados de interés en la interfaz de "The Umbrella" (TU®) mejora su experiencia de análisis de la evidencia?	Acceso a datos
P. 11	¿Piensa que la metodología "The Umbrella" (TU®) permite una interpretación más rápida de los datos en comparación con los Umbrella Reviews Tradicionales (URT)?	Interpretación rápida
P. 12	¿La información proporcionada por "The Umbrella" (TU®) es útil en su ámbito de trabajo (clínica, investigadora o docente) para la toma de decisiones basadas en la evidencia?	Útil toma decisiones
P. 13	¿La facilidad de uso de la interfaz de "The Umbrella" (TU®) facilita una mayor exploración de los datos en comparación con los métodos tradicionales?	Facilidad explorar datos
P. 14	¿Considera útil la evolución histórica de la evidencia que proporciona la metodología "The Umbrella" (TU®)?	Evolución Histórica
P. 15	¿La metodología "The Umbrella" (TU®) requiere menos conocimientos estadísticos para interpretar los resultados en comparación con los Umbrella Reviews Tradicionales (URT)?	Menos conocimientos
P. 16	En general, ¿está satisfecho con la metodología "The Umbrella" (TU®) como herramienta para la síntesis terciaria de evidencia?	Satisfacción

En la encuesta se segmentaron datos de filiación de los encuestados para reflejar la diversidad y las características específicas de los participantes del estudio. Las variables recopiladas fueron esenciales para analizar la percepción de la metodología experimental TU® comparada con la metodología tradicional representadas por las URT. A continuación, se detalla cada una de estas variables:

1. Área Primordial de Trabajo: se clasificó a los participantes según su campo profesional principal para evaluar diferencias en la percepción y aplicación de la metodología TU®. Las categorías incluyen Medicina o Enfermería.
2. Especialidad Médica: para los participantes de la categoría profesional “médico”, se recogió información sobre su especialidad médica según el listado de especialidades de la Unión Europea de Médicos Especialistas (157). Esta variable permitió analizar si existían diferencias en la percepción de la metodología TU® entre las diversas especialidades médicas.
3. Género: se determinó el género de los participantes con las opciones de hombre, mujer y no determinado, permitiendo estudios comparativos de percepción entre diferentes grupos de género.
4. Edad: la variable de edad, categorizada en cuatro categorías, se incluyó como un factor demográfico importante que podía influir en la aceptación y uso de nuevas tecnologías y metodologías en el campo médico. Las cuatro categorías de edad fueron: < 33 años, de 34 a 44 años, de 45 a 54 años y de 55 a 65 años.
5. País de Residencia: esta variable proporcionó contexto geográfico y cultural, crucial para entender la variabilidad en la adopción de nuevas prácticas médicas y tecnológicas en diferentes regiones.
6. Hospital de Trabajo: se identificó el hospital universitario específico dentro de la Comunidad de Madrid donde trabaja el participante, de una lista predeterminada de cinco hospitales. Esta información es útil para correlacionar la infraestructura institucional y los recursos disponibles con la percepción y uso de la metodología TU®.

Estas variables principales y desagregados fueron analizadas para determinar el impacto en la percepción de profesionales sanitarios en la potencial adopción de la metodología TU® en su práctica clínica diaria. Además, el análisis proporcionó una comprensión más profunda de cómo diferentes factores demográficos y profesionales pueden influir en la percepción de la utilidad de la metodología en estudio.

3.4 PROCEDIMIENTO EXPERIMENTAL

3.4.1 Procedimiento Experimental E-I.

3.4.1.1 Selección y análisis de URT en el área de geriatría:

El estudio comenzó con la identificación de URT específicas de geriatría, utilizando como base de datos Medline a través de PubMed. La búsqueda se realizó el 5 de marzo de 2023 con los términos “*umbrella*” AND “*geriatric*”. El enlace de la búsqueda fue: <https://pubmed.ncbi.nlm.nih.gov/?term=umbrella+geriatric&sort=date>. Se procedió a una primera selección de las referencias obtenidas, con los criterios de inclusión (URT del área de geriatría) y de exclusión (no ser una URT, no ser del área de la geriatría, ser un protocolo, no ser sobre intervenciones o riesgos). Del listado final de referencias incluidas se seleccionó de forma aleatoria los URT incluidos en el análisis final. Se procedió a una evaluación exhaustiva de todas las URT incluidas en el análisis final, accediendo a la publicación completa. Los datos extraídos abarcaron:

1. Referencia Completa: detalle bibliográfico de cada URT.
2. Pregunta de Investigación: clasificada en efectividad de intervenciones o exposiciones a riesgos.
3. Se analizó la estrategia de búsqueda de cada URT, incluyendo si los URT se limitaban a resultados de interés limitados y predefinidos o abarcaban todos los Rdl descritos en las RS/MA incluidos en las URT.
4. RS/MA incluidos en las URTs: listado y referencias completas junto con PMID (PubMed Unique Identifier) de las RS/MA utilizadas para derivar resultados en cada URT.
5. Resultados de Interés (Rdl): identificación del Rdl, TE, dirección del efecto (favorable/desfavorable a la intervención, aumento/disminución de riesgo), significación estadística y métodos de validación de calidad empleados (p. ej., validación bioestadística de Papatheodorou, calidad metodológica de AMSTAR-2 o certeza de la evidencia con GRADE. Identificación mediante el PMID de las RS/MA de las cuales se obtienen los resultados sobre la identificación de Rdl y sus propiedades.

3.4.1.2 Configuración de Proyectos en Paralelo con la metodología experimental The Umbrella Collaboration®

Se realizaron proyectos con la metodología TU® con la misma pregunta de investigación que las URT seleccionadas. De forma esquemática se explica el proceso seguido para la realización de un proyecto dentro de TU®.

Los proyectos en TU® se estructuran alrededor de preguntas de investigación específicas, replicando las preguntas de investigación de las URT seleccionadas para el análisis. Utilizando el software, The Umbrella®, desarrollado *exprofeso*, que incorpora modelos avanzados de web *scraping*, procesamiento de lenguaje natural, estructuración de información y análisis de sentimiento, que simplifica y automatiza la búsqueda de literatura relevante, exclusivamente RS/MA, y agrega y presenta en un informe gráfico y visual la información relevante a Rdl con una intervención humana mínima. Se describen los procedimientos fundamentales de la metodología experimental TU® a continuación.

3.4.1.3 Proceso de Búsqueda y Selección por The Umbrella®

La búsqueda se realiza únicamente en la base de datos Medline (vía PubMed), partiendo de términos específicos como, por ejemplo: "Artrosis de rodilla y eventos adversos de salud". Con el apoyo de LLM, se expande la búsqueda incorporando términos de búsqueda relacionados, permitiendo múltiples consultas básicas vía Pubmed de forma simultánea. TU® automatiza la revisión de *abstracts*, exclusivamente de RS/MA, que constituyen la unidad de estudio, seleccionando aquellos que cumplen con los criterios de inclusión predefinidos y presentando los resultados de análisis inicial sin supervisión humana en dos listados en orden cronológico según la relevancia y cumplimiento de los criterios de inclusión. Los dos listados de RS/MA son en función de si el programa identificó correctamente en el *abstract* de las RS/MA Rdl con métricas de magnitudes de efecto correspondientes quedando estas RS/MA expuestas en un listado primario, en caso de que el programa identifique la RS/MA, pero no localice Rdl con sus métricas de magnitudes de efecto, estas RS/MA irán a un listado secundario. Ambos listados son revisados por el revisor humano responsable del proyecto TU®.

3.4.1.4 Verificación de Datos por el Revisor Humano

El revisor humano responsable del proyecto revisa cada RS/MA seleccionada por TU® para confirmar su pertinencia con la pregunta de investigación y su inclusión final en los resultados de síntesis del proyecto. Esta revisión se realiza primero por título y, en caso de dudas, se extiende al *abstract* completo. A continuación, para terminar el proceso, el revisor humano corrige, completa y verifica la información extraída de los RS/MA incluidos por TU®, asegurando la precisión de:

- Número de estudios primarios y participantes en cada RS/MA.
- Identificación y clasificación correcta de los Rdl.
- Métricas de magnitudes de efecto y sus respectivos valores e intervalos de confianza.
- Evaluación del sentimiento del análisis.
- Para concluir el revisor humano tiene que realizar dos operaciones de confirmación: en primer lugar, determinar el sentido del efecto del Rdl, y luego confirmar que ese resultado de interés ha sido revisado.

3.4.1.5 Normalización y Representación de Datos

Una vez revisados, verificados y aceptados por el revisor humano todos los Rdl, encontrados en cada RS/MA, TU® incorpora cada Rdl individual de los RS/MA al pool de estudios incluidos definitivamente en los resultados finales del proyecto, y normaliza con fórmulas propias los datos para la integración de los resultados del proyecto para la visualización de los resultados en el interfaz online del programa.

3.4.1.6 Integración y Análisis Continuo (nuevas búsquedas automáticas cada 24h)

Los proyectos en TU® se mantienen permanentemente activos y se actualizan constantemente a diario de forma automática o a demanda del revisor humano responsable del proyecto, agregando nuevas RS/MA según se vayan publicando y coincidan con los criterios de búsqueda del proyecto. Para este estudio se limitó el tiempo de búsqueda a una fecha, sin realizar actualizaciones de los proyectos que hubiese permitido obtener más datos.

3.4.2 Procedimiento Experimental E-II.

En el E-II, diseñada para valorar la comprensión y utilidad de la presentación de los resultados de TU® frente a URT, se llevó a cabo una evaluación mediante una encuesta dirigida a profesionales de la salud de 5 servicios de geriatría de hospitales universitarios acreditados para la formación MIR de la Comunidad de Madrid (Hospital Universitario Puerta de Hierro - Majadahonda, Hospital Universitario Clínico San Carlos, Hospital Universitario La Paz, Hospital Universitario Ramón y Cajal y Hospital Universitario Cruz Roja). Se invitó a todos los servicios hospitalarios de Geriatría de la CM con acreditación docente MIR en Geriatría que pudieran participar dentro del periodo del estudio del 1 de enero al 1 de julio de 2024. La encuesta se realizó en un entorno específico y controlado inmediatamente después de una presentación detallada del proyecto a los participantes. La presentación del proyecto incluyó una presentación teórica de la síntesis terciaria de la evidencia, junto con una explicación detallada de la metodología y aplicación The Umbrella Collaboration®.

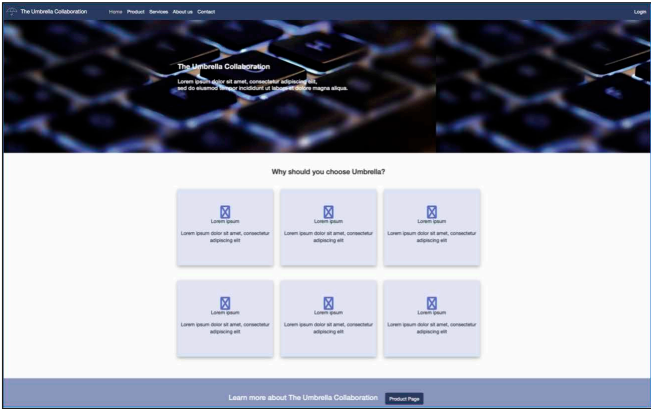
Este método permitió a los encuestados aclarar dudas sobre las metodologías y la encuesta, y recoger respuestas de manera inmediata, lo que contribuyó a una alta tasa de respuesta y minimizó el riesgo de olvido o sesgos asociados con el retraso en la respuesta. Durante la exposición se analizó la forma de presentación de resultados de ambas metodologías, con acceso a la plataforma interactiva de TU® por parte de los encuestados. Enlace de acceso a la página web "The Umbrella Collaboration®": (<https://theumbrellacollaboration.org/>). Enlace QR de acceso a la encuesta: (<https://forms.gle/vi4QvwWAZSUDokNPA>) (**Figura 1**).

Figura 1. Enlace QR de acceso a la encuesta.



La página de The Umbrella Collaboration® es una página interactiva (<https://theumbrellacollaboration.org/>) de acceso libre que permite acceder a los resultados de todos los proyectos realizados dentro de la metodología The Umbrella Collaboration® (**Figura 2**).

Figura 2. Página inicio The Umbrella Collaboration®.



Para que un usuario pueda acceder a la información sobre los distintos proyectos de síntesis terciaria realizados con la metodología de TU®, debe seguir una serie de pasos. Primero debe registrarse y una vez dentro de la aplicación, debe visitar la página de especialidades médicas, (<https://theumbrellacollaboration.org/#/select-speciality>). Esta página clasifica los proyectos según las especialidades médicas basadas en la clasificación de la Unión Europea de Especialidades Médicas, con algunas modificaciones (**Figura 3**).

Figura 3. Listado de especialidades página web The Umbrella Collaboration®.

1 0 - Tests	2 Acupuncture	3 Allergology
4 Anaesthesiology	5 Cardiology	6 Cardiothoracic Surgery
7 Child and Adolescent Psychiatry	8 Clinical Neurophysiology	9 Dermatology and Venereology
10 Emergency Medicine	11 Endocrinology	12 Gastroenterology
13 Geriatrics	14 Gynaecology and Obstetrics	15 Hematology
16 Infectious Diseases	17 Internal Medicine	18 Laboratory Medicine / Medical Biopathology
19 Medical Genetics	20 Medical Microbiology	21 Medical Oncology
22 Nephrology	23 Neurology	24 Neurosurgery
25 Nuclear Medicine	26 Occupational Medicine	27 Ophthalmology
28 Oro-Maxillo-Facial Surgery	29 Orthopaedics & Traumatology	30 Otorhinolaryngology
31 Paediatric Surgery	32 Paediatrics	33 Pathology

Durante la presentación los encuestados podían buscar y seleccionar dentro del listado de especialidades el área denominada "Test", que alberga los proyectos desarrollados durante la realización de esta tesis doctoral. Al seleccionar "Test", se despliega una lista de proyectos. Cada proyecto incluye los resultados del proceso de síntesis de evidencia utilizando la metodología de TU®. El usuario puede acceder a información detallada sobre cada proyecto (**Figura 4**).

Figura 4. Listado de proyectos dentro del área Test The Umbrella Collaboration®.

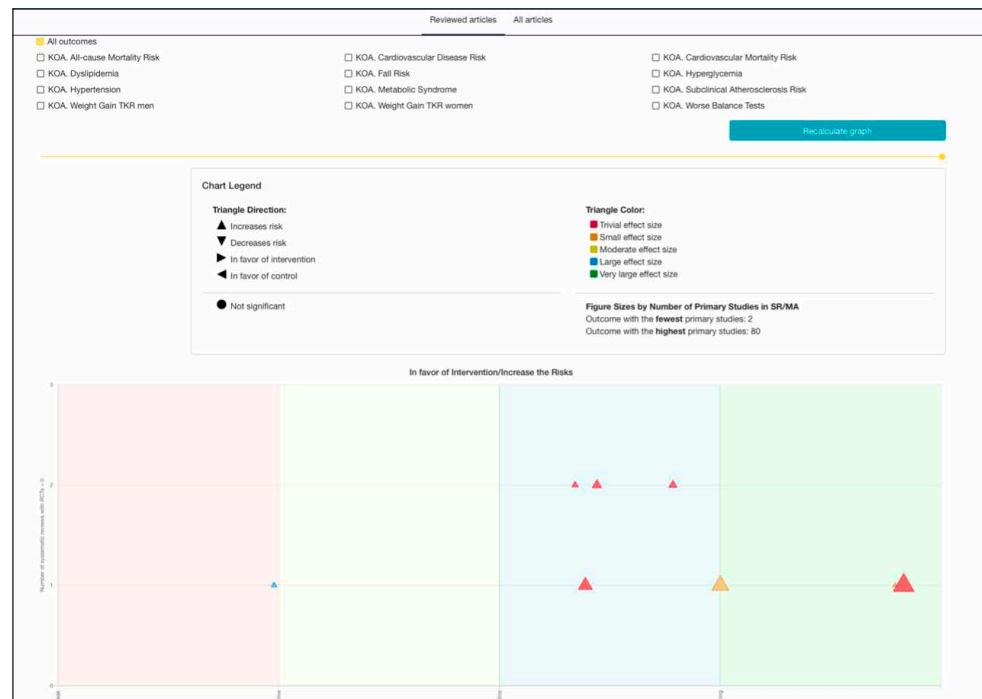
1. interventions reduce adverse outcmes older adults Emergency Department	2. Physical fitness training older adults	3. elderly life expectancy
4. Bone density Exercise Adult	5. Vitamin D and Falls	6. urinary incontinence elderly
7. Adverse outcomes older patients emergency department interventions	8. older adults, gait, balance, and functional mobility assessments, and falls prediction	9. SARS-CoV-2 infection, severe COVID-19 illness COVID-19 mortality pre-existing mental disorders
10. Effect of digoxin mortality atrial fibrillation heart failure	11. Modifiable predictive factors and all-cause mortality in the non-hospitalized elderly population	12. Operative Risk Factors Associated with Post-Operative Cognitive Dysfunction
13. Knee osteoarthritis and adverse health outcomes	14. Zinc Intakes and Health Outcomes	15. Predicting falls in older adults
16. Effect of low-dose aspirin on health outcomes	17. Exercise in Older People With Sarcopenia	18. Physical Activity and Exercise in Mild Cognitive Impairment and Dementia
19. Knee Osteoarthritis and Adverse Health Outcomes 2	20. Predicting Falls in Older Adults	21. Menopausal hormone therapy and women health
22. The Dietary Inflammatory Index and Human Health	23. Eye disease and health outcomes	24. Eye Disease and Health Outcomes 2

Para seleccionar un proyecto de interés, el usuario debe visitar <https://theumbrellacollaboration.org/#/select-project>. Aquí se encuentra una lista de todos los proyectos disponibles clasificados bajo el área de "Test". Cada proyecto está listado con su título. El usuario puede explorar los títulos para identificar los proyectos de su interés, hacer clic en el título del proyecto de interés para acceder a más detalles y examinar la información completa que incluye de los resultados del proceso de síntesis terciaria de evidencia. Este proceso permite al usuario obtener información de cada proyecto y sus hallazgos.

Como ejemplo, al acceder al proyecto "Condiciones de salud relacionadas con la Artrosis de Rodilla" en TU®, el usuario encontrará en primer lugar una pantalla denominada “Bubble Plot Principal” (**Figura 5**). Esta pantalla se organiza de la siguiente manera: en la parte superior, se encuentra una lista de todos los Rdl encontrados en el proyecto. A continuación, se presenta una leyenda de la gráfica que detalla el significado de las figuras y colores en la gráfica. Las figuras incluyen triángulos para resultados significativos, con la orientación indicando el sentido del efecto, y círculos para resultados no significativos. Los colores

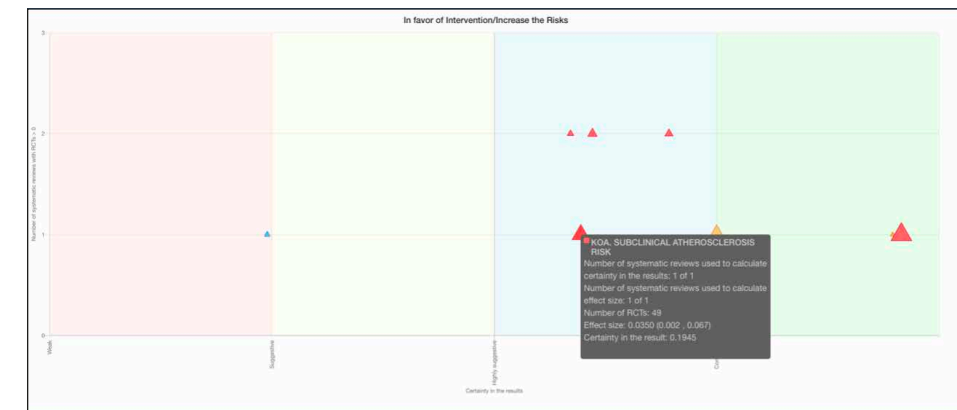
representan diferentes tamaños de efecto: rojo para trivial, naranja para efecto pequeño, amarillo para efecto moderado, azul para efecto grande y verde para efecto muy grande. El tamaño de la figura está relacionado con el número de estudios primarios incluidos en las revisiones sistemáticas que aportan datos para esos Rdl. El “*Bubble Plot* Principal” muestra de manera visual la distribución de los Rdl hallados en el proyecto, estructurada en cuatro áreas de distintos colores que representan la certeza en los resultados: *Weak* (Débil), *Suggestive* (Sugestivo), *Highly Suggestive* (Altamente sugestivo) y *Convincing* (Convincente). El eje de las ordenadas indica el número de revisiones sistemáticas que han participado en cada Rdl. Este esquema proporciona una comprensión visual clara de la evidencia sintetizada, facilitando la evaluación de la certeza y relevancia de los resultados del proyecto.

Figura 5. Bubble Plot principal Knee osteoarthritis and health outcomes.



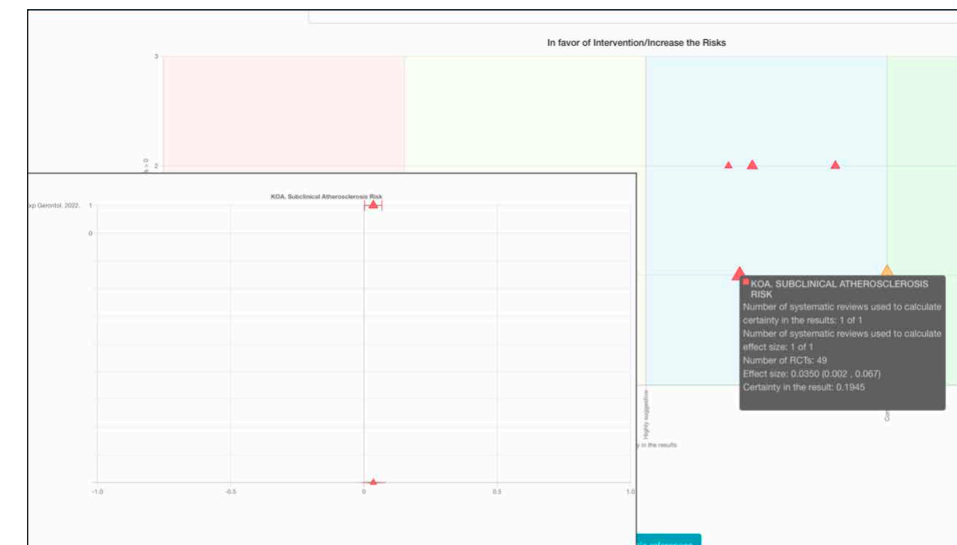
Al posicionar el cursor sobre las figuras en el *Bubble Plot* Principal, se despliega una pestaña con información detallada sobre ese Rdl. La información incluye el nombre identificativo del Rdl, que es un botón interactivo que lleva a pantallas secundarias como el “*Forest Plot*” y el “*Bubble Plot* Secundario”, el número de revisiones sistemáticas usadas para calcular la certeza y el tamaño del efecto, el número de estudios primarios incluidos entre todas las revisiones sistemáticas que se incluyen en ese Rdl, el tamaño del efecto R_{TU} , con intervalos de confianza y el valor de certeza según el sentimiento de análisis en el resultado. Este desglose permite al usuario comprender la robustez y significancia de los resultados presentados (**Figura 6**).

Figura 6. Despliegue ventana información de resultado de interés KOA. Subclinical atherosclerosis risk en Knee osteoarthritis and health outcomes.



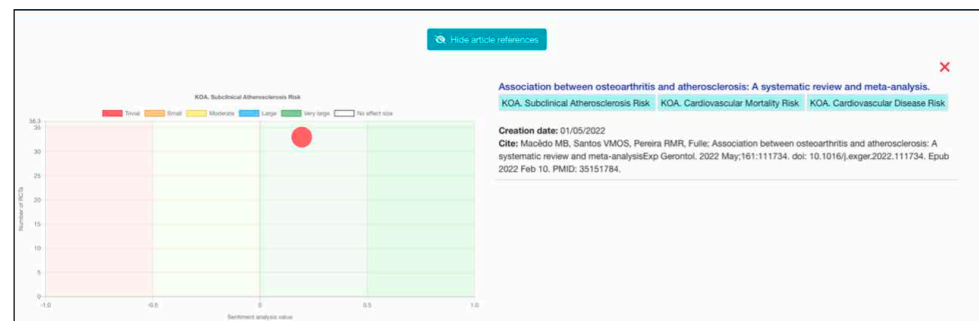
Al hacer clic sobre la figura en el *Bubble Plot* Principal, se abren tres pantallas secundarias: el “*Forest Plot*”, el “*Bubble Plot* Secundario” y el “Listado de Revisiones Sistemáticas”. El “*Forest Plot*” presenta el TE R_{TU} de los resultados de cada RS/MA incluido para ese Rdl de manera similar a los *forest plots* utilizados en metaanálisis tradicionales. El eje horizontal muestra la métrica R_{TU} , con un rango de -1 a +1 y un punto nulo en 0. Si el intervalo de confianza cruza 0, el resultado se considera no significativo. Las revisiones sistemáticas se presentan en orden cronológico y cada figura de TE es un botón interactivo que lleva al *abstract* de la revisión según figura en PubMed. Cada Rdl individual de cada revisión sistemática incluida en el Rdl de TU®, figura en el “*Forest Plot*” con una figura que representa su significación (triángulo para significativos / círculos para no significativos), los triángulos orientados según el sentido del efecto, diferentes colores en función del tamaño de efecto y una barra con los extremos del intervalo de confianza del valor del tamaño de efecto. Las revisiones sistemáticas vienen ordenadas de manera cronológica, apareciendo en la base de la figura la representación del Rdl final obtenido por TU® de la agregación de todos los Rdl individuales de cada RS/MA incluida para ese Rdl (**Figura 7**).

Figura 7. Forest Plot de resultado de interés KOA. Subclinical atherosclerosis risk en Knee osteoarthritis and health outcomes.



El “*Bubble Plot* Secundario” muestra, en su gráfico todos los Rdl individuales de las RS/MA que participan en el Rdl resultante del análisis de TU®. En el eje de abscisas viene representado el valor de certeza del análisis de sentimiento otorgado por TU®, de -1 a +1, y el eje de ordenadas el número de estudios primarios incluidos por cada RS/MA para el cálculo del Rdl individual. Los Rdl individuales se disponen según su certeza (de izquierda a derecha), número de estudios primarios, significación estadística (triángulo si es significativo, círculo si no), sentido del efecto (orientación del triángulo) y tamaño del efecto (color y tamaño de la figura) (**Figura 8**).

Figura 8. Bubble plot secundario y listado de RS/MA de resultado de interés KOA. Subclinical atherosclerosis risk in Knee osteoarthritis and health outcomes.



El “Listado de Revisiones Sistemáticas” que participan en el Rdl que estamos explorando, incluye datos básicos de la revisión sistemática que han sido incluidos en ese Rdl resultante del análisis, como el título, listado de todos los Rdl incluidos en esa revisión sistemática y la fecha de publicación con la cita completa. El título es un botón interactivo que lleva al *abstract* en PubMed. Este esquema facilita una comprensión detallada de los resultados y su relevancia estadística, permitiendo una exploración interactiva y profunda de la evidencia sintetizada.

3.5 RECOPIACIÓN DE DATOS

3.5.1 Recopilación de datos E-I

Una vez completadas las fases de lectura detallada de las URT y la implementación de proyectos paralelos en TU®, se procedió a la recopilación de los datos de interés, que fueron almacenados en una base de datos creada con Microsoft Excel. Estos datos, que incluyen datos de los resultados de ambas metodologías, como la identificación por PMID de las RS/MA incluidas en la síntesis asociados a los Rdl, la identificación del Rdl, TE, direcciones del efecto, significancias estadísticas, evaluaciones de calidad y tiempos de ejecución de los todos los proyectos con TU®. Los datos fueron extraídos manualmente de los URT, y en TU® fueron obtenidos directamente de la interfaz de la aplicación online TU®.

Cada dato relevante fue tabulado en una hoja de cálculo de Microsoft Excel. Este paso, realizado personalmente por el autor del estudio, facilitó la organización sistemática de

la información necesaria para realizar comparaciones analíticas entre las metodologías evaluadas. La estructuración de los datos en Microsoft Excel, y el procedimiento de realizarse por el propio autor, no solo permitió un acceso y manejo eficiente durante el análisis estadístico, sino que también aseguró la integridad y la precisión de los datos al minimizar posibles errores.

3.5.2 Recopilación de datos E-II

La recolección de datos para el estudio de la comprensión y utilidad de la metodología y la aplicación TU® se realizó mediante una encuesta online distribuida a través de Google Survey. La encuesta fue aplicada en cinco hospitales universitarios acreditados en la formación MIR de Geriátrica en la Comunidad de Madrid. Las presentaciones del proyecto, que incluían la explicación y demostración de la metodología, se realizaron en diferentes fechas entre mayo y julio de 2024 en cada uno de los hospitales participantes: Hospital Universitario Puerta de Hierro-Majadahonda, Hospital Universitario La Paz, Hospital Universitario Clínico San Carlos, Hospital Universitario Ramón y Cajal, y Hospital Universitario Central de la Cruz Roja "San José y Santa Adela".

La participación en la encuesta fue voluntaria y anónima, dirigida a los profesionales sanitarios presentes en las presentaciones, tanto personal del propio servicio como profesionales visitantes. Cada participante recibió un enlace a la encuesta de Google Survey durante la presentación del proyecto y la encuesta, permitiéndoles completar el cuestionario en un momento conveniente.

Una vez que se dio por concluida la posibilidad de participar en la encuesta con fecha de 10 de julio de 2024, todos los datos recogidos fueron transferidos a la base de datos creada con Microsoft Excel para su posterior organización y estructuración. Esta base de datos facilitó la clasificación de los participantes según hospital, especialidad, género y grupos de edad. La sistematización de los datos en Microsoft Excel permitió un manejo eficiente y seguro de la información recopilada, asegurando la integridad y confidencialidad de los datos de los encuestados.

El uso de Google Survey proporcionó una plataforma accesible y fácil de usar para los participantes, buscaba a una alta tasa de respuesta y una recolección rápida y eficiente de los datos. La combinación de una metodología de encuesta online con la organización en una base de datos de Microsoft Excel garantizó que los datos estuvieran listos para cualquier análisis posterior, manteniendo la calidad y la coherencia de la información recopilada.

3.6 ANÁLISIS ESTADÍSTICO: ANÁLISIS DE DATOS Y MÉTODOS ESTADÍSTICOS

3.6.1 Análisis Estadístico: Análisis de Datos y Métodos Estadísticos E-I

3.6.1.1 Análisis Cualitativo de Resultados de Interés: identificación de Rdl, tamaño de efecto categorizado, la dirección del efecto de Rdl y la significación estadística del efecto de Rdl.

Para el análisis de los valores cualitativos, Rdl identificados, los TE categorizados, la dirección y la significación estadística, se emplearon tablas cruzadas o de contingencia, para el contraste de hipótesis se empleó el test Chi-Cuadrado y como tamaño o magnitud del efecto la V de Cramer. Las tablas de contingencia permitieron analizar la coincidencia entre las categorías de TE (trivial, pequeño, moderado, grande) de ambas metodologías. La comparación cualitativa de concordancia se dividió en tres niveles: concordancia plena, concordancia parcial y discordancia según descrito anteriormente. Para la dirección del efecto se catalogaron en las tablas de contingencia los resultados obtenidos por ambas metodologías en tres categorías: favorable, desfavorable o desconocido para intervenciones, y aumento, disminución o desconocido del riesgo para exposiciones. Para la significación estadística del efecto de los Rdl se catalogaron en las tablas de contingencia los resultados obtenidos por ambas metodologías en tres categorías: significativo, no-significativo y desconocido.

La prueba de chi-cuadrado se empleó para determinar si las diferencias en la distribución de categorías entre TU® y URT eran estadísticamente significativas. Esta prueba es adecuada para comparar frecuencias observadas en celdas de una tabla de contingencia con las frecuencias esperadas bajo la hipótesis nula (que no haya diferencias entre metodologías).

Para medir el tamaño de efecto en las variables cualitativas expresadas en las tablas de contingencia, empleamos el indicador más comúnmente empleado que es el coeficiente V de Cramer. Este coeficiente tiene una escala de varía de 0 a 1, lo que permite una interpretación clara y precisa de la magnitud del efecto. Los valores del coeficiente V de Cramer < 0.1 se consideran tamaños de efecto insignificantes, los valores entre 0.1 y 0.3 se consideran pequeños, los valores entre 0.3 y 0.5 se consideran tamaños de efecto moderados, mientras que los valores > 0.5 se consideran grandes. El tamaño de efecto moderado expresa diferencias que son reales y que se pueden detectar hasta con muestras tan pequeñas como $50 < N < 100$. El tamaño de efecto pequeño es indicativo de que puede haber algo, pero para detectarlas con cierta seguridad se necesitarían muestras de alrededor de $N=500$. El tamaño grande evidencia diferencias que existen con elevado grado de confianza y que se detectan casi a “simple vista” y hasta con muestras muy pequeñas (158,159).

Existen una serie de limitaciones de las tablas de contingencia y la prueba de chi-cuadrado que son relevantes en nuestro estudio, por lo que lo desarrollamos en el plano teórico a continuación (160):

Limitaciones de las Tablas de Contingencia

Escalabilidad: Las tablas de contingencia son las más apropiadas con un número limitado de categorías (por ejemplo, 2x2, 3x3 o 4x4), pero a medida que el número de categorías aumenta, las tablas se vuelven más difíciles de interpretar y manejar. Si las variables tienen muchas categorías o niveles, las tablas pueden volverse muy grandes, lo que dificulta una interpretación clara y precisa de los datos.

Datos agrupados: Las tablas de contingencia no proporcionan detalles sobre los valores individuales. Los datos se agrupan en categorías discretas, lo que significa que se pierde información sobre la variabilidad dentro de cada categoría. Esto puede ocultar diferencias importantes en los datos originales que podrían ser relevantes.

Frecuencias pequeñas o nulas: Cuando algunas celdas en una tabla de contingencia tienen frecuencias muy bajas o incluso valores nulos (ceros), los resultados del análisis pueden no ser fiables. La presencia de frecuencias bajas puede distorsionar el cálculo del valor chi-cuadrado.

Dependencia del tamaño de la muestra: Las tablas de contingencia son sensibles al tamaño de la muestra. Las conclusiones pueden variar drásticamente dependiendo del número total de observaciones. Por ejemplo, con muestras pequeñas, la prueba de chi-cuadrado puede carecer de potencia para detectar una asociación significativa, mientras que, con muestras grandes, incluso asociaciones pequeñas pueden resultar estadísticamente significativas, aunque sean irrelevantes desde un punto de vista práctico.

Limitaciones del Test de Chi-Cuadrado

Tamaño de la muestra: La prueba de chi-cuadrado es muy sensible al tamaño de la muestra. En muestras pequeñas, puede no tener suficiente potencia estadística para detectar una relación entre las variables. Por otro lado, con muestras muy grandes, incluso diferencias muy pequeñas entre las frecuencias observadas y esperadas pueden resultar en un valor de chi-cuadrado significativo, lo que podría llevar a interpretar una asociación como significativa, aunque, en la práctica, sea trivial.

No mide la fuerza de la asociación: Aunque la prueba de chi-cuadrado indica si existe una asociación significativa entre las variables, no proporciona información sobre la magnitud o fuerza de dicha asociación. Para medir la fuerza de la asociación se deben utilizar medidas adicionales, como el V de Cramer o el coeficiente Phi.

Supuesto de independencia: La prueba de chi-cuadrado asume que las observaciones en la tabla de contingencia son independientes entre sí. Si las observaciones no son independientes (por ejemplo, si se tienen datos repetidos o datos emparejados), los resultados del test pueden ser inválidos. En estos casos, se deben utilizar otros métodos estadísticos que manejen la dependencia de los datos, como los modelos de regresión logística para datos categóricos.

Requiere un número mínimo de observaciones por celda: Se recomienda que, para que la prueba de chi-cuadrado sea fiable, todas las celdas de la tabla de contingencia tengan frecuencias esperadas mayores de 5. Si hay celdas con frecuencias observadas o esperadas muy bajas, se deben considerar ajustes (como la corrección de Yates) o recurrir a pruebas alternativas, como la prueba exacta de Fisher, que es más adecuado para tablas con frecuencias pequeñas.

La prueba de chi-cuadrado requiere que las frecuencias esperadas en al menos el 80% de las celdas de la tabla de contingencia sean mayores o iguales a 5, y ninguna celda debe tener una frecuencia esperada de 0. Si esto no se cumple, se pueden tomar dos medidas: aplicar la corrección de Yates (para tablas 2x2), que ajusta el valor del chi-cuadrado, pero puede subestimar la significación, o usar la prueba exacta de Fisher, recomendado cuando las frecuencias son pequeñas, especialmente para tablas 2x2.

En cuanto al tamaño de la muestra, aunque no existe un número absoluto de observaciones necesarias, el tamaño total debe ser lo suficientemente grande para cumplir con el criterio de frecuencias esperadas. Se sugiere un mínimo de 20 observaciones en total, asegurando que las frecuencias esperadas en cada celda sean al menos 5.

Si las tablas tienen celdas vacías o frecuencias bajas (menores de 5), esto puede distorsionar los resultados. Para mitigar este problema, se pueden agrupar categorías adyacentes para reducir el número de celdas, o recurrir al test exacto de Fisher cuando las frecuencias son muy bajas para aplicar el test de chi-cuadrado de manera confiable.

No aplicable a datos continuos: La prueba de chi-cuadrado solo es adecuado para datos categóricos. Si los datos son continuos, primero se deben agrupar en categorías, lo que puede llevar a una pérdida de información y a resultados menos precisos. En estos casos, otros métodos, como la correlación o la regresión, podrían ser más apropiados.

Asume una distribución homogénea: La prueba de chi-cuadrado compara las frecuencias observadas con las frecuencias esperadas bajo la hipótesis nula de distribución homogénea (sin asociación). Si los datos no cumplen con esta suposición, es decir, si las categorías no están distribuidas uniformemente o no son comparables en tamaño, esto puede afectar la validez del resultado.

3.6.1.2 Análisis Cuantitativo de Resultados de Interés: tamaño de efecto y certeza

Se realizó un análisis detallado de los datos cuantitativos obtenidos en el estudio, utilizando diversas herramientas estadísticas para evaluar la relación, predicción y diferencias numéricas entre las dos metodologías. Este análisis incluyó el uso de correlaciones y regresión lineal.

Correlación y regresión lineal

Se emplearon los coeficientes de correlación de Pearson y Spearman para evaluar la relación entre los datos cuantitativos del estudio, tanto en las magnitudes cuantitativas

de los TE como en los valores de certeza reclasificados de GRADE en URT y de análisis de sentimiento de TU® y en la relación entre el tiempo de ejecución de los proyectos y el número de referencias incluidas. Este análisis permitió identificar si las dos metodologías tienden a producir valores similares en términos de magnitud del efecto y certeza. Para evaluar la normalidad de los conjuntos de datos utilizados en los análisis de correlación, se empleó la prueba de Shapiro-Wilk. Esta prueba permitió determinar si los datos seguían una distribución normal ($p \geq 0,05$) o no ($p < 0,05$). En función de estos resultados, se seleccionó el coeficiente de correlación correspondiente: Pearson para datos con distribución normal y Spearman para aquellos que no cumplían este criterio.

Además, se empleó un modelo de regresión lineal para evaluar la capacidad de los valores generados por URT de predecir los valores correspondientes de TU®. Se analizó la intersección, como valor que representa el punto en que la línea de regresión cruza el eje de las ordenadas. Un valor cercano a cero sugiere que ambas metodologías parten de un punto de referencia similar. Se analizó igualmente la pendiente de la línea de regresión, que indica cuánto cambia el TE de URT por cada unidad de cambio en el TE de TU®. Una pendiente de 1 indicará una correspondencia perfecta entre ambas metodologías, mientras que un pendiente diferente de 1 sugerirá una variación sistemática en cómo se relacionan los TE. Finalmente se analizó el coeficiente de determinación R^2 fue utilizado para cuantificar el grado de variabilidad en TU® explicada por los valores de URT, proporcionando un indicador de cuán fuerte es la relación entre ambas metodologías en términos predictivos.

3.6.1.3 Procesamiento y análisis de datos.

Los datos se procesaron y analizaron en IBM SPSS Statistics 26, aplicando un nivel alfa de 0.05 para determinar la significación estadística. Las correlaciones se complementaron con la V de Cramer para estimar la magnitud del efecto en las tablas de contingencia, proporcionando una medida de la fuerza de asociación entre las metodologías evaluadas.

3.6.2 Análisis Estadístico Estudio II (E-II)

Para el análisis estadístico de los resultados de la encuesta aplicada en los servicios de geriatría de hospitales universitarios de la Comunidad de Madrid, se realizaron diversos procedimientos utilizando software estadístico especializado IBM SPSS Statistics.

Inicialmente, se llevó a cabo un análisis descriptivo de la muestra de estudio, reportando la media y la desviación estándar para las variables cuantitativas, y las frecuencias y porcentajes para las variables cualitativas. Este análisis permitió obtener una caracterización detallada de los participantes en términos de profesión, especialidad, género y grupo etario.

Posteriormente, se examinaron las asociaciones entre variables utilizando diferentes pruebas estadísticas, seleccionadas según la naturaleza de las variables involucradas. Para la comparación de variables cualitativas, se emplearon las pruebas de chi-cuadrado con corrección de Yates y el test exacto de Fisher, asegurando así la validez de los resultados aun en casos de frecuencias esperadas bajas.

Para comparar una variable cuantitativa con una variable dicotómica se utilizó la prueba t de Student, bajo el supuesto de normalidad de la variable cuantitativa y homogeneidad de varianzas entre los grupos. En el caso de comparaciones entre una variable cuantitativa y una variable cualitativa no dicotómica, se empleó el análisis de varianza de un factor (ANOVA) tras verificar los supuestos de normalidad y homogeneidad de varianzas, permitiendo identificar diferencias significativas entre múltiples grupos.

El análisis se realizó en varias etapas:

1. Análisis Global: se consideró toda la muestra para identificar los resultados globales.
2. Subanálisis de diferentes categorías dentro de las poblaciones que respondieron a la encuesta: se analizaron diferencias por categorías profesionales (médicos / enfermeros), por género, por grupos de edad, por hospitales, médicos geriatras frente a otras especialidades médicas y un análisis de los médicos geriatras de manera aislada.

Se estableció un nivel de significación del 95% ($p < 0.05$) para todas las pruebas estadísticas, asegurando así la robustez de los hallazgos. Adicionalmente, se calcularon intervalos de confianza del 95% para proporcionar una medida de precisión para las estimaciones obtenidas.

3.7 CONSIDERACIONES ÉTICAS

Este estudio no involucró animales ni el uso de datos personales sensibles. La mayor parte de los datos utilizados provienen de fuentes publicadas y se analizaron siguiendo los estándares éticos establecidos para el análisis de datos secundarios. Adicionalmente, se incluyó una encuesta realizada de forma anónima y voluntaria, garantizando la confidencialidad de las respuestas y obteniendo el consentimiento informado de los participantes. Por estas razones, no se consideró necesario solicitar la aprobación de un comité de ética, ya que el estudio se llevó a cabo cumpliendo con las prácticas éticas estándar para investigaciones de esta naturaleza.

4

RESULTADOS

4.1 RESULTADOS DE LA COMPARACIÓN ENTRE METODOLOGIAS COMPARATIVAS: TU® FRENTE A URT (ESTUDIO-I)

4.1.1 Identificadion de umbrella reviews tradicionales en el área de geriatría para utilizar como modelo de referencia

En este apartado se detalla el proceso empleado para la identificación y selección de URT en el área de geriatría, utilizados como referencia o patrón oro para comparar la validez del modelo experimental, TU®.

Para facilitar esta comparación, se realizó una búsqueda selectiva en PubMed con el fin de identificar revisiones tipo URT relevantes en el área de geriatría. Esta búsqueda se orientó hacia la identificación de revisiones representativas, en lugar de realizar una búsqueda exhaustiva. Nuestro enfoque consistió en una búsqueda focalizada en PubMed, utilizando los términos 'Umbrella' y 'Geriatric', con el objetivo de encontrar URT que sirvieran como referencia para este análisis comparativo. Esta búsqueda selectiva fue suficiente para nuestra comparación metodológica y no requería la estrategia de búsqueda bibliográfica rigurosa y estructurada típica de las revisiones sistemáticas, ya que nuestro objetivo no era cubrir todo el espectro de la literatura disponible, sino habilitar una evaluación paralela de las metodologías de síntesis. Por lo tanto, aunque la estrategia de búsqueda pueda parecer básica, fue diseñada intencionadamente para satisfacer las necesidades específicas de nuestro proyecto sin pretender una recuperación integral de la literatura, lo cual está fuera del alcance de este estudio.

Se llevó a cabo la búsqueda en PubMed en la fecha del 5 de marzo de 2023. El enlace a la búsqueda realizada es:

<https://pubmed.ncbi.nlm.nih.gov/?term=umbrella+geriatric&sort=date>

Este proceso resultó en un total de 111 referencias. Para afinar la selección, se procedió a una revisión detallada de ellas, resultando en la exclusión de 75 referencias por no ajustarse a los criterios de inclusión, de ser específicamente del área de geriatría o por no constituir revisiones bajo el formato *umbrella* o por ser protocolos o no ser sobre intervenciones o exposiciones a riesgos. Este proceso de exclusión está documentado en el **ANEXO 3**.

De la búsqueda realizada, 36 referencias cumplían los criterios de selección de ser URTs en el área de la geriatría, de las cuales 8 fueron seleccionadas de forma aleatoria, dando prioridad a los URTs más recientes, ya que estos suelen cumplir mejor con los estándares metodológicos actuales de las síntesis terciarias, para realizar el análisis comparativo detallado con proyectos paralelos desarrollados bajo el enfoque de TU®. Este flujograma y los criterios de selección aseguraron un proceso replicable, clave para la integridad metodológica de la comparativa entre ambas metodologías.

Figura 9. Flujograma búsqueda y selección URT geriatría.

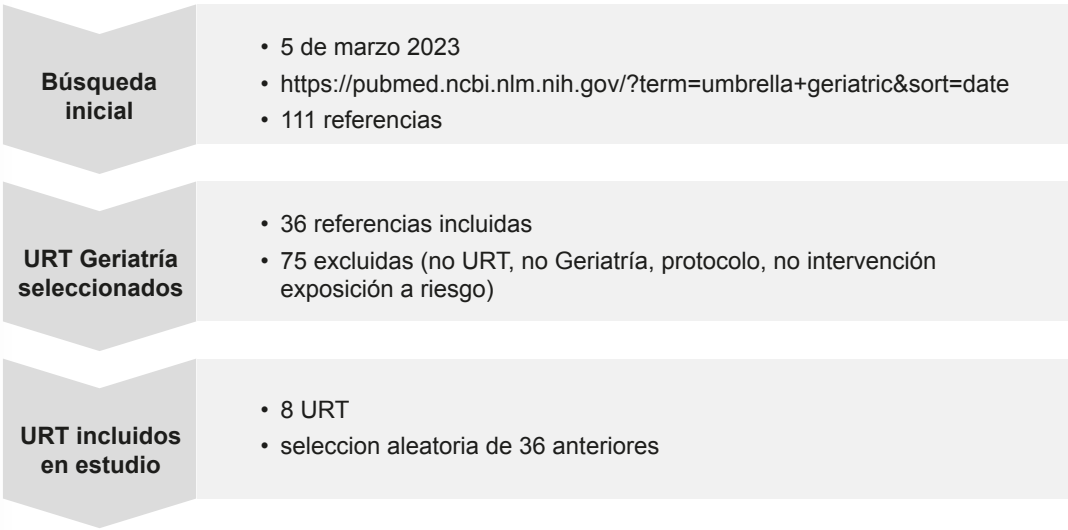


Tabla 3. URT incluidos búsqueda inicial / URT incluidos para la comparación de resultados.

URT búsqueda inicial 36/111	URT incluidos comparación 8/36 (nº proyecto)	Autor (año)	Título URT
1	4	Conneely, 2022 (161)	The effectiveness of interventions to reduce adverse outcomes among older adults following Emergency Department discharge: umbrella review.
2		Visser, 2022 (162)	Effectiveness and characteristics of physical fitness training on aerobic fitness in vulnerable older adults: an umbrella review of systematic reviews.
3		Yuan, 2022 (163)	Modifiable predictive factors and all-cause mortality in the non-hospitalized elderly population: An umbrella review of meta-analyses.
4		Travica, 2023 (164)	Peri-Operative Risk Factors Associated with Post-Operative Cognitive Dysfunction (POCD): An Umbrella Review of Meta-Analyses of Observational Studies.
5	1	Veronese, 2023 (165)	Knee osteoarthritis and adverse health outcomes: an umbrella review of meta-analyses of observational studies.
6		Li, 2022 (166)	Zinc Intakes and Health Outcomes: An Umbrella Review.
7	3	Shen, 2022 (167)	Effects of Exercise on Patients Important Outcomes in Older People With Sarcopenia: An Umbrella Review of Meta-Analyses of Randomized Controlled Trials.
8		Beck Jepsen, 2022	Predicting falls in older adults: an umbrella review of instruments assessing gait, balance, and functional mobility.
9	5	Gazzaniga, 2023 (169)	Effect of digoxin on all-cause and cardiovascular mortality in patients with atrial fibrillation with and without heart failure: an umbrella review of systematic reviews and 12 meta-analyses.

URT búsqueda inicial 36/111	URT incluidos comparación 8/36 (nº proyecto)	Autor (año)	Título URT
10	6	Musazadeh, 2023 (170)	Effect of vitamin D supplementation on type 2 diabetes biomarkers: an umbrella of interventional meta-analyses.
11	7	Veronese, 2021 (1) (171)	Interventions for reducing loneliness: An umbrella review of intervention studies.
12	8	Veronese, 2021 (2) (172)	Efficacy of conservative treatments for hand osteoarthritis: An umbrella review of interventional studies.
13		Veronese, 2020 (173)	Effect of low-dose aspirin on health outcomes: An umbrella review of systematic reviews and meta-analyses.
14		Soysal, 2023 (174)	The impact of urinary incontinence on multiple health outcomes: an umbrella review of meta-analysis of observational studies.
15		Zhang, 2021 (175)	Menopausal hormone therapy and women's health: An umbrella review.
16	2	Marx, 2021 (176)	The Dietary Inflammatory Index and Human Health: An Umbrella Review of Meta-Analyses of Observational Studies.
17		Trott, 2022 (177)	Eye disease and mortality, cognition, disease, and modifiable risk factors: an umbrella review of meta-analyses of observational studies.
18		Celotto, 2021 (178)	An umbrella review of systematic reviews with meta-analyses evaluating positive and negative outcomes of Hydroxychloroquine and chloroquine therapy.
19		Demurtas, 2021 (179)	Stem cells for treatment of cardiovascular diseases: An umbrella review of randomized controlled trials.
20		Demurtas, 2020 (180)	Physical Activity and Exercise in Mild Cognitive Impairment and Dementia: An Umbrella Review of Intervention and Observational Studies.
21		Obaid, 2020 (181)	Can we prevent poststroke cognitive impairment? An umbrella review of risk factors and treatments.
22		Veronese, 2020 (182)	Glucosamine sulphate: an umbrella review of health outcomes.
23		Trott, 2021 (183)	Hearing impairment and diverse health outcomes: An umbrella review of meta-analyses of observational studies.
24		Neil-Sztramko, 2022 (184)	Community-based group physical activity and/or nutrition interventions to promote mobility in older adults: an umbrella review.

URT búsqueda inicial 36/111	URT incluidos comparación 8/36 (nº proyecto)	Autor (año)	Título URT
25		Antonazzo, 2022 (185)	Impact of arterial hypertension and its management strategies on cognitive function and dementia: a comprehensive umbrella review.
26		Demurtas, 2020 (186)	The efficacy and safety of influenza vaccination in older people: An umbrella review of evidence from meta-analyses of both observational and randomized controlled studies.
27		Conroy, 2020 (187)	In-hospital interventions for reducing readmissions to acute care for adults aged 65 and over: An umbrella review.
28		Gielen, 2021 (188)	Nutritional interventions to improve muscle mass, muscle strength, and physical performance in older people: an umbrella review of systematic reviews and meta-analyses.
29		Biondi-Zoccai, 2020 (189)	Oral antiplatelet therapy in the elderly undergoing percutaneous coronary intervention: an umbrella review.
30		Ofori-Asenso, 2020 (190)	Frailty Confers High Mortality Risk across Different Populations: Evidence from an Overview of Systematic Reviews and Meta-Analyses.
31		Veronese, 2019 (191)	Sarcopenia and health-related outcomes: an umbrella review of observational studies.
32		Beckwée, 2019 (192)	Exercise Interventions for the Prevention and Treatment of Sarcopenia. A Systematic Umbrella Review.
33		De Spiegeleer, 2018 (193)	Pharmacological Interventions to Improve Muscle Mass, Muscle Strength and Physical Performance in Older People: An Umbrella Review of Systematic Reviews and Meta-analyses.
34		Jadczak, 2018 (194)	Effectiveness of exercise interventions on physical function in community-dwelling frail older people: an umbrella review of systematic reviews.
35		Poscia, 2018 (195)	Effectiveness of nutritional interventions addressed to elderly persons: umbrella systematic review with meta-analysis.
36		Stubbs, 2015 (196)	What works to prevent falls in older adults dwelling in long term care facilities and hospitals? An umbrella review of meta-analyses of randomised controlled trials.

En el **ANEXO 3** se detallan las referencias de la búsqueda inicial que fueron excluidos y el motivo de su exclusión. Los motivos de exclusión fueron: no ser un URT (1), no ser del área de geriatría (2), ser un protocolo (3) o no ser sobre intervenciones o riesgos (4).

4.1.2 Identificación de resultados de interés

En este apartado presentamos una exposición de los Rdl identificados en cada uno de los ocho proyectos comparativos analizados. El Rdl en un estudio de síntesis de evidencia es un aspecto específico evaluado en los estudios objeto de síntesis, sobre el que se objetiva o determina una relación de la intervención o exposición a riesgo entre los dos grupos

expuestos, y cuyos valores cuantitativos se pueden agregar o no de forma ponderada en el metaanálisis. Se comparan los resultados obtenidos por las dos metodologías para la identificación de Rdl: las URT, considerados como patrón oro en este estudio, y el modelo experimental TU®.

Presentaremos en primer lugar los resultados de forma global. A continuación, presentamos el análisis de concordancia en la identificación de Rdl. Luego, presentamos el factor de incremento de identificación de Rdl entre TU® y URT. Finalmente, en el **ANEXO 4** presentamos los resultados de identificación de Rdl por proyectos de forma individualizada. En la presentación de resultados por proyectos se describen inicialmente de forma global los resultados del proyecto y luego en dos tablas adjuntas el detalle de los Rdl identificados; primero los Rdl identificado por URT y el nivel de concordancia de TU®, y en segundo lugar los RDI identificados por TU®, no identificados por URT.

4.1.2.1 Resultados globales de Rdl identificados por TU® y URT

Se representan los resultados de Rdl identificados por cada proyecto, el nivel de concordancia y no concordancia de Rdl entre metodologías, así como los resultados globales (**Tabla 4**). En total los 8 URT incluidos en el estudio identificaron 86 Rdl, y los proyectos en paralelo ante la misma pregunta de investigación, pero con la metodología TU® identificaron 410 Rdl. De los 86 Rdl identificados por los URT, TU® coincide en identificar 73 de ellos (84.88%), mientras que no logra identificar 13 Rdl (15.11%) (**Figura 10**). De los 410 Rdl que identifica la metodología TU®, los URT coinciden en identificar 73 (17.80%), mientras que los URT dejaron de identificar 337 Rdl (82.19%) de los que identificó TU® (**Figura 11**). El nivel de concordancia de Rdl de URT por TU® por proyectos varía entre el 70% del proyecto 4 y el 100% de los proyectos 1, 5, 6, 7 y 8. Los URT de los proyectos 2, 3 y 4 son los que presentaron el mayor número de Rdl no identificados por TU®, con un 21.05% (N=8), 28.57% (N=2) y 30.00% (N=3) de Rdl no identificados por TU® respectivamente.

Tabla 4. Tabla global de Rdl identificados por URT y TU®: totales, coincidencia y no coincidencia.

Proyecto	1		2		3		4		5		6		7		8		Total	
	URT	TU®	URT	TU®	URT	TU®	URT	TU®	URT	TU®	URT	TU®	URT	TU®	URT	TU®	URT	TU®
Rdl Identificados	5	12	38	71	7	34	10	22	4	25	8	149	6	49	8	48	86	410
Rdl de URT identificados por TU® (N) (% de URT)	5 (100%)		30 (78.94%)		5 (71.42%)		7 (70.0%)		4 (100%)		8 (100%)		6 (100%)		8 (100%)		73 (84.88%)	
Rdl de URT NO identificados por TU® (N) (% de URT)	-		8 (21.05%)		2 (28.57%)		3 (30.00%)		-		-		-		-		13 (15.11%)	
Rdl de TU® identificados por URT (N) (% de TU®)	5 (41.66%)		30 42.25%		5 (14.70%)		7 (31.81%)		4 (16.00%)		8 (5.36%)		6 (12.24%)		8 (16.66%)		73 (17.80%)	
Rdl de TU® NO identificados por URT (N) (% de TU®)	7 (58.33%)		41 (57.74%)		29 (85.29%)		15 (68.18%)		21 (84.00%)		141 (94.63%)		43 (87.75%)		40 (83.33%)		337 (82.19%)	

Figura 10. Rdl identificados por URT concordancia TU®.

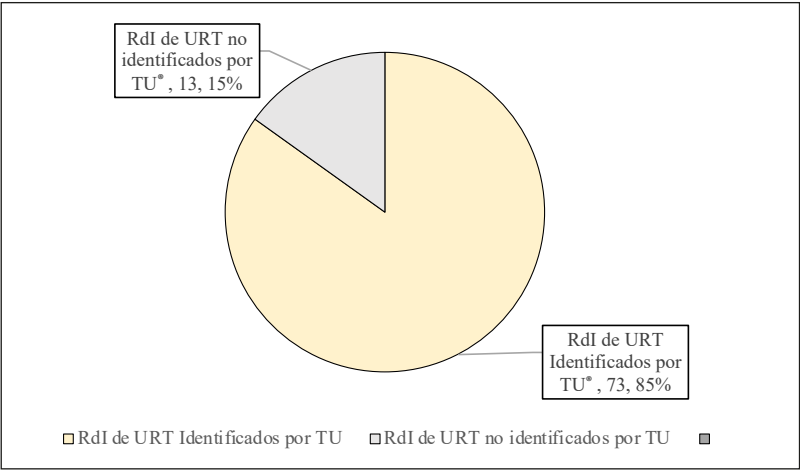
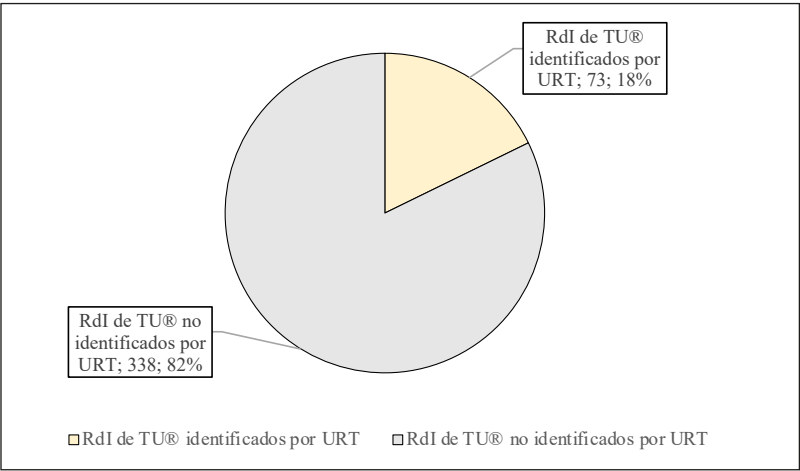


Figura 11. Rdl identificados por TU® concordancia URT.



4.1.2.2 Análisis de concordancia en la identificación de Rdl entre TU® y URT

Se realizó un análisis de la concordancia en la identificación de Rdl entre las dos metodologías, TU® y URT (**Tabla 5**).

Tabla 5. Tabla de contingencia. Número de Rdl identificados por URT frente número de Rdl identificados por TU®.

TU®	URT		
	No identificado por URT	Identificado por URT	
No identificado por TU®	-	15.1% (13)	3.1% (13)
Identificado por TU®	100% (337)	84.9% (73)	96.9% (410)
	79.7% (337)	20.3% (86)	

4.1.2.3 Factor de incremento en la identificación de Rdl por TU®

Se expone la **Tabla 6** el número de Rdl identificados por las dos metodologías, y el factor de incremento de identificación de TU® respecto a URT. Este factor multiplicador se define como el número de veces que TU® identifica más Rdl en relación con URT, proporcionando una medida directa de incremento en la identificación. Se observa que en todos los proyectos la metodología experimental TU® logra identificar un número mayor de Rdl frente a los URT. En total suponen 324 Rdl adicionales identificados por TU® que no fueron identificados por los URT, lo que supone un factor de identificación global de Rdl en la metodología TU® frente a URT de 4.77 veces, con un intervalo por proyectos que varía entre 1.87 en el proyecto 2, a 18.63 en el proyecto 6.

Tabla 6. Rdl identificados por TU® y URT, y Rdl adicionales y factor de identificación de TU®.

Proyecto	N.º Rdl identificados por TU®	N.º Rdl identificados por URT	Rdl adicional identificado por TU®	Factor de identificación de Rdl de TU® frente URT
Proyecto 1	12	5	7	2,40
Proyecto 2	71	38	33	1,87
Proyecto 3	34	7	27	4,86
Proyecto 4	22	10	12	2,20
Proyecto 5	25	4	21	6,25
Proyecto 6	149	8	141	18,63
Proyecto 7	49	6	43	8,17
Proyecto 8	48	8	40	6,00
TOTAL	410	86	324	4,77

4.1.3 TIEMPOS DE EJECUCIÓN

En este apartado analizaremos el tiempo de ejecución para realizar los proyectos de síntesis terciaria utilizando el método experimental TU®. No fue posible medir ni analizar los tiempos de ejecución de las URT, ya que no se lleva un registro de este dato en los proyectos seleccionados ni en la literatura disponible. Aunque se acepta como referencia los tiempos reportados por expertos y organizaciones internacionales, que estiman un rango de 6 a 12 meses para completar una URT, esta información no permite realizar una comparación directa con los resultados de TU®. Por ello, nos centramos exclusivamente en el análisis de los tiempos de ejecución registrados con la metodología TU®.

En la **Tabla 7** se presenta un resumen de los 8 proyectos incluidos y el tiempo empleado por TU® para la obtención de los resultados, de media el tiempo empleado fue de 4h 46' (±2h 30), empleando de mediana un tiempo de búsqueda de 4h 36'. Como se puede apreciar, existe cierta dispersión en los tiempos de ejecución, ya que encontramos mínimos de 1h 34' como es el caso del proyecto de Gazzaniga, 2023 (169) y máximos como el de Musazadeh, 2023 (170) con un tiempo de ejecución de 10h 08'.

Tabla 7. Número de proyecto, fecha, tiempo y números de términos de búsqueda y coincidencia referencias entre metodologías.

Proyecto	Fecha Búsqueda	Tiempo	Nº Términos Búsqueda TU®	Nº Referencias URT	Nº Referencias por TU®	Coincidencia Referencias (%)
P. 1 Veronese, 2023	01/08/2023	2:37:27	16	5	8	100
P. 2 Marx, 2021	01/05/2023	4:54:57	60	15	54	100
P. 3 Shen 2022	01/05/2023	5:29:22	19	6	18	100
P. 4 Conneely, 2022	01/05/2023	5:37:14	6	16	23	100
P. 5 Gazzaniga, 2023	01/05/2023	1:34:37	15	11	19	90.9
P. 6 Musazadeh, 2023	12/11/2023	10:08:24	41	37	90	97.2
P. 7 Veronese, 2021 (1)	12/12/2023	4:17:43	53	7	35	100
P. 8 Veronese 2021 (2)	23/12/2023	3:33:54	16	8	22	88.8

4.1.3.1 Análisis de la relación entre el número de referencias por TU® y el tiempo de ejecución

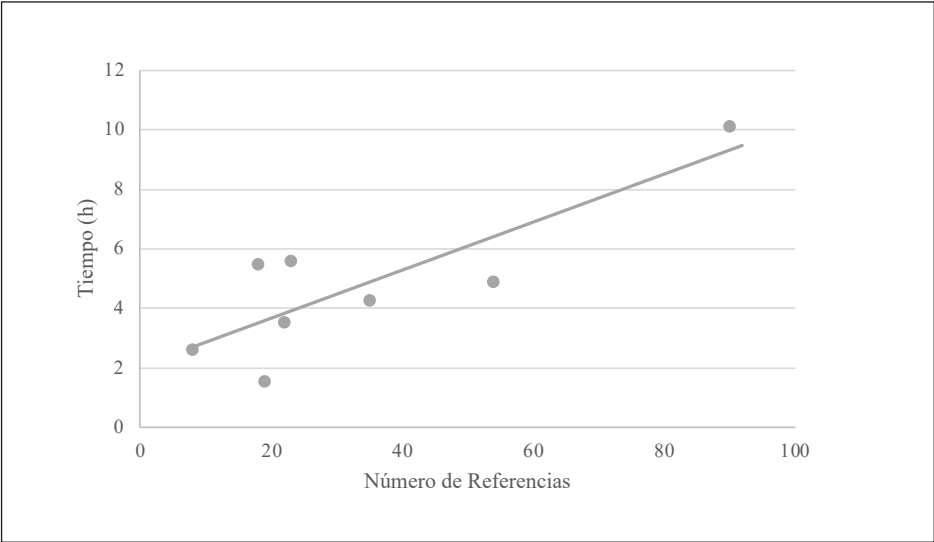
Este análisis busca explorar si existe una relación entre el número de referencias incluidas en los proyectos de TU® y el tiempo de ejecución de dichos proyectos (**Figura 12**).

El análisis muestra un coeficiente de determinación R² de 0.693, lo cual indica una relación considerable entre el número de referencias incluidas y el tiempo de ejecución del proyecto. El valor de p = 0.01 sugiere que esta relación es estadísticamente significativa, lo que indica que el número de referencias por TU® tiene un impacto notable en la duración del tiempo de ejecución.

Estos resultados sugieren que, en los proyectos TU®, el tiempo de ejecución tiende a aumentar de manera significativa conforme incrementa el número de referencias incluidas. Esto apunta a que la carga de trabajo asociada al procesamiento de un mayor volumen de

referencias influye en la duración total del proyecto. En metodologías de síntesis terciaria, también en la metodología TU®, esto resulta esperable, ya que una revisión exhaustiva de cada referencia impacta directamente en el tiempo requerido para completar el proyecto.

Figura 12. Gráfico dispersión de la relación entre el Número de Referencias incluidas en los proyecto TU® y el Tiempo de ejecución (h) de los proyectos con regresión lineal.



Se realiza, a pesar del número bajo de datos, un análisis de la correlación entre ambos valores con el coeficiente de Pearson (**Tabla 8**). (Nota: el cálculo de normalidad de los datos del número de referencias de TU® y el tiempo de ejecución no es < 0.05, se observa que los datos siguen una distribución normal, por lo que se emplea el coeficiente de Pearson (**Tabla 9**)).

Coeficiente de Correlación de Pearson presenta un valor de 0.833, lo cual representa una fuerte correlación positiva entre el número de referencias y el tiempo de ejecución. Esto significa que, a medida que aumenta el número de referencias incluidas en el proyecto, también tiende a aumentar el tiempo de ejecución. El nivel de significación estadística (valor-p = 0.01), es estadísticamente significativo, indica que es poco probable que esta relación observada sea producto del azar y respalda la existencia de una correlación real entre el número de referencias y el tiempo de ejecución.

Tabla 8. Medidas de Asociación y Significación de Correlación entre Número de Referencias y Tiempo para proyectos TU®.

Estadístico	Valor (Est – p-valor)	Resultado	Fuerza
Pearson (r)	.833* (.01)	Significativo	Fuerte

^{NS} No significativo, [†] Casi significativo ($p < .10$), ^{*} Significativo ($p < .05$) y ^{**} Alt. significativo ($p < .01$)

Tabla 9. Cálculo normalidad de variables relacionados con tiempo de ejecución TU®.

	N Términos TU	Nº Referencias URT	Nº Ref TU	Tiempo Búsqueda
W de Shapiro-Wilk	0.854	0.757	0.823	0.901
Valor p de Shapiro-Wilk	0.104	0.010	0.050	0.296

Es importante señalar que, aunque los resultados estadísticos son correctos, el análisis se ha realizado con un tamaño de muestra reducido (N = 8). Esto limita la capacidad para realizar inferencias o extrapolar estos resultados a una población mayor. Sin embargo, se puede concluir que parece probable que exista una relación directa y positiva entre el número de referencias y el tiempo de trabajo en TU® (con un coeficiente de Pearson de 0.833).

4.1.4 Evaluación de la concordancia y correlación de los tamaños de efecto derivados de metodologías comparativas

Es necesario aclarar que esta evaluación comparativa se ha realizado en 3 niveles, con variación de datos según se ha ido avanzando del primer conjunto (Nivel 0), al segundo (Nivel A) hasta el tercero (Nivel B).

El Nivel O se corresponde con los 66 TE de Rdl identificados por ambas metodologías, con todas las métricas de URT transformados a la *d* de Cohen, y comparadas con la métrica R_{TU} de TU®.

El Nivel A donde se mantuvo las métricas de Rdl de los URT transformados a la *d* de Cohen, excepto en los casos en que se presentaban con una métrica de SMD que fueron utilizados directamente como TE sin transformar en la *d* de Cohen. Además, un valor fue eliminado por no quedar adecuadamente justificado su procedencia en el URT (correspondiente al Rdl identificado con el código FO15 en la base de datos del proyecto), lo que resultó en un total de 65 datos incluidos. Estos ajustes fueron realizados tras el análisis de los TE discordantes en dos o más niveles, identificados en el Análisis Nivel O. Se observó que las *d* de Cohen discordantes provenientes de SMD no se alineaba con los resultados y conclusiones sobre el TE de las RS/MA incluidas en las URT. Como primer ajuste, se decidió no transformar los SMD a *d* de Cohen; utilizándolos directamente como TE (**ANEXO 5**).

En el Nivel B, además fueron eliminados los TE de URT que originalmente estuviesen en MD o WMD. Lo que reduce el número de datos a 48 en la evaluación de este nivel. La falta de los tamaños muestrales impidió la estandarización o ajuste de las MD o WMD a SMD y posterior transformación a la *d* de Cohen. Además, se observó que algunos de los TE que diferían en más de un nivel de categoría en el Análisis Nivel O de TU® y URT correspondían a casos donde la extracción es a partir de MD o WMD.

4.1.4.1 Evaluación de la concordancia y correlación derivadas de metodologías comparativas. Nivel O

En la evaluación Nivel O se utilizan todos los Rdl que ambas metodologías coinciden en identificar, y todas las métricas de TE de URT transformados a la *d* de Cohen. El número total de valores es de 66. El motivo de limitar los 86 Rdl identificados por los URT a solo 66 proviene de dos circunstancias. Por un lado, en el proyecto 4 el URT no expresa el TE de los Rdl (N=10). Por otro lado, TU® no identifica 10 de los Rdl identificados por los URT.

4.1.4.1.1 Evaluación de la concordancia de los TE derivados de metodologías comparativas. Nivel O

En la **Tabla 10** se muestran la tabla de contingencia de los resultados para la prueba de contraste entre la categorización en términos cualitativos de los TE de URT frente a los TE de TU®. Centrado la evaluación de concordancia desde TU®, observamos los siguientes niveles de concordancia. TU® categoriza de TE cualitativo Trivial a 18 valores (27.3%), de los cuales URT coincide en 10 valores como Triviales (55.6%), 7 valores (38.9%) en la categoría Pequeño y uno como Grande (5.6%). TU® categoriza de TE cualitativo Pequeño a 30 valores (45.5%), de los cuales URT coincide de forma plena en 7 (23.3%), considera Triviales 4 (13.3%), Moderado 5 (16.7%) y Grande 14 (46.7%). TU® categoriza de TE cualitativo Moderado a 16 valores (24.2%), de los cuales URT solo coincide con 1 (6.1%), considera Pequeño 1 (6.3%) y Grande a 14 (87.5%). TU® categoriza de TE cualitativo Grande a 2 valores (3.0%), de los cuales URT coincide en 1 (50.0%) y considera de TE Moderado el restante (50.0%).

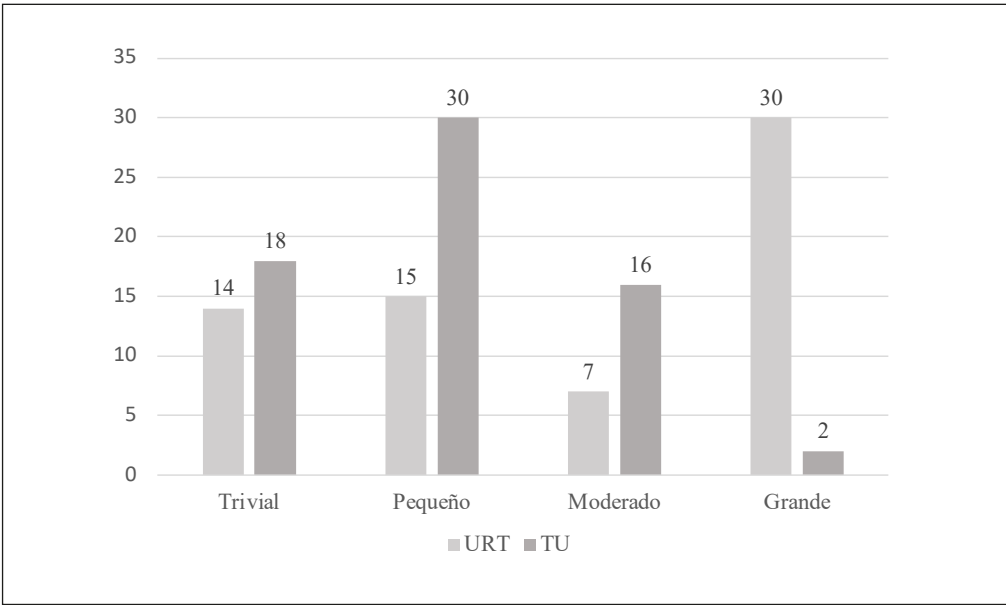
Tabla 10. Tabla de contingencia. Variables TE términos cualitativos: TE URT vs TE TU®. Nivel O.

URT - TE	TU - TE				
	1.Trivial	2.Pequeño	3.Moderado	4.Grande	
1. Trivial	55.6% (10)	13.3% (4)	-	-	21.1% (14)
2.Pequeño	38.9% (7)	23.3% (7)	6.3% (1)	-	22.7% (15)
3.Moderado	-	16.7% (5)	6.3% (1)	50.0% (1)	10.6% (7)
4.Grande	5.6% (1)	46.7% (14)	87.5% (14)	50.0% (1)	45.5% (30)
	27.3% (18)	45.5% (30)	24.2% (16)	3.0% (2)	(66)
$\chi^2 = 37.85^{**}$ ($p < 0.001$), $V = 0.437$ (Moderado)					
Notas					
^{NS} No significativo, † Casi significativo ($p < .10$), * Significativo ($p < .05$) y ** Altamente significativo ($p < .01$) Test estadístico/contraste: χ^2 - Chi Cuadrado. Tamaño o magnitud del efecto: V - V de Cramer.					
Concordancia Plena	Concordancia Parcial (Discrepancia de 1 nivel)		Discordancia (Discrepancia de > 1 nivel)		

Representado en la **Figura 13**, se presentan los valores por categoría de TE de ambas metodologías. URT considera categoría Triviales 21,2% (N=14), Pequeño 22.7% (N=15), Moderado 10.6% (N=7) y Grande 45.5% (N=30). TU® considera categoría Triviales 27.3%

(18), Pequeño 45.5% (N= 30), Moderado 24.2% (N=16) y Grande el 3.0% (N=2). Destaca, que URT valora como Grande a la mayor parte de sus TE (N=30; 45,45%); mientras que TU® clasifica como Pequeño a la mayor parte de sus TE (N=30; 45,45%). Lo que nos puede inicialmente llegar a indicar que TU® valora de forma menor el TE comparado con URT.

Figura 13. Categorías de TE cualitativo según metodologías. Nivel O (N=66).



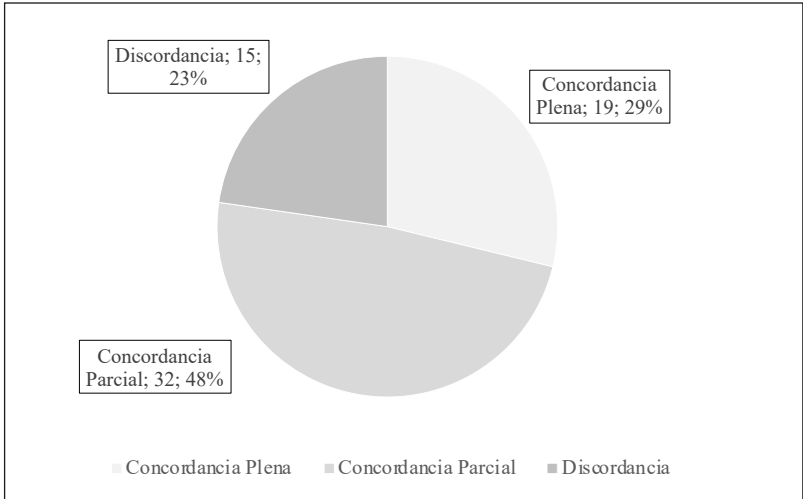
Se observa que, de los 66 valores, en 19 casos (28.7%) se produce una Concordancia Plena, en 32 casos (48.8%) se produce una Concordancia Parcial y en 15 casos (22.7%) se produce Discordancia. El conjunto de Coincidencia Consistente incluye 51 de los 66 valores (77.2%) (**Tabla 11**) (**Figura 14**).

Nota sobre nivel de concordancia de categorías de TE entre metodologías. Consideramos como “**Concordancia Plena**” cuando ambas metodologías categorizan a los TE en la misma categoría. Consideramos como “**Concordancia Parcial**” cuando la categorización de los TE entre las dos metodologías discrepan en un nivel, por ejemplo, si una metodología considera un TE Pequeño, la otra metodología no lo categoriza como Pequeño, pero lo categoriza como Trivial o Moderado, nunca como Grande. El rango de diferencia entre las categorías de ambas metodologías no es mayor que el rango de una categoría completa. Y, por último, consideramos “**Discordancia**” cuando existe una diferencia en la clasificación de categorías de TE de más de un nivel, por ejemplo, si una metodología considera un TE como Pequeño la otra metodología lo considera Grande. En este nivel el rango de diferencias entre categorías es mayor que el rango de una categoría completa. Esta diferencia de categorización observada en los TE “Discordantes” consideramos que es lo suficientemente grande para precisar un análisis para intentar entender su motivo. El conjunto de “Concordancia Plena” y “Concordancia Parcial” lo denominaremos “**Concordancia Consistente**”.

Tabla 11. Distribución de Concordancia y Discrepancia entre Metodologías según Nivel de Acuerdo. Nivel O.

	Frec. Absoluta	Frec. Absoluta Acumulada	Frec. Relativa	Frec. Relativa Acumulada
Diferente TE	47	47	71.2%	71.2%
Mismo TE	19	66	23.5%	100%
Mismo TE	19	19	28.8%	28.8%
Discrepancia 1 Nivel	32	51	48.5%	77.3%
Discrepancia > 1 Nivel	15	66	22.7%	100%

Figura 14. Distribución de Concordancia TE entre Metodologías Nivel O (N=66).



4.1.4.1.2 Evaluación de la correlación de los TE derivados de metodologías comparativas. Nivel O

En esta sección vamos a realizar un análisis de la correlación, de la relación entre las dos variables, los TE de ambas metodologías, en valor cuantitativo. Con ello conoceremos la “fuerza” y la “dirección” de la relación entre ambas variables. Para evaluar si al cambiar un valor de una metodología, el valor en la otra metodología tiende a hacerlo en la misma dirección (correlación positiva) o no (correlación negativa).

Se realiza un análisis de correlación utilizando el coeficiente para valores de distribución no normal de Spearman (Tabla 12) (Tabla 13).

Tabla 12. Distribución de Concordancia y Discrepancia entre Metodologías según Nivel de Acuerdo. Nivel O.

Estadístico	Valor (Est – p-valor)	Resultado	Fuerza
Rho Spearman	.637** (<.001)	Alt. Significativo	Moderada-Fuerte

^{NS} No significativo, † Casi significativo (p<.10), * Significativo (p<.05) y ** Alt. significativo (p<.01)

Tabla 13. Cálculo normalidad variables cuantitativos Rdl Nivel O.

Nivel O	TU® - TE (R _{TU})	URT - TE (d Cohen)
W de Shapiro-Wilk	.862	.767
Valor p de Shapiro-Wilk	<.001**	<.001**

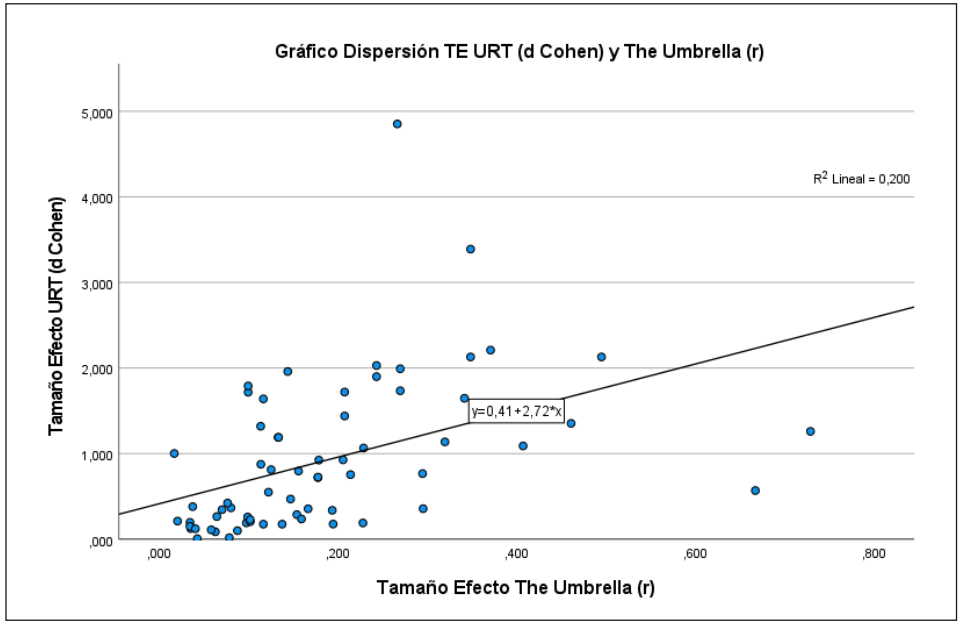
^{NS} No significativo, † Casi significativo (p<.10), * Significativo (p<.05) y ** Alt. significativo (p<.01)

Los análisis de correlación muestran resultados consistentes entre las dos metodologías evaluadas. El coeficiente Rho de Spearman (Rho = 0.637, p < 0.001) refleja una correlación positiva, significativa y de fuerza moderada a fuerte. El coeficiente de Spearman, al ser no paramétricos, es robusto frente a *outliers* y distribuciones no normales, lo que proporciona una mayor confianza en la solidez de las correlaciones observadas entre las dos metodologías. Esto destaca la consistencia de R_{TU} en relación con la métrica establecida d de Cohen.

Para complementar el análisis presentamos la gráfica de dispersión (Figura 15), que nos explica la relación entre los valores TE de URT vs TU®, en el eje de las ordenadas están los TE de URT en el eje de abscisas los TE de TU®, en este caso, además, se añade la regresión lineal entre las puntuaciones que nos habla de los resultados esperados.

URT TE vs TU TE con regresión lineal

Figura 15. Gráfico dispersión de la relación entre el TE de URT (d de Cohen) y The Umbrella (r) con regresión lineal. Nivel O.



La gráfica de dispersión muestra una relación positiva entre los tamaños de efecto de URT y TU®. La intersección de la línea de regresión, situada en 0.41, sugiere que cuando el tamaño de efecto de TU® es cero, el tamaño de efecto esperado en URT es de 0.41. La pendiente de 2.72 indica que, por cada unidad adicional en el tamaño de efecto de TU®, el tamaño de efecto en URT tiende a incrementarse en 2.72 unidades, reflejando una relación positiva, pero pronunciada entre ambas métricas. El coeficiente de determinación, R² = 0.20, señala que el 20% de la variabilidad en los tamaños de efecto de URT puede explicarse por los tamaños de efecto de TU®, lo cual sugiere que, aunque existe una relación, una parte significativa de la variabilidad en URT se debe a otros factores no explicados por esta relación lineal.

4.1.4.1.3 Conclusión Nivel O

En el Nivel O, se evaluaron la concordancia y correlación entre los TE cualitativos y cuantitativos obtenidos por las dos metodologías en estudio: URT, transformando todas las métricas en la métrica consolidada *d* de Cohen, y la R_{TU} de TU®, una métrica desarrollada específicamente para el contexto de The Umbrella Collaboration®. A partir de los 66 valores comunes identificados entre ambas metodologías, se observó que en un 28.7% de los casos hubo concordancia plena (clasificación categórica idéntica de los TE), mientras que un 48.8% mostró concordancia parcial (discrepancia de un nivel entre categorías). En el 22.7% restante se registraron discrepancias significativas (diferencia de más de un nivel). Al considerar tanto la concordancia plena como la parcial, el porcentaje de la concordancia consistente alcanza un 77.2%, lo cual indica una tendencia general de ambas metodologías a clasificar los TE de manera similar, aunque con diferencias en algunos casos.

La prueba de χ^2 resultó altamente significativa ($p < 0.001$), lo que indica que las discrepancias observadas entre las frecuencias de ambas metodologías no son producto del azar, sugiriendo una asociación estadísticamente significativa en la categorización de los TE. La V de Cramer (0.437) señala una asociación de magnitud moderada entre las categorías de ambas metodologías. En cuanto a la correlación cuantitativa, el coeficiente de Spearman (Rho = 0.637) mostró una correlación positiva y significativa, de moderada-fuerte, lo cual implica que, en general, a medida que cambia el TE en una metodología, también lo hace en la otra en la misma dirección. La gráfica de dispersión con regresión lineal refuerza esta observación, aunque con una pendiente de 2.72 y un R² de 0.20, lo que sugiere que, si bien existe una relación positiva, una gran parte de la variabilidad en los TE de URT no es explicada por los TE de TU®.

4.1.4.1.4 Análisis de los 15 valores con discrepancia de más de un nivel detectado en el Nivel O

Se realizó un análisis detallado de los 15 valores de TE de Rdl con discrepancias de más de una categoría que se observan en el Análisis Nivel O entre URT y TU® (ANEXO 5).

De los 15 valores analizados encontramos 8 (53.3%), identificados con los códigos siguientes en la base de datos del trabajo (F184, F015, F016, F333, F380, F381, F378, F382) y en que determinamos que el valor del TE del Rdl del URT transformado a la *d* de Cohen no se puede asumir como cierta. Los motivos son diversos (Tabla 14).

Por lo que, para no sesgar los estudios y mantener criterios comunes, llegamos a los siguientes acuerdos metodológicos:

- 1. En el Análisis Nivel A, los SMD se utiliza directamente la magnitud expresada en el URT como TE, sin transformarlos a la *d* de Cohen.
- 2. En el Análisis Nivel B, no se utiliza ningún valor MD ni WMD.

Tabla 14. Evaluación y Decisiones sobre Valores Discordantes (N=15) en el TE entre URT y TU en el Nivel O.

Ítem	Explicación	Decisión	Justificación
F184	No es correcto que el WMD de 1.083 sea GRANDE, al menos clínicamente no es un efecto grande	Eliminar en la parte B, se mantiene en A	Se utiliza WMD que no es un TE por sí solo, la transformación no ajusta del todo bien, sesgando el estudio.
F015	El valor utilizado por el URT no figura en ninguna de las referencias del trabajo, por lo que debemos asumir que no es cierto.	Eliminar del estudio, tanto en A como en B	Al no existir la referencia no se puede corroborar que el resultado sea correcto, por lo que, se toma en consideración eliminarlo del estudio.
F016	El URT utiliza la referencia 29 (...) En dicha referencia, sin embargo, el RR que aparece es: pancreatic cancer, 1.16 (95% CI: 1.05, 1.28; I2 = 61.6%; n = 2).	Se sustituye en valor RR 2.54 por 1.16, pasando de Grande a Pequeño.	El resultado que aparece y al que se referencia en el URT no es el correcto, tras revisarlo, se cambia.
F028	Es correcta la diferencia (Es un OR)	Ninguna	Son correctos los valores aportados tras la revisión.
F099	Se trata de un valor MD	Eliminar en la parte B, se mantiene en A	Se trata de un valor MD.
F184	Es correcta la diferencia.	Eliminar en la parte B, se mantiene en A	Se trata de un valor WMD.
F185	Es correcta la diferencia (Es un SMD)	Ninguna	Son correctos los valores aportados tras la revisión.
F330	Es correcta la diferencia.	Eliminar en la parte B, se mantiene en A	Se trata de un valor MD.
F184	Es correcta la diferencia.	Eliminar en la parte B, se mantiene en A	Se trata de un valor WMD.
F333	Se cambia la forma de tratar SMD, utilizamos el valor literal de SMD como TE.	Rehacer cálculos y mantener en A y en B.	En el ajuste de SMD en algunos casos se obtienen valores con SD próximos al valor del SE, al dividir salen valores muy altos (más de 1) que sería un TE grande.
F380			
F381			
F378			
F382			
F377	Es correcta la diferencia (Es un SMD)	Ninguna	Son correctos los valores aportados tras la revisión.

Nota. El valor F332 toma el valor 0.49 (.485), lo que en cierta forma lo hace estar muy próximo de la valoración de mediano (a partir de 0.50) en vez de pequeño.

4.1.4.2 Evaluación de la concordancia y correlación derivadas de metodologías comparativas. Nivel A

En la evaluación Nivel A se realiza la comparación de los valores en formato cualitativo y cuantitativo de los TE de los Rdl coincidentes entre ambas metodologías con todos los datos (excepto el valor F015, que fue eliminado por no estar bien documentado en la URT)

(N=65), y todas las métricas de los TE de URT convertidos a la *d* de Cohen, excepto las SMD (que quedan con la magnitud en que aparecen en el URT como TE).

4.1.4.2.1 Evaluación de la concordancia derivada de metodologías comparativas. Nivel A

En la **Tabla 15** se muestran la tabla de contingencia de los resultados para la prueba de contraste entre la categorización en términos cualitativos (términos categóricos) de los TE de URT frente a los TE de TU®. Centrado la evaluación de concordancia desde TU®, observamos los siguientes niveles de concordancia. TU® categoriza de TE Trivial a 18 valores (27.3%), de los cuales URT coincide en 10 valores como Triviales (55.6%), 7 valores (38.9%) de la categoría Pequeño y un valor como Grande (5.6%). TU® categoriza de TE Pequeño a 29 valores (45.5%), de los cuales URT coincide de forma plena en 13 (44.3%); mientras considera Triviales 6 (20.7%) y Moderado 6 (20.7%) y Grande 4 (13.8%). TU® categoriza de TE Moderado a 16 valores (24.2%), de los cuales URT solo coincide con 2 (12.5%); mientras que considera Trivial 1 (6.3%), Pequeño 4 (25.0%) y Grande a 9 (56.3%). TU® categoriza de TE Grande a 2 valores (3.0%), de los cuales URT coincide en 1 (50.0%) y considera de TE Moderado el restante (50.0%).

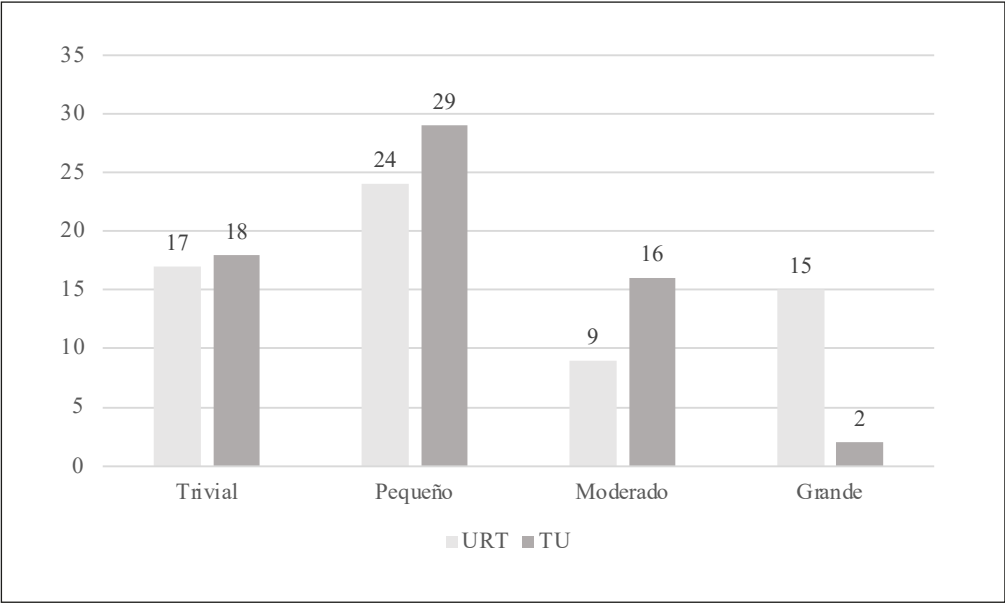
Tabla 15. Tabla de contingencia. Variables TE términos cualitativos: TE URT vs TE TU®. Nivel A.

URT - TE	TU® - TE				
	1.Trivial	2.Pequeño	3.Moderado	4.Grande	
1. Trivial	55.6% (10)	20.7% (6)	6.3% (1)	-	26.2% (17)
2.Pequeño	38.9% (7)	44.3% (13)	25.0% (4)	-	36.9% (24)
3.Moderado	-	20.7% (6)	12.5% (2)	50.0% (1)	13.8% (9)
4.Grande	5.6% (1)	13.8% (4)	56.3% (9)	50.0% (1)	23.1% (15)
	27.3% (18)	45.5% (29)	24.2% (16)	3.1% (2)	(65)
$\chi^2 = 28.21^{**}$ ($p < 0.001$), $V = 0.380$ (Moderado)					
Notas					
^{ns} No significativo, † Casi significativo ($p < .10$), * Significativo ($p < .05$) y ** Altamente significativo ($p < .01$) Test estadístico/contraste: χ^2 - Chi Cuadrado. Tamaño o magnitud del efecto: V - V de Cramer.					
Concordancia Plena	Concordancia Parcial (Discrepancia de 1 nivel)		Discordancia (Discrepancia de > 1 nivel)		

Representado en la **Figura 16**, se presentan los valores por categoría de TE de ambas metodologías. URT considera valores Triviales 26,2% (N=17), Pequeño 36.9% (N=24), Moderado 13.8% (N=9) y Grande 23.1% (N=15). TU® considera valores Triviales 27.7%

(N=18), Pequeño 44.6% (N= 29), Moderado 24.2% (N=16) y Grande el 3.1% (N=2). Destaca, en URT la disminución de valores categorizados como Grandes en Nivel O 45.5% (N=30) a 23.1% (N=15).

Figura 16. Categorías de TE cualitativo según metodologías. Nivel A (N=65).

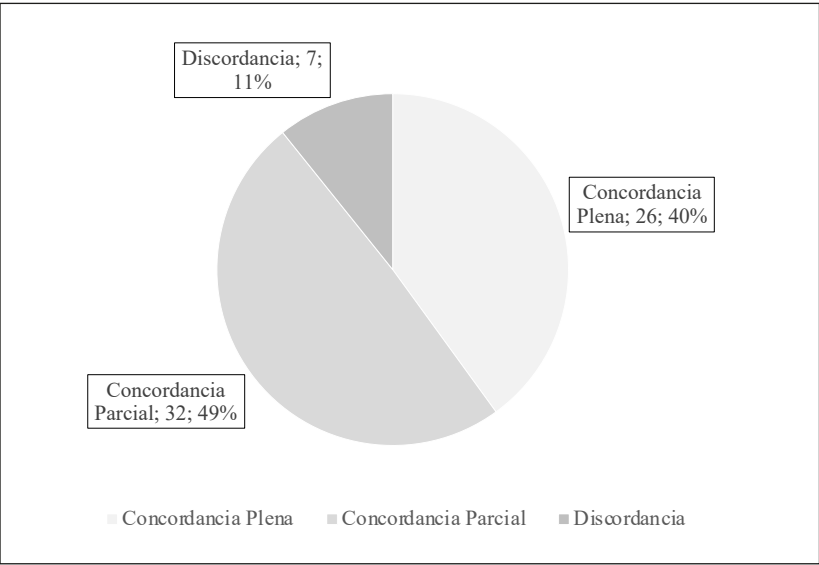


Se observa que, de los 65 valores, en 26 casos (40.0%) se produce una Concordancia Plena; en 32 casos (49.2%) se produce una Concordancia Parcial y en 7 casos (10.8%) se produce Discordancia. El conjunto de Coincidencia Consistente incluye 58 de los 65 valores (89.2%) (**Tabla 16**) (**Figura 17**).

Tabla 16. Distribución de Concordancia y Discrepancia entre Metodologías según Nivel de Acuerdo. Nivel A.

	Frec. Absoluta	Frec. Absoluta Acumulada	Frec. Relativa	Frec. Relativa Acumulada
Diferente TE	39	39	60.0%	60.0%
Mismo TE	26	65	40.0%	100%
Mismo TE	26	26	40.0%	40.0%
Discrepancia 1 Nivel	32	58	49.2%	89.2%
Discrepancia > 1 Nivel	7	65	10.8%	100%

Figura 17. Distribución de Concordancia TE entre Metodologías. Nivel A (N=65).



4.1.4.2.2 Evaluación de la correlación derivada de metodologías comparativas. Nivel A

En esta sección se realizó un análisis de la correlación entre las dos variables, los TE de ambas metodologías, en valor cuantitativo. Con ello conocemos la “fuerza” y la “dirección” de la relación entre ambas variables. Para evaluar si al cambiar un valor de una metodología, el valor en la otra metodología tiende a hacerlo en la misma dirección (correlación positiva) o no (correlación negativa). Se realiza un análisis de correlación utilizando el coeficiente de Spearman (Tabla 17), se presenta el cálculo de normalidad (Tabla 18).

Tabla 17. Medidas de Asociación y Significación de Correlación de TE entre Metodologías. Nivel A.

Estadístico	Valor (Est – p-valor)	Resultado	Fuerza
Rho Spearman	.512** (<.001)	Alt. Significativo	Moderada-Fuerte

NS No significativo, † Casi significativo (p<.10), * Significativo (p<.05) y ** Alt. significativo (p<.01)

Tabla 18. Cálculo normalidad variables cuantitativos RdI Nivel A.

Nivel A	TU® - TE (R _{TU})	URT - TE (d Cohen)
W de Shapiro-Wilk	.865	.770
Valor p de Shapiro-Wilk	<.001**	<.001**

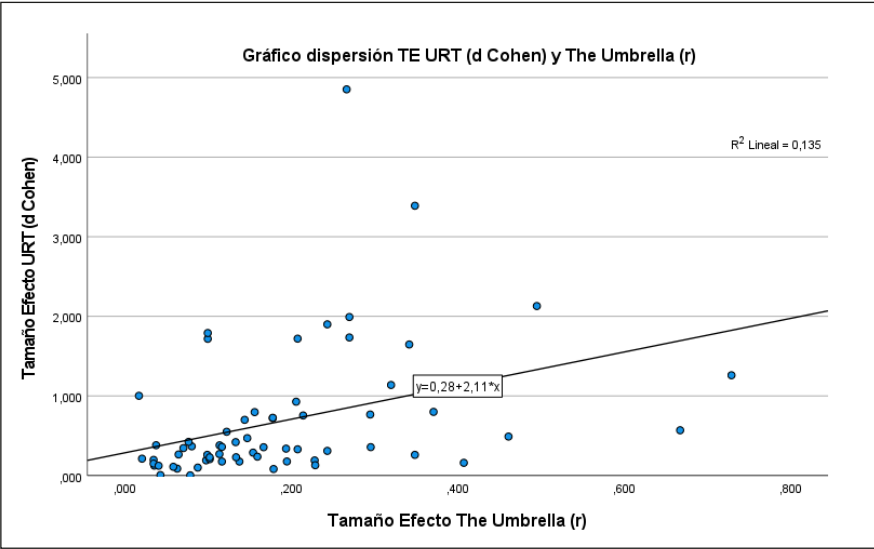
NS No significativo, † Casi significativo (p<.10), * Significativo (p<.05) y ** Alt. significativo (p<.01)

Los análisis de correlación en este nivel muestran una ligera disminución respecto al Nivel O, pero mantienen resultados positivos, altamente significativos y de fuerza moderada. El coeficiente Rho de Spearman (Rho = 0.512, p < 0.001) indica una correlación moderada a fuerte. Al ser un coeficiente no paramétrico, es robusto ante outliers y distribuciones no normales, lo que refuerza la fiabilidad y solidez de las correlaciones observadas. Esto destaca la consistencia de R_{TU} en relación con la métrica consolidada d de Cohen.

Para complementar el análisis presentamos la gráfica de dispersión que nos explica la relación entre los valores TE cuantitativos de URT vs TU®, en el eje de las ordenadas están los TE de URT y en el eje de abscisas los TE de TU®, en este caso, además, se añade la regresión lineal entre las puntuaciones que nos habla de los resultados esperados (Figura 18).

URT TE vs TU TE con regresión lineal

Figura 18. Gráfico dispersión de la relación entre el TE de URT (d de Cohen) y The Umbrella (r) con regresión lineal. Nivel A.



En el análisis de la correlación entre TE en el Nivel A, observamos mejoras en ciertos aspectos clave con respecto al análisis de correlación en Nivel O. La intersección ha disminuido de 0.41 en Nivel O a 0.28, lo que indica una mayor alineación inicial entre URT y TU®. La pendiente se ha reducido a 2.11 (desde 2.72 en Nivel O), lo que refleja una relación discretamente menos pronunciada. Estos resultados refuerzan la coherencia y consistencia entre ambas metodologías, a pesar de una R² = 0.135 (Nivel O: R² = 0.200) que sugiere que el modelo de regresión lineal no explica bien la variabilidad entre los valores numéricos de los TE de ambas metodologías.

4.1.4.2.3 Conclusión Nivel A

En el Nivel A, la evaluación de la concordancia y correlación entre los TE de ambas metodologías muestra mejoras significativas en comparación con el Nivel O. En esta

ocasión, se trabajó con 65 valores tras eliminar un valor no plenamente justificado en el URT, y de mantener los valores de los Rdl que emplean el SMD, sin transformarlos a la *d* de Cohen. En términos de concordancia, el Nivel A registra un aumento en la concordancia plena, que asciende al 40.0% (frente al 28.7% en el Nivel O), mientras que la concordancia parcial se mantiene estable en un 49.2%. Esto da lugar a una concordancia consistente total del 89.2%, un incremento notable respecto al 77.2% observado en el Nivel O. La discordancia también se reduce drásticamente a solo un 10.8%, indicando una mejora en la alineación de ambas metodologías y una menor discrepancia en la categorización de los TE. Este ajuste sugiere que los cambios en el manejo de las métricas de TE han tenido un impacto positivo, reduciendo la categorización de "Grande" en URT y acercando la clasificación de ambas metodologías.

La prueba de χ^2 sigue siendo altamente significativa ($p < 0.001$), lo que confirma que las diferencias observadas no son producto del azar. Sin embargo, el valor de V de Cramer disminuye ligeramente a 0.380, indicando una asociación moderada pero ligeramente inferior a la del Nivel O (0.437). En cuanto a la correlación cuantitativa, el coeficiente de Spearman ($Rho = 0.512$) siguen siendo positivo y altamente significativo, sugiriendo una relación de moderada a fuerte. La gráfica de dispersión también refleja una mejora en la alineación inicial entre ambas metodologías, con una intersección de 0.28 frente a 0.41 en el Nivel O y una ligera reducción de la pendiente (2.11), lo que sugiere una relación menos pronunciada y una mayor consistencia. El R^2 disminuye a 0.135, lo que indica que aún queda un margen de variabilidad sin explicar, aunque los ajustes en las métricas han llevado a una mayor coherencia general entre los TE de ambas metodologías.

4.1.4.3 Evaluación de la concordancia y correlación derivadas de metodologías comparativas Nivel B

En este Análisis Nivel B se mantuvo la modificación del Nivel A (mantener magnitud de SMD en valor original) y se eliminaron los Rdl con TE en formato de MD y WMD. Debido a que, la falta de los tamaños muestrales no permite la estandarización o ajuste correcta a SMD y posterior transformación a la *d* de Cohen, además, se observó que en algunos de los casos en los que difieren los TE de TU® y URT (Nivel O) se encontró que el valor de la métrica, MD o WMD, no estaba correctamente anotada en el URT (ANEXO 5). El conjunto de datos queda reducido a N=48. Se ha reducido el número de datos con respecto al Nivel O (N=66) en un 27.27% y con respecto al Nivel A (N=65) en un 26.15%.

4.1.4.3.1 Evaluación de la concordancia derivada de metodologías comparativas. Nivel B

En la **Tabla 19** se muestran la tabla de contingencia de los resultados para la prueba de contraste entre la categorización en términos cualitativos (término categórico) de los TE de URT frente a los TE de TU®. Centrado la evaluación de concordancia desde TU®, observamos los siguientes niveles de concordancia. TU® categoriza de TE Trivial a 17 valores (35.4%), de los cuales URT coincide en 10 valores como Triviales (58.8%) y 7 valores (41.2%) los URT los categorizan como Pequeño. TU® categoriza de TE Pequeño a 26 valores (54.2%), de los cuales URT coincide de forma plena en 13 (50.0%), mientras considera Triviales a

5 (19.2%), 6 valores como Moderado (23.1%) y 2 como Grandes (7.7%). TU® categoriza de TE Moderado a 5 valores (10.4%), de los cuales URT solo coincide con 1 (20.0%), considera Pequeño 3 (60.0%) y Trivial a 1 (20.0%). TU® ni URT clasifican a ningún TE en la categoría Grande.

Tabla 19. Tabla de contingencia. Variables TE términos cualitativos: TE URT vs TE TU®. Nivel B.

URT - TE	TU® - TE				
	1.Trivial	2.Pequeño	3.Moderado	4.Grande	
1. Trivial	58.8% (10)	19.2% (5)	20.0% (1)	-	33.3% (16)
2.Pequeño	41.2% (7)	50.0% (13)	60.0% (3)	-	47.9% (23)
3.Moderado	-	23.1% (6)	20.0% (1)	-	14.6% (7)
4.Grande	-	7.7% (2)	-	-	4.1% (2)
	35.4% (17)	54.2% (26)	10.4% (5)	-	(48)
$\chi^2 = 28.21^{**}$ ($p < 0.001$), $V = 0.380$ (Moderado)					
Notas ^{NS} No significativo, † Casi significativo ($p < .10$), * Significativo ($p < .05$) y ** Altamente significativo ($p < .01$) Test estadístico/contraste: χ^2 - Chi Cuadrado. Tamaño o magnitud del efecto: V - V de Cramer.					
Concordancia Plena	Concordancia Parcial (Discrepancia de 1 nivel)		Discordancia (Discrepancia de > 1 nivel)		

Representado en la **Figura 19** se presentan los valores por categoría de TE de ambas metodologías. URT considera valores Triviales 33.3% (N=16), Pequeño 47.9% (N=23), Moderado 14.6% (N=7) y Grande 4.2% (N=2). TU® considera valores Triviales 35.4% (N=17), Pequeño 54.2% (N= 26), Moderado 10.4% (N=5) y Grande el 0.0% (N=0). Destaca, en la mayor alineación en la clasificación de las categorías de TE entre las dos metodologías en este Nivel B.

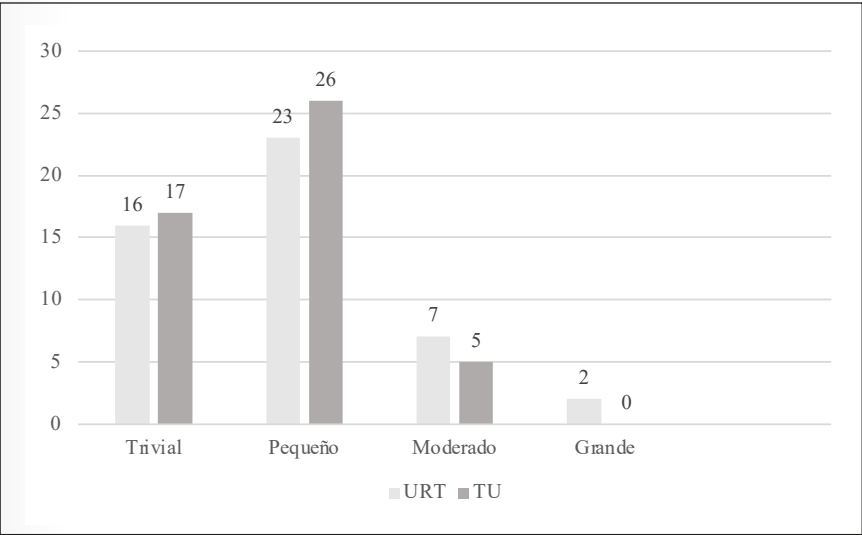


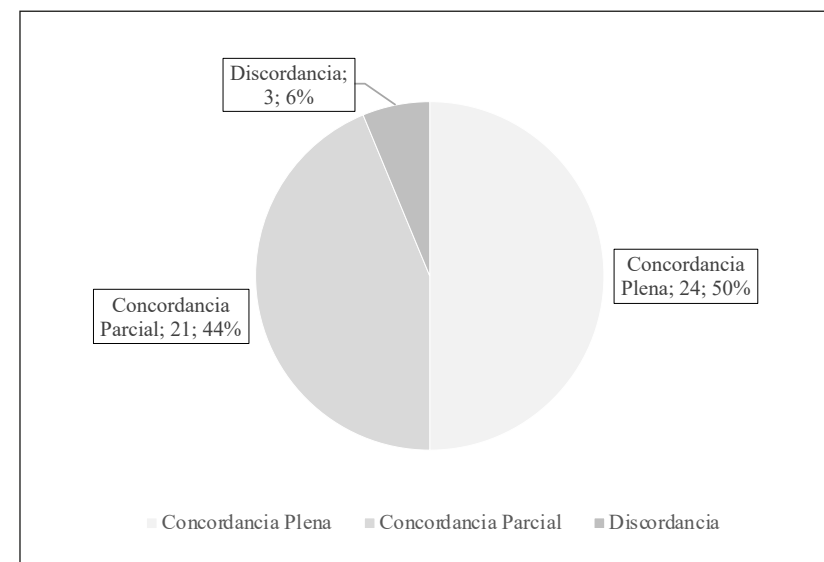
Figura 19. Categorías de TE cualitativo según metodologías. Nivel B (N=48).

Se observa que, de los 48 valores, en 24 casos (50.0%) se produce una Concordancia Plena; en 21 casos (43.8%) se produce una Concordancia Parcial y en 3 casos (6.3%) se produce Discordancia. El conjunto de Concordancia Consistente incluye 45 de los 48 valores (93.8%) (**Tabla 20**) (**Figura 20**).

Tabla 20. Distribución de Concordancia y Discrepancia entre Metodologías según Nivel de Acuerdo. Nivel B.

	Frec. Absoluta	Frec. Absoluta Acumulada	Frec. Relativa	Frec. Relativa Acumulada
Diferente TE	24	24	50.0%	50.0%
Mismo TE	24	48	50.0%	100%
Mismo TE	24	24	50.0%	50.0%
Discrepancia 1 Nivel	21	45	43.8%	93.8%
Discrepancia > 1 Nivel	3	48	6.3%	100%

Figura 20. Categorías de TE cualitativo según metodologías. Nivel B (N=48).



En el análisis del Nivel B, el estadístico χ^2 de 11.03 ($p = 0.088$), queda por debajo del valor crítico para un nivel de significación de $\alpha = 0.05$ y $\alpha = 0.01$, cuyos valores correspondientes son 16.92 y 21.67 respectivamente. Esto indica que el estadístico χ^2 no es lo suficientemente grande para rechazar la hipótesis nula, que postula la ausencia de asociación entre las variables, es decir, no existe una relación significativa entre las dos metodologías para categorizar los TE de los Rdl. Aunque se observa una distribución de frecuencias relativamente similar entre ambas metodologías en el Nivel B, llegando a una Concordancia Consistente del 93.8% (**Figura 20**), la reducción en el número total de

observaciones afecta considerablemente el valor de la prueba estadística. Este valor más bajo de χ^2 en el Nivel B, en comparación con el Nivel O y A, sugiere que la discrepancia entre los valores observados y los esperados bajo la hipótesis nula es menor. Por lo tanto, la evidencia para rechazar la hipótesis nula es más débil en este nivel, lo que indica una menor probabilidad de que la relación observada sea distinta de lo que podría ocurrir por azar. Esta pérdida de significación estadística en el Nivel B puede atribuirse principalmente a dos factores:

1. Disminución del número de observaciones. En este análisis, se trabaja con una muestra más pequeña en comparación con el Nivel O y A. Al reducir el tamaño de la muestra, las diferencias en las frecuencias observadas y esperadas se vuelven menos detectables estadísticamente, lo cual disminuye la potencia de la prueba. Las pruebas de χ^2 son sensibles al tamaño de la muestra; un número reducido de datos limita la capacidad de la prueba para identificar asociaciones significativas, incluso si estas están presentes. Esto es especialmente relevante en nuestro estudio, donde la muestra restringida posiblemente afecta la precisión y robustez del análisis estadístico.
2. Frecuencias bajas y celdas vacías de la tabla de contingencia. En el Nivel B, varias celdas de la tabla de contingencia tienen frecuencias observadas pequeñas (4/16 celdas) o están vacías (7/16 celdas), lo que reduce la estabilidad de la prueba χ^2 y que las discrepancias entre las metodologías se perciban como menos pronunciadas. Esta distribución limita la validez de la prueba, ya que las celdas vacías o con datos muy bajos afectan la comparación entre las frecuencias observadas y las esperadas, disminuyendo el valor de χ^2 y, por tanto, la significación estadística. Aunque no necesariamente implica que no haya una relación entre las metodologías, sino que la evidencia no es lo suficientemente fuerte para declararla significativa en este contexto (197).

A pesar de observarse una disminución en el valor del estadístico χ^2 , que implica la no significación en el Nivel B, debemos valorar como la disminución del tamaño muestral, así como el aumento de celdas vacías o de un recuento inferior a 5 observaciones por celda puede estar sesgando el resultado, es por ello que debemos mirar hacia el tamaño de efecto, la V de Cramer de 0.339 sugiere que existe una asociación de magnitud moderada entre ambas metodologías. Este valor indica que, aunque el análisis no es estadísticamente significativo en este nivel, existen evidencias, desde la perspectiva estadística, para afirmar que existe una consistencia moderada en la clasificación de los TE entre las metodologías. En investigación con tamaños muestrales limitados, la V de Cramer sigue proporcionando una indicación de la fuerza de la relación observada, independientemente de la significación estadística, y permite evaluar la relevancia práctica de los resultados en ausencia de evidencia concluyente (198).

4.4.3.2 Evaluación de la correlación derivada de metodologías comparativas. Nivel B

En esta sección vamos a realizar un análisis de la correlación, de la relación entre las dos variables, los TE de ambas metodologías, en valor cuantitativo. Con ello conoceremos la “fuerza” y la “dirección” de la relación entre ambas variables. Para evaluar si al cambiar un valor de una metodología, el valor en la otra metodología tiende a hacerlo en la misma dirección (correlación positiva) o no (correlación negativa). Se realiza un análisis de correlación utilizando el coeficiente de Spearman (Tabla 21), y cálculo de normalidad (Tabla 22).

Tabla 21. Medidas de Asociación y Significación de Correlación de TE entre Metodologías. Nivel B.

Estadístico	Valor (Est – p-valor)	Resultado	Fuerza
Rho Spearman	.399** (.005)	Alt. Significativo	Baja-Moderada

^{NS} No significativo, † Casi significativo (p<.10), * Significativo (p<.05) y ** Alt. significativo (p<.01)

Tabla 22. Medidas de Asociación y Significación de Correlación de TE entre Metodologías. Nivel B.

Nivel A	TU® - TE (R _{TU})	URT - TE (d Cohen)
W de Shapiro-Wilk	.874	.847
Valor p de Shapiro-Wilk	<.001**	<.001**

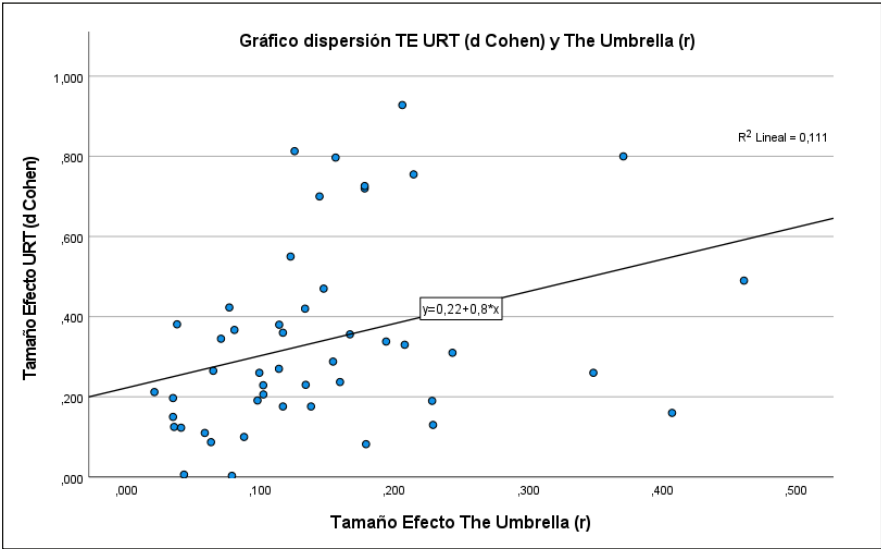
^{NS} No significativo, † Casi significativo (p<.10), * Significativo (p<.05) y ** Alt. significativo (p<.01)

Los análisis de correlación en el Nivel B muestran una ligera disminución respecto al Nivel A, pero mantienen resultados positivos, altamente significativos y de fuerza baja a moderada. El coeficiente Rho de Spearman (Rho = 0.399, p = 0.005) refleja esa correlación. Al ser no paramétricos, son robustos frente a outliers y distribuciones no normales, lo que refuerza la fiabilidad de las correlaciones observadas. Esto subraya la consistencia de R_{TU} en relación con la métrica consolidada d de Cohen.

Para complementar el análisis presentamos la gráfica de dispersión que nos explica la relación entre los valores TE cuantitativos de URT vs TU®, en el eje de las ordenadas están los TE de URT y en el eje de abscisas los TE de TU®, en este caso, además, se añade la regresión lineal entre las puntuaciones que nos habla de los resultados esperados (Figura 21).

URT TE vs TU TE con regresión lineal

Figura 21. Gráfico dispersión de la relación entre el TE de URT (d de Cohen) y The Umbrella (r) con regresión lineal. Nivel B.



En el análisis de correlación de los TE en el Nivel B, se observan ciertos cambios relevantes en comparación con los niveles anteriores. La intersección de la recta de regresión es ahora de 0.22, lo cual representa una reducción en el valor inicial de partida entre las metodologías URT y TU® en los Niveles O y A. Este descenso puede indicar una menor discrepancia en el origen de las mediciones de TE entre ambas metodologías.

La pendiente de la línea de regresión se ha reducido a 0.8, un valor que se aproxima más a 1 en comparación con los niveles previos donde las pendientes eran mayores a 2. Esto sugiere una relación más consistente y menos exagerada entre ambas metodologías, ya que una pendiente más cercana a 1 implica una correspondencia más directa en la escala de los tamaños de efecto estimados por cada metodología. Sin embargo, el coeficiente de determinación R² = 0.11 indica que el modelo de regresión lineal explica solo una pequeña parte de la variabilidad entre las mediciones de tamaño de efecto de URT y TU®, lo que sugiere que existen otros factores no capturados por esta relación lineal que podrían estar influyendo en las diferencias observadas entre ambas metodologías.

4.1.4.3.3 Conclusión Nivel B.

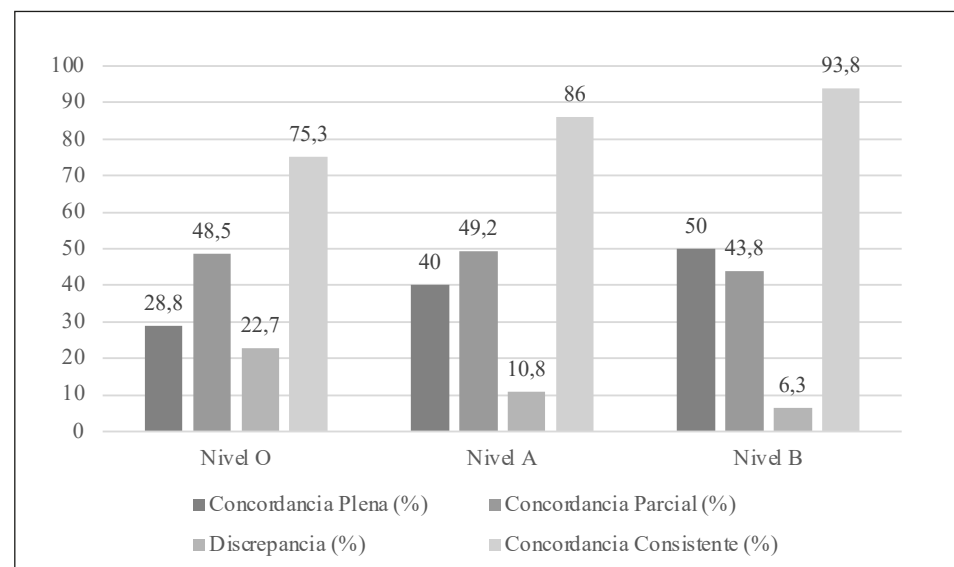
En el Nivel B, tras aplicar ajustes adicionales eliminando métricas sin posibilidad de estandarización completa, la evaluación de concordancia y correlación entre los TE de URT y TU® presenta un total de 48 valores. Este ajuste se traduce en una mayor Concordancia Consistente, que alcanza el 93.8%, la más alta entre los tres niveles, lo que refleja una mayor alineación en la categorización de los TE entre ambas metodologías. La Concordancia Plena se incrementa al 50.0%, en comparación con el 40.0% en el Nivel A y el 28.7% en el Nivel O, mientras que la Discordancia se reduce a un mínimo de 6.3%, mostrando una mejora significativa en la clasificación conjunta de los TE categóricos y una reducción en las discrepancias. Sin embargo, esta muestra más pequeña, junto con las frecuencias bajas y algunas celdas vacías en la tabla de contingencia, ha llevado a una prueba de χ² que no alcanza la significación estadística (p = 0.088), aunque el valor de V de Cramer de 0.339 aún indica una asociación de magnitud moderada.

En cuanto a la correlación cuantitativa, el coeficiente de Spearman ($Rho = 0.399$) sigue siendo positivo y altamente significativo, aunque disminuyen en magnitud con respecto a los Niveles O y A, lo que sugiere una correlación baja-moderada. La gráfica de dispersión muestra que la intersección se reduce aún más a 0.22 y la pendiente a 0.8, un valor cercano a 1, lo cual indica una relación más equilibrada y directa entre ambas metodologías. Sin embargo, el coeficiente de determinación R^2 de 0.11 señala que la variabilidad en los TE de URT explicada por TU[®] es limitada, lo que sugiere la influencia de otros factores no reflejados en esta relación lineal. En conjunto, estos ajustes han mejorado la concordancia y consistencia general entre las metodologías, aunque la pérdida de significación estadística en la prueba de χ^2 limita la robustez de las conclusiones.

4.1.4.4 Consecuencias de las modificaciones en los niveles de análisis: concordancia, intersección, pendiente y coeficiente de determinación

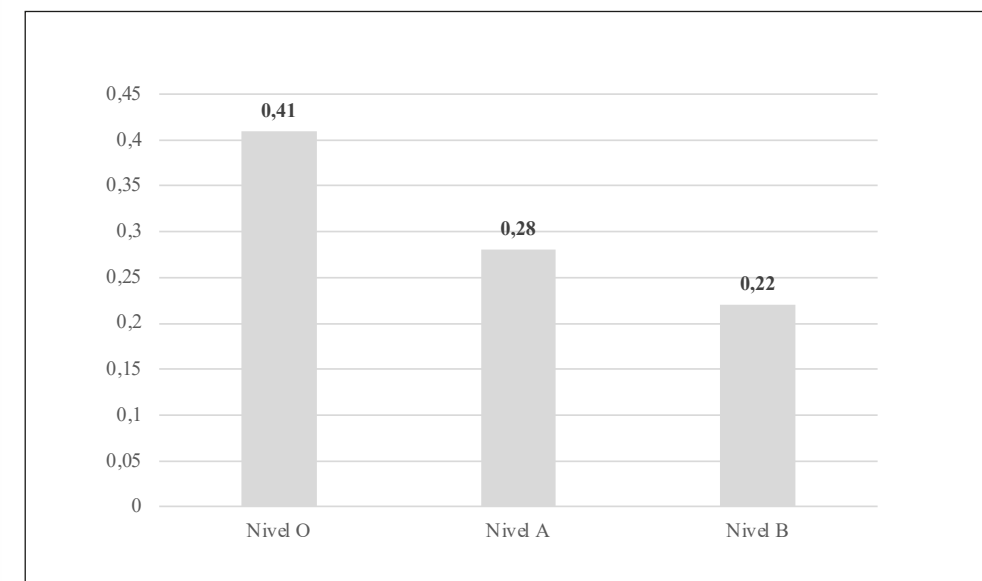
Las modificaciones adoptadas en los Análisis de Nivel A y B han producido modificaciones significativas en el grado de concordancia de la clasificación de categorías de TE entre ambas metodologías. Se representa en el porcentaje de concordancia en las categorías de TE según ambas metodologías en los diferentes niveles de análisis, O, A y B. Se observa como el nivel de Concordancia Plena en la clasificación de categorías de TE es del 28.8% en el Nivel O, aumentando a 40.0% en Nivel A y hasta el 50.0% en el nivel B. De forma contraria el grado de Discordancia es del 22.7% en Nivel O, disminuye a 10.8% en Nivel A, y al 6.3% en Nivel B. Los grados de Concordancia Parcial se muestran estables a lo largo de las modificaciones introducidas en los distintos niveles de análisis, siendo de 48.5% en Nivel O, aumenta discretamente al 49.2% en Nivel A, y al 43.8% en Nivel B. La Concordancia Consistente (suma de Concordancia Plena y Concordancia Parcial), encuentra un aumento progresivo en los distintos niveles de análisis, de manera que en el Nivel O supone el 77.3%, en el Nivel A el 89.2% y en el Nivel B el 93.8% (Figura 22).

Figura 22. Grados de concordancia de TE (%) según nivel de análisis (O, A, B).



La intersección (también llamada ordenada en el origen o constante) es el valor de “y” (TE de URT) cuando “x” = 0 (TE de TU[®]). Representa el punto en el cual la línea de regresión cruza el eje de las ordenadas. En Nivel O, es 0.41, lo que refleja una mayor discrepancia inicial entre URT y TU[®]. Disminuye a 0.28 en Nivel A y finalmente a 0.22 en Nivel B, indicando una mejora progresiva en la alineación inicial entre las metodologías con las modificaciones realizadas al suprimir los valores extremos que no se ajustaban al sentido en que se presentaban en las URTs. El hecho de que la intersección disminuya progresivamente a medida que se ajustan los datos tiene un significado importante. Indica que las dos metodologías, URT y TU[®], son cada vez más coherentes en las estimaciones de los TE, justificándose por la eliminación de valores que no se ajustaban por distintos motivos y que distorsionaban el conjunto de valores. Una intersección menor sugiere que, cuando TU[®] TE es cercano a 0, URT TE también se aproxima más a 0. Esto refleja una reducción en la discrepancia inicial observada entre ambas metodologías, sugiriendo que los ajustes han mejorado la comparabilidad entre las dos herramientas, eliminando sesgos o discrepancias extremas (Figura 23).

Figura 23. Valor intersección en los distintos niveles de análisis (O, A y B).

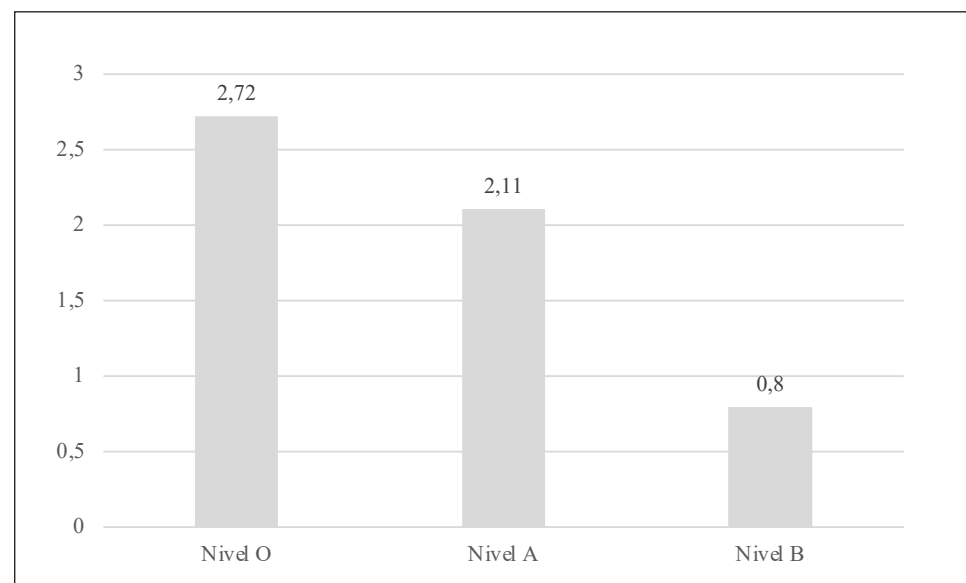


La pendiente de la línea de regresión (recta de ajuste) representa la tasa de cambio entre las dos variables, en este caso, los TE medidos por las metodologías URT y TU[®]. Idealmente, si ambas metodologías fueran completamente equivalentes en sus mediciones, esperaríamos una pendiente cercana a 1, lo que indicaría que un aumento en el TE medido por TU[®] se correspondería de manera directa y proporcional con un aumento similar en el TE medido por URT. En los resultados observados, la pendiente muestra una disminución progresiva a lo largo de los niveles: de 2.72 en el Nivel O, a 2.11 en el Nivel A, y finalmente a 0.8 en el Nivel B. Este cambio implica una tendencia hacia una relación más equilibrada y proporcional entre ambas metodologías a medida que se implementaron los cambios en los distintos niveles. En el Nivel O, con una pendiente de 2.72, se observa que el TE medido

por URT tiende a ser significativamente mayor que el TE medido por TU[®]. Esto sugiere que, para cada unidad de incremento en el TE de TU[®], el TE de URT aumenta en 2.72 unidades, indicando una relación desproporcionada donde URT amplifica los tamaños de efecto en comparación con TU[®]. En el Nivel A, la pendiente disminuye a 2.11, lo que sigue indicando una tendencia de URT a reportar TE mayores, pero con una menor desproporción en comparación con el Nivel O. Este ajuste refleja un cierto grado de alineación adicional entre las metodologías, aunque URT continúa mostrando valores de TE más altos que TU[®]. Finalmente, en el Nivel B, la pendiente baja a 0.8, acercándose al valor ideal de 1, lo que sugiere una relación mucho más equilibrada entre las dos metodologías. Una pendiente de 0.8 implica que, en promedio, por cada unidad de cambio en el TE de TU[®], el TE de URT cambia en 0.8 unidades, indicando que URT tiende a clasificar los tamaños de efecto de forma ligeramente menor en comparación con TU[®], aunque la relación ya no es tan amplificada como en los niveles anteriores (**Figura 24**).

En conjunto, esta disminución progresiva de la pendiente a lo largo de los niveles refleja un ajuste hacia una mayor consistencia y proporcionalidad entre las metodologías una vez ajustados los datos, con una mejora notable en el Nivel B, donde la relación entre los TE de URT y TU[®] es más cercana a la equivalencia ideal.

Figura 24. Valor de pendiente de recta de ajuste en los distintos niveles de análisis (Nivel O, A y B).



Coefficiente de determinación (R^2), es una medida que indica qué proporción de la variabilidad en la variable dependiente (y) puede ser explicada por la variable independiente (x) en un modelo de regresión lineal. Su valor oscila entre 0 y 1: $R^2 = 0$ significa que el modelo no explica ninguna variabilidad de la variable dependiente. $R^2 = 1$ significa que el modelo explica toda la variabilidad. En términos simples, R^2 mide la bondad de ajuste del modelo, es decir, qué tan bien los datos se ajustan a la línea de regresión. En Nivel O, R^2 es 0.200, lo que significa que el 20% de la variabilidad en URT TE puede explicarse por TU[®] TE. En

Nivel A, R^2 baja a 0.135, mientras que en Nivel B disminuye a 0.111. Cuando el R^2 es bajo, indica que la regresión lineal no está capturando bien la relación entre las variables. Esto puede suceder cuando la relación no es lineal o cuando hay otros factores no incluidos en el modelo que influyen en la variable dependiente. Este dato sugiere que los datos podrían estar siguiendo un patrón no lineal (**Figura 25**).

Los valores bajos de R^2 sugieren que la relación entre los tamaños de efecto de URT y TU[®] no es completamente lineal y que existen otros factores que influyen en el TE de URT y que no están representados en el modelo. Esto es comprensible en el contexto de nuestro análisis, principalmente por las diferencias metodológicas entre URT y TU[®]. Aunque ambas metodologías miden tamaños de efecto, la métrica de TU[®], R_{TU} , fue desarrollada específicamente para The Umbrella Collaboration[®], mientras que URT utiliza la métrica consolidada d de Cohen. Estas diferencias en la forma en que cada metodología calcula el TE pueden introducir variabilidad que no se alinea perfectamente de forma lineal entre ambas. El objetivo de utilizar la regresión lineal era explicar de manera sencilla, y conocida en el ámbito de la investigación, qué sucede cuando aumenta el TE en un caso y en el otro. La tendencia observada es clara, ya que los valores de correlación positivos indican una relación significativa y en el mismo sentido. Optamos por la regresión lineal porque facilita la interpretación, mientras que modelos más complejos, como la regresión cuadrática, habrían complicado la comprensión para el lector. Por ejemplo, una fórmula donde el aumento no es lineal, sino que depende de un cuadrado o términos más complejos habría dificultado la interpretación y facilidad de comprensión. Aunque se podrían haber explorado alternativas como la transformación de los valores (por ejemplo, una transformación logística), decidimos priorizar la claridad en la presentación de los resultados. Esto es especialmente relevante en un estudio exploratorio, como es el caso, que busca validar un nuevo procedimiento, donde es crucial que los hallazgos sean accesibles y comprensibles para todos los lectores. En resumen, los valores de R^2 obtenidos en los Niveles O, A y B sugieren que, aunque existe cierta relación entre los tamaños de efecto de ambas metodologías, esta relación es parcial y probablemente influenciada por otros factores no considerados en el modelo lineal.

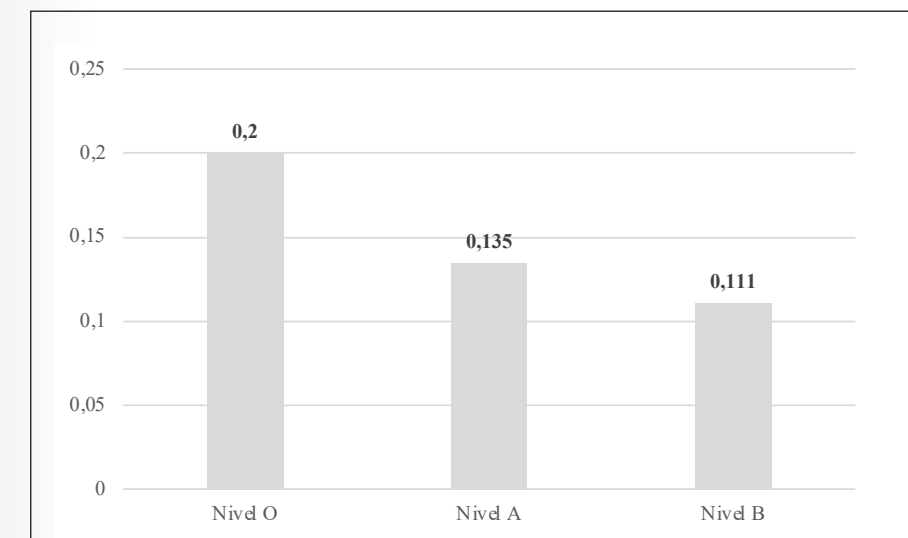


Figura 25.

Valores de coeficiente de determinación en los distintos niveles de análisis (O, A y B).

4.1.5 Direccion y significación estadística de los resultados de interés

En este apartado se presentan los resultados obtenidos de la comparación de la dirección y significación estadística del efecto de los Rdl. Se presentan de forma conjunta la dirección y la significación estadística de los Rdl por el hecho no habitual de presentarse resultados similares en ambos conjuntos de datos.

Presentaremos los resultados de la dirección y significación estadística del efecto de los Rdl y su análisis estadístico. Este primer conjunto de datos presenta los resultados según figuran originalmente en los URT, por lo que se denominarán “originales”. Se realizó un análisis en detalle de los URT en que se presentaban resultados discordantes entre la dirección y la significación de los Rdl según metodologías, que en total eran 13 resultados provenientes de 5 UTR. Este análisis supuso la revisión en detalle de las RS/MA incluidos en las URT que presentaban resultados discordantes, y llevó a proponer una revisión de la dirección y significación estadística de los efectos de 4 Rdl (ANEXO 6). Esto nos permitió generar un nuevo conjunto resultados de dirección y significación estadística de Rdl que denominaremos “revisados”, y presentaremos en segundo lugar de forma descriptiva y con su correspondiente análisis estadístico.

4.1.5.1 Análisis de la dirección y significación estadística de los Rdl con datos “originales”

Presentamos el grado de concordancia que tiene la dirección y la significación estadística del efecto del Rdl que identifica TU® frente la dirección y significación estadística del efecto que presenta en esos mismos Rdl el URT por proyectos y global (Tabla 23).

Se observa que de los 73 Rdl en que ambas metodologías coinciden en identificar y clasificar la dirección y la significación estadística, los resultados muestran un porcentaje de concordancia global de 60 resultados (82.19%), con una horquilla que varía entre 57.14% en el proyecto 4 y un 100% en los proyectos 5, 6 y 8.

Tabla 23. Concordancia de dirección y significación estadística del efecto de Rdl entre metodologías (The Umbrella frente a URT original).

Proyecto	Dirección del efecto Rdl Concordancia TU®/URT (%)	Significación del efecto Rdl Concordancia TU®/URT (%)
Proyecto 1	3/5 (60.00)	3/5 (60.00)
Proyecto 2	26/30 (86.67)	26/30 (86.67)
Proyecto 3	3/5 (60.00)	3/5 (60.00)
Proyecto 4	4/7 (57.14)	4/7 (57.14)
Proyecto 5	4/4 (100.00)	4/4 (100.00)
Proyecto 6	8/8 (100.00)	8/8 (100.00)
Proyecto 7	4/6 (66.67)	4/6 (66.67)
Proyecto 8	8/8 (100.00)	8/8 (100.00)
TOTAL	60/73 (82.19)	60/73 (82.19)

4.1.5.2 Análisis estadístico de la dirección de efecto de Rdl de metodologías comparativas

La tabla de contingencia (Tabla 24) compara como cada metodología clasifica la dirección del efecto del Rdl. De tal manera, que cuando URT clasifica la dirección del efecto en “a favor de la intervención” o “↑ riesgo” en total 56/73 (76.70%), TU® clasifica en esta categoría 67/73 (91.80%), coincidiendo con URT en 55/67 (82.10%), y no coincidiendo en 12/67 (17.90%). Por otro lado, URT clasifica la dirección de efecto como desconocido en 17/73 (23.3%) de los Rdl, mientras que TU® lo hace en 6/73 (8.2%), coincidiendo en 5/6 Rdl con URT (83.3%).

Los resultados de la prueba de la χ^2 muestran una relación altamente significativa entre las clasificaciones de las direcciones de los efectos de los Rdl coincidentes entre ambas metodologías. Como lo muestra el estadístico χ^2 13.19 y valor-p de <0.001. Se acompaña de un tamaño de efecto, cuantificado por la V de Cramer, de 0.425, considerado como una asociación moderada. El nivel de concordancia entre las dos metodologías, empleando el índice Kappa de Cohen (κ), muestra un índice de 0.357, considerado un nivel de concordancia débil.

Tabla 24. Tabla de contingencia. Variables: dirección efecto Rdl URT frente TU®.

Dirección efecto Rdl URT	Dirección efecto Rdl TU®		
	A Favor Intervención / ↑ riesgo	Desconocido	
A Favor Intervención / ↑ riesgo	82.1% (55)	16.7% (1)	76.7% (56)
Desconocido	17.9% (12)	83.3% (5)	23.3% (17)
	91.8% (67)	8.2% (6)	(73)
$\chi^2 = 13.19^{**}$ (<0.001), V = 0.425 (Moderado), $\kappa = 0.357$ (Débil)			
Notas ^{NS} No significativo, † Casi significativo ($p < .10$), * Significativo ($p < .05$) y ** Altamente significativo ($p < .01$) Test estadístico/contraste: χ^2 - Chi Cuadrado. Tamaño o magnitud del efecto: V – V de Cramer, κ - Índice Kappa de Cohen			

4.1.5.3 Análisis estadístico de la significación estadística del efecto de Rdl de metodologías comparativas

La tabla de contingencia (Tabla 25) compara como cada metodología clasifica la significación estadística del efecto del Rdl. De tal manera, que cuando URT clasifica la significación estadística del efecto en “Significativo” en total 56/73 (76.70%), TU® clasifica en esta categoría 67/73 (91.80%), coincidiendo con URT en 55/67 (82.10%), y no coincidiendo en 12/67 (17.90%). Por otro lado, URT clasifica la significación estadística de efecto como “No-significativo” en 17/73 (23.3%) de los Rdl, mientras que TU® lo hace en 6/73 (8.2%), coincidiendo en 5/6 Rdl con URT (83.3%).

Los resultados de la prueba de la χ^2 muestran una relación altamente significativa entre las clasificaciones de la significación estadística de los efectos de los Rdl coincidentes entre ambas metodologías. Como lo muestra el estadístico χ^2 13.19 y p-valor de < 0.001. Se acompaña de un tamaño de efecto, cuantificado por la V de Cramer, de 0.425, considerado como una asociación moderada. El nivel de concordancia entre las dos metodologías, empleando el índice Kappa de Cohen (κ), muestra un índice de 0.357, considerado un nivel de concordancia débil.

Tabla 25. Tabla de contingencia. Variables: significación estadística del efecto de Rdl URT frente TU.

Significación estadística del efecto Rdl URT	Significación estadística efecto Rdl TU®		
	Significativo	No significativo	
Significativo	82.1% (55)	16.7% (1)	76.7% (56)
No significativo	17.9% (12)	83.3% (5)	23.3% (17)
	91.8% (67)	8.2% (6)	(73)
$\chi^2 = 13.19^{**}$ (<0.001), V = 0.425 (Moderado), $\kappa = 0.357$ (Débil)			
Notas ^{NS} No significativo, † Casi significativo ($p < .10$), * Significativo ($p < .05$) y ** Altamente significativo ($p < .01$) Test estadístico/contraste: χ^2 - Chi Cuadrado. Tamaño o magnitud del efecto: V – V de Cramer, κ - Índice Kappa de Cohen			

4.1.5.4 Análisis de la dirección y significación estadística de los Rdl con datos “revisados”

Se identificaron 5 proyectos donde existen discrepancias entre las metodologías en la dirección y la significación estadística del efecto de los Rdl según los datos originales (**ANEXO 6**). Son los proyectos 1, 2, 3, 4 y 7. En total son 13 resultados discordantes. Después de un análisis detallado que incluyó un análisis de las RS/MA incluidas en los URT, se objetiva que en los proyectos 1, 3 y 4 existe una clasificación en los URT no ajustado a lo referido en las RS/MA en cuanto a la dirección y significación estadística del efecto de 4 Rdl. Por lo que en la revisión propuesta el grado de concordancia en el proyecto 1 pasa de 3/5 a 4/5, en el proyecto 3 de 3/5 a 5/5 y en el proyecto 4 de 4/7 a 5/7. No encontrándose desajustes en la clasificación en los proyectos 2 y 7.

Con estas modificaciones propuestas con la revisión de datos se observa que de los 73 Rdl en que ambas metodologías coinciden en identificar y clasificar la dirección y la significación estadística, los resultados muestran un porcentaje de concordancia global de 64 resultados (87.67%), con una horquilla que varía entre 66.67% en el proyecto 7 y un 100% en los proyectos 3, 5, 6 y 8 (**Tabla 26**).

Tabla 26. Concordancia de dirección y significación estadística del efecto de Rdl entre metodologías (The Umbrella frente a URT revisado).

Proyecto	Dirección del efecto Rdl Concordancia TU®/URT (%)	Significación del efecto Rdl Concordancia TU®/URT (%)
Proyecto 1	4/5 (80.00)	4/5 (80.00)
Proyecto 2	26/30 (86.67)	26/30 (86.67)
Proyecto 3	5/5 (100.00)	5/5 (100.00)
Proyecto 4	5/7 (71.42)	5/7 (71.42)
Proyecto 5	4/4 (100.00)	4/4 (100.00)
Proyecto 6	8/8 (100.00)	8/8 (100.00)
Proyecto 7	4/6 (66.67)	4/6 (66.67)
Proyecto 8	8/8 (100.00)	8/8 (100.00)
TOTAL	64/73 (87.67)	64/73 (87.67)

4.1.5.5 Análisis estadístico de la dirección de efecto de Rdl de metodologías comparativas

La tabla de contingencia **Tabla 27**, compara como cada metodología clasifica la dirección del efecto del Rdl. En este caso el conjunto de datos de los URT son la revisión propuesta tras el análisis detallado en el **ANEXO 6**. De tal manera, que cuando URT clasifica la dirección del efecto en “a favor de la intervención” o “↑ riesgo” en total 58/73 (79.50%), TU® clasifica en esta categoría 67/73 (91.80%), coincidiendo con URT en 58/67 (86.60%), y no coincidiendo en 9/67 (13.40%). Por otro lado, URT clasifica la dirección de efecto como desconocido en 15/73 (20.50%) de los Rdl, mientras que TU® lo hace en 6/73 (8.2%), coincidiendo en 6/6 Rdl con URT (100.00%).

Los resultados de la prueba de la χ^2 muestran una relación altamente significativa entre las clasificaciones de las direcciones de los efectos de los Rdl coincidentes entre ambas metodologías. Como lo muestra el estadístico χ^2 25.28 y valor-p de <0.001. Se acompaña de un tamaño de efecto, cuantificado por la V de Cramer, de 0.588, considerado como una asociación grande. El nivel de concordancia entre las dos metodologías, empleando el índice Kappa de Cohen (κ), muestra un índice de 0.514, considerado un nivel de concordancia moderado.

Tabla 27. Tabla de contingencia. Variables: dirección efecto Rdl URT frente TU.

Dirección efecto Rdl URT	Dirección efecto Rdl TU®		
	A Favor Intervención / ↑ riesgo	Desconocido	
A Favor Intervención / ↑ riesgo	86.6% (58)	-	79.5% (58)
Desconocido	13.4% (9)	100.0% (6)	20.5% (15)
	91.8% (67)	8.2% (6)	(73)
$\chi^2 = 25.28^{**}$ (<.001), V = .588 (Grande), $\kappa = .514$ (Moderado)			
Notas ^{NS} No significativo, † Casi significativo (p<.10), * Significativo (p<.05) y ** Altamente significativo (p<.01) Test estadístico/contraste: χ^2 - Chi Cuadrado. Tamaño o magnitud del efecto: V – V de Cramer, κ - Índice Kappa de Cohen			

4.1.5.6 Análisis estadístico de la significación estadística del efecto de Rdl de metodologías comparativas

La tabla de contingencia **Tabla 28**, compara como cada metodología clasifica la significación estadística del efecto del Rdl. En este caso el conjunto de datos de los URT son la revisión propuesta tras el análisis detallado en el **ANEXO 6**. De tal manera, que cuando URT clasifica la dirección del efecto en “significativo” en total 58/73 (79.50%), TU® clasifica en esta categoría 67/73 (91.80%), coincidiendo con URT en 58/67 (86.60%), y no coincidiendo en 9/67 (13.40%). Por otro lado, URT clasifica la significación estadística de efecto como no-significativo en 15/73 (20.50%) de los Rdl, mientras que TU® lo hace en 6/73 (8.2%), coincidiendo en 6/6 Rdl con URT (100.00%).

Los resultados de la prueba de la χ^2 muestran una relación altamente significativa entre las clasificaciones de las direcciones de los efectos de los Rdl coincidentes entre ambas metodologías. Como lo muestra el estadístico χ^2 25.28 y valor-p de <0.001. Se acompaña de un tamaño de efecto, cuantificado por la V de Cramer, de 0.588, considerado como una asociación grande. El nivel de concordancia entre las dos metodologías, empleando el índice Kappa de Cohen (κ), muestra un índice de 0.514, considerado un nivel de concordancia moderado.

Tabla 28. Tabla de contingencia. Variables: significación estadística del efecto de Rdl URT (revisados) frente TU®.

Significación estadística del efecto Rdl URT	Significación estadística efecto Rdl TU®		
	Significativo	No significativo	
Significativo	86.6% (58)	-	79.5% (58)
No significativo	13.4% (9)	100% (6)	20.5% (15)
	91.8% (67)	8.2% (6)	(73)
$\chi^2 = 25.28^{**}$ (<.001), V = .588 (Grande), $\kappa = .514$ (Moderado)			
Notas ^{NS} No significativo, † Casi significativo (p<.10), * Significativo (p<.05) y ** Altamente significativo (p<.01) Test estadístico/contraste: χ^2 - Chi Cuadrado. Tamaño o magnitud del efecto: V – V de Cramer, κ - Índice Kappa de Cohen			

4.1.6 Resultados concordancia de niveles de certeza de Rdl entre metodologías

El objetivo de este análisis es evaluar la concordancia entre dos metodologías para valorar la certeza de los Rdl en estudios de síntesis terciaria: URT que utiliza el método GRADE, y la metodología TU®, emplea un análisis automatizado de sentimiento aplicado a los *abstracts* de las RS/MA. Con este análisis, buscamos determinar si la evaluación experimental de certeza realizada por TU® se comporta de manera similar a la clasificación establecida por GRADE.

Seleccionamos de entre los 8 URT utilizados en este trabajo, solo los URT que utilizan GRADE para valorar la certeza. Son los 5 URT de los proyectos 3 (167) Shen, 2023; 4 (161) Conneely, 2022; 5 (169) Gazzaniga, 2023; 7 (171) Veronese, 2021 (1) y 8 (172) Veronese, 2021 (2). Finalmente quedan 25 valores de certeza para poder realizar la comparación.

Se muestra en la **Tabla 29** el listado de valores de certeza de Rdl transformados, permitiendo una comparación directa entre las valoraciones de ambas metodologías.

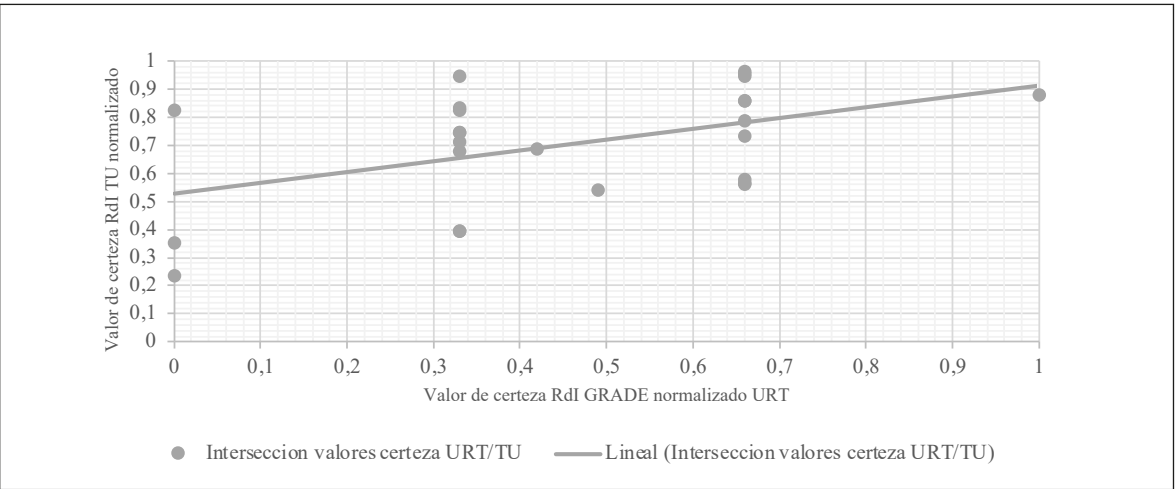
Tabla 29. Listado de valores de certeza transformados de los Rdl de ambas metodologías

Proyecto	Nº Rdl	Valor certeza transformado GRADE de Rdl URT	Valor de certeza transformado de Rdl TU®
3	F092	0.66	0.8584
	F093	1	0.8793
	F097	0.66	0.9468
4	F098	0.33	0.9468
	F131	0.33	0.8264
	F132	0.33	0.8314

Proyecto	Nº Rdl	Valor certeza transformado GRADE de Rdl URT	Valor de certeza transformado de Rdl TU®
	F133	0.33	0.7433
	F135	0.42	0.6869
	F136	0.49	0.5385
	F137	0.33	0.7469
	F138	0.33	0.7107
5	F153	0.66	0.5795
	F154	0.66	0.7307
	F155	0.33	0.6784
	F156	0.66	0.7867
7	F330	0	0.8243
	F332	0	0.2366
8	F377	0.33	0.3944
	F378	0.33	0.3944
	F379	0.66	0.9550
	F380	0.66	0.9623
	F381	0.66	0.5598
	F382	0.66	0.5674
	F383	0	0.3509
	F384	0.66	0.8592

Para visualizar la concordancia entre las dos metodologías, utilizamos una gráfica de dispersión (**Figura 26**) que permite ver la relación directa entre los valores de certeza transformados de los Rdl asignados por los URT con GRADE y por TU®. En el eje de las abscisas se representa los valores de certeza reclasificados de los Rdl por parte de los URT con GRADE. En el eje de las ordenadas se representa los valores de certeza reclasificados de los Rdl según TU®.

Figura 26. Gráfica dispersión valores transformados certeza Rdl (N=25).



En el análisis de correlación realizado, hemos utilizado el coeficiente de correlación de Spearman para evaluar la relación entre los valores de certeza transformados de los Rdl según GRADE para los URT y según la certeza otorgada por TU®.

Como se observa en la **Tabla 30**, el coeficiente de Spearman presenta un valor positivo y significativo, lo que indica una relación directa entre las dos metodologías. El coeficiente de Spearman de 0.446, con un valor-p de 0.025, confirmando también una correlación significativa de fuerza moderada. Cálculo distribución normalidad variables certeza Rdl transformados de ambas metodologías (**Tabla 31**).

Estos resultados sugieren que existe una relación positiva y estadísticamente significativa entre las valoraciones de certeza proporcionadas por URT y TU®, lo que refuerza la concordancia observada entre ambas metodologías.

Tabla 30. Análisis de correlación. Variables: valores de certeza Rdl transformados TU® frente URT.

Coeficiente Spearman			
N	Coeficiente	Valor-p	Resultado
25	0.446*	0.025	Significativo

Tabla 31. Cálculo normalidad variables certeza transformados Rdl.

	Valor certeza GRADE transformado de Rdl URT	Valor de certeza transformado de Rdl TU®
W de Shapiro-Wilk	0.874	0.928
Valor p de Shapiro-Wilk	0.005	0.049

La gráfica de dispersión revela una tendencia positiva, donde los aumentos en los niveles de certeza transformado de GRADE de los URT están generalmente asociados con incrementos en los niveles de certeza transformados de TU®. Los puntos de datos están distribuidos a lo largo del rango, pero se observa una agrupación notable a lo largo de una trayectoria que asciende desde la esquina inferior izquierda hacia la superior derecha del gráfico. Esta tendencia es indicativa de una correlación directa, alineándose con los resultados cuantitativos reportados en la gráfica de dispersión, donde el coeficiente de correlación de Spearman (0.469) sugiere una relación positiva estadísticamente significativa de fuerza moderada.

4.2 RESULTADOS ENCUESTA DE SATISFACCIÓN Y UTILIDAD CLÍNICA DE LA METODOLOGÍA Y APLICACIÓN The Umbrella Collaboration® (ESTUDIO-II)

El proyecto y especialmente la aplicabilidad clínica, comprensión y utilidad de la metodología y aplicación TU® de síntesis terciaria de evidencia, fue presentado en 5 servicios de Geriátría

de Hospitales Universitarios y acreditados en formación MIR de Geriátría de la Comunidad de Madrid:

- Hospital Universitario Puerta de Hierro- Majadahonda, el 17 de mayo de 2024.
- Hospital Universitario La Paz, el 13 de junio de 2024.
- Hospital Universitario Clínico San Carlos, el 19 de junio de 2024.
- Hospital Universitario Ramón y Cajal, el 3 de julio de 2024.
- Hospital Universitario Central de la Cruz Roja “San José y Santa Adela” el 5 de julio 2024.

Al final de la presentación se realizó una encuesta online voluntaria y anónima a los profesionales sanitarios presentes, personal del propio servicio y profesionales visitantes, sobre la comprensión y utilidad de la herramienta y la forma de presentación de los resultados de TU®. A continuación, se exponen los resultados de dicha encuesta.

4.2.1 Datos descriptivos del total de la muestra

La encuesta fue cumplimentada por 121 profesionales sanitarios presentes durante la presentación de TU® en los 5 servicios de Geriátría de Hospitales Universitarios de la Comunidad Autónoma de Madrid descritos anteriormente.

La **Tabla 32** muestra la descripción de las características de los profesionales, con predominio de médicos (85.96%) y dentro de éstos, mayoritariamente fueron geriatras (89.44%) (**Tabla 33**). Nuestra muestra coincide con la feminización de la medicina y especialmente de la especialidad de la geriatría, ya que un 79.33% fueron mujeres. Por grupos de edad predominaron de jóvenes, menores de 33 años (56.2%), con un pequeño grupo de mayor edad (55-65 años) (13.2%) (**Figura 27**).

Tabla 32. Características de la muestra.

N 121 (%)	
Médico	104 (85.96%)
Geriatra	93 (76.8%)
Otros	11 (9.16%)
Enfermero	17 (14.04%)
Género	
Mujer	96 (79.33%)
Hombre	25 (20.67%)
Subgrupos de edad	
< 33 años	68 (56.2%)
33-43 años	26 (21.5%)
44-54 años	11 (9.1%)
55-65 años	16 (13.2%)

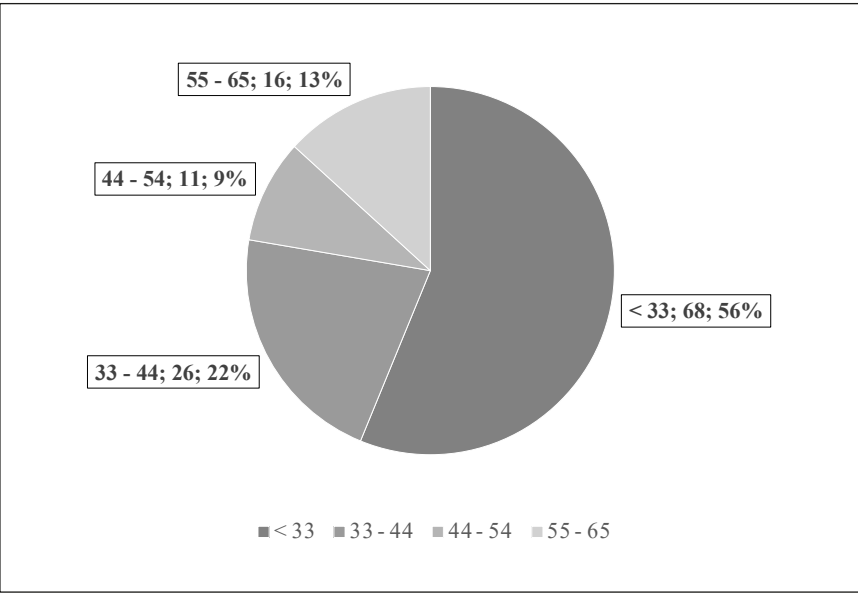
N 121 (%)	
Hospital	
H U Puerta de Hierro-Majadahonda	13 (10.7%)
H U La Paz	25 (20.7%)
H U Clínico San Carlos	21 (17.4%)
H U Ramón y Cajal	22 (18.2%)
H U Central de la Cruz Roja “San José y Santa Adela”	40 (33.0%)

*H U: Hospital Universitario

Tabla 33. Especialidad de los médicos participantes.

N 104 (%)	
Geriatras	92 (89.44%)
Medicina Familiar y Comunitaria	5 (4.80%)
Medicina Interna	2 (1.92%)
Traumatología y Ortopedia	3 (2.88%)
Urología	1 (0.96%)

Figura 27. Distribución grupos de edades (N=121).



Por último, en relación con los hospitales la muestra fue muy semejante en el HU la Paz, HU Ramón y Cajal y HU Clínico San Carlos (entre 17-21%), siendo más reducida en el HU Puerta de Hierro Majadahonda (10.7%) y la mayor en el HU Central de la Cruz Roja (33%), todo ello en relación con el tamaño y recursos humanos del Servicio de Geriátría que ha participado (**Figura 28**).

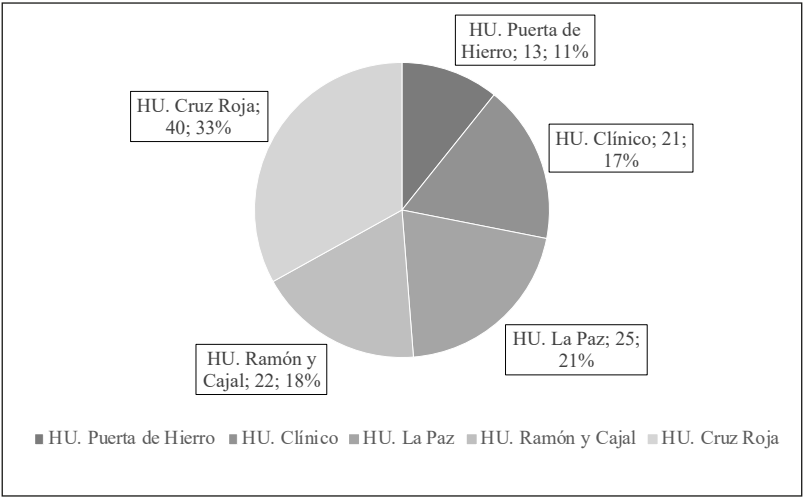


Figura 28. Participantes por hospitales (N; %).

4.2.2 Datos de la valoración global de la utilidad clínica de la metodología y del formato de presentacion de resultados de TU®

4.2.2.1 Análisis de las puntuaciones medias, desviaciones estándar por ítem y promedio de la encuesta TU®.

Se presentan los resultados, según la muestra total de los participantes en la encuesta, de la puntuación media recibido por cada uno de los 10 ítems (P.7 – P.16) de la encuesta y la media global (Tabla 34). Los datos empleados para esta valoración son la puntuación media obtenida para cada ítem, con su correspondiente desviación estándar (DE), se valora en una escala de 1 a 5, donde 1 indica total desacuerdo y 5 indica total acuerdo con la afirmación de cada ítem. La media global de todas las respuestas sobre la utilidad de TU® es de 4.35 ± 0.860. El análisis detallado muestra que la puntuación mínima corresponde al ítem P. 7 (Facilidad comprensión) con un valor de 4.27 ± 0,876, mientras que la puntuación máxima se observa en el ítem P. 16 (Satisfacción) con un valor de 4.42 ± 0.854.

Tabla 34. Valoración media y DE de cada ítem de la encuesta sobre la utilidad de TU®.

Código Pregunta. Descripción	Media	DE
P7. Facilidad comprensión	4.27	0.876
P8. Bubble plots ayuda identificar resultados	4.38	0.799
P9. Representación gráfica mejora evaluación	4.32	0.819
P10. Acceso a datos mejora evaluación	4.36	0.806
P11. Interpretación más rápida de resultados	4.36	0.874
P12. Es útil en toma de decisiones	4.36	0.876
P13. Facilita más la exploración de los datos	4.36	0.855
P14. Evolución Histórica es útil	4.36	0.912
P15. Requiere menos conocimientos estadísticos	4.31	0.930
P16. Satisfacción con metodología	4.42	0.854
Promedio	4.35	0.8601

4.2.2.2 Análisis descriptivo de frecuencias de puntuación por ítem de la encuesta TU® según la muestra total de participantes

Presentamos los resultados globales de puntuaciones recibidas para los 10 ítems de la encuesta sobre la utilidad de la herramienta TU®, de manera individual (P. 7, P. 8, P. 9, P. 10, P. 11, P. 12, P. 13, P. 14, P. 15 y P. 16), con las respuestas de toda la población de la encuesta en la Tabla 35. En ella se resumen las frecuencias de puntuaciones recibidas por cada ítem de la muestra total de encuestados.

Los resultados indican que la mayoría de las respuestas se concentran en las categorías de valoración 4 y valoración 5. Los resultados de cada pregunta fueron los siguientes:

- La P.7 (Facilidad comprensión) recibió 58 respuestas en valoración 5 y 44 en valoración 4, representando el 47.93% y el 36.36% del total respectivamente. En contraste, solo tres respuestas correspondieron a valoración 1 (2.48%).
- La P. 8 (Bubble plots ayuda a identificar resultados) recibió 63 respuestas en valoración 5 (52.07%) y 46 en valoración 4 (38.02%), mientras que recibió 3 respuestas en valoración 1 (2.48%) y 1 en valoración 2 (0.83%).
- La P. 9 (Representación gráfica mejora evaluación) recibió 59 respuestas en valoración 5 (48.76%) y 47 en valoración 4 (38.02%), recibiendo 2 en valoración 1 (1.65%) y 1 en valoración 2 (0.83%).
- La P. 10 (Acceso a datos mejora evaluación) recibió 62 valoraciones 5 (51.24%) y 46 valoraciones 4 (38.02%), recibiendo 2 valoraciones 1 (1.65%) y 1 valoración 2 (0.83%).
- La P. 11 (Interpretación más rápida de resultados) recibió 66 valoraciones 5 (54.55%) y 38 valoraciones 4 (31.4%), recibiendo 3 valoraciones 1 (2.48%) y 0 valoraciones 2.
- La P. 12 (Es útil en la toma decisiones) recibió 67 valoraciones 5 (55.37%) y 37 valoraciones 4 (30.58%), recibiendo 3 valoraciones 1 (2.48%) y 0 valoraciones 2.
- La P. 13 (Facilita más la exploración de los datos) recibió 64 valoraciones 5 (52.89%) y 42 valoraciones 4 (34.71%), recibiendo 3 valoraciones 1 (2.48%) y 0 valoraciones 2.
- La P. 14 (Evolución histórica es útil) recibió 71 valoraciones 5 (58.68%) y 28 valoraciones 4 (23.14%), recibiendo 2 valoraciones 1 y 2 (1.65%).
- La P. 15 (Requiere menos conocimientos estadísticos) recibió 67 valoraciones 5 (55.37%) y 31 valoraciones 4 (25.62%), recibiendo 2 valoraciones 1 (1.65%) y 3 valoraciones 2 (2.48%).
- La P. 16 (Satisfacción con metodología) recibió 71 valoraciones 5 (58.68%) y 36 valoraciones 4 (29.75%).

Tabla 35. Frecuencia global de las valoraciones de cada ítem de la encuesta.

Pregunta (N.º)	Valoración 1 (%)	Valoración 2 (%)	Valoración 3 (%)	Valoración 4 (%)	Valoración 5 (%)
P. 7	3 (2.48)	0 (0.00)	16 (13.22)	44 (36.36)	58 (47.93)
P. 8	2 (1.65)	1 (0.83)	9 (7.44)	46 (38.02)	63 (52.07)
P. 9	2 (1.65)	1 (0.83)	12 (9.92)	47 (38.84)	59 (48.76)
P. 10	2 (1.65)	1 (0.83)	10 (8.26)	46 (38.02)	62 (51.24)
P. 11	3 (2.48)	0 (0.00)	14 (11.57)	38 (31.40)	66 (54.55)
P. 12	3 (2.48)	0 (0.00)	14 (11.57)	37 (30.58)	67 (52.37)
P. 13	3 (2.48)	0 (0.00)	12 (9.92)	42 (34.71)	64 (52.89)
P. 14	2 (1.65)	2 (1.65)	18 (14.88)	28 (23.14)	71 (58.68)
P. 15	2 (1.65)	3 (2.48)	18 (14.88)	31 (25.62)	67 (55.37)
P. 16	3 (2.48)	0 (0.00)	11 (9.09)	36 (29.75)	71 (58.68)
Total	25 (2.07)	8 (0.66)	134 (11.07)	395 (32.64)	648 (53.55)

En cuanto a los resultados totales, es importante destacar que, entre las valoraciones máximas, valoración 4 y valoración 5, se acumula un porcentaje significativo de respuestas. En concreto, de las 1210 respuestas totales registradas en la encuesta, 395 (32.64%) correspondieron a valoración 4 y 648 (53.55%) a valoración 5. Esto significa que más del 85% de las respuestas globales (86.19%) se encuentran en las categorías más altas de satisfacción y acuerdo con los ítems evaluados.

4.2.3 Análisis por categoría profesional: médicos y enfermeros

4.2.3.1 Análisis de las puntuaciones medias, DE y significación de la encuesta de comprensión y utilidad TU® entre categorías profesionales (médicos y enfermeros)

La media de la valoración de la utilidad de TU® medida mediante la encuesta analizada es de 4.35 +/- 0,860. Se presenta un desglose de las puntuaciones medias, DE y significación estadística de los 10 ítems de la encuesta de utilidad de TU®, categorizados por los dos grupos de profesionales sanitarios participantes: médicos y enfermeros (Tabla 36).

En la categoría de médicos (N=104), El promedio de todas las respuestas de este grupo es de 4.36 ± 0.878, discretamente superior a la media de valoración de la utilidad de TU® por todos los participantes. La media de las puntuaciones por ítem oscila entre 4.29 ± 0.878 para la P. 7 (Facilita la comprensión) y 4.43 ± 0.879 para la P. 16 (Satisfacción metodología).

En la categoría profesional de enfermeros (N=17), la media de valoración global de la utilidad de TU® en este grupo es ligeramente inferior, con un valor de 4.26 ± 0.756. Las puntuaciones medias por ítem varían entre 4.18 ± 0.883 para el ítem P. 7 (Facilita la comprensión), la P. 13 (Facilidad explora datos) con una media y DE de 4.18 ± 0.809 y la P.

15 (Menos conocimientos) 4.18 ± 0.809, y en el rango superior se encuentran las medias a las P. 9 (Representación gráfica), P. 10 (Acceso a datos), P. 12 (Toma de decisiones) y P. 16 (Satisfacción) con medias y DE de 4.35 ± 0.606, 4.35 ± 0.606, 4.35 ± 0.786 y 4.35 ± 0,702, respectivamente.

Aunque las puntuaciones son comparativamente similares entre los dos grupos profesionales, las desviaciones estándar son ligeramente superiores en varios ítems entre los médicos con un promedio de DE en el grupo de los médicos de 0.878 (IC: 0.830 – 0.950), mientras que en el grupo de enfermeros el promedio de DE es de 0.756 (IC: 0.562 – 0.951), reflejando una variabilidad algo más baja en las respuestas de los enfermeros.

En ambas categorías profesionales, la pregunta que obtuvo la mayor puntuación fue la P. 16 (Satisfacción), mientras que la P. 7 (Facilita la comprensión) obtuvo la menor puntuación.

Para todos los ítems evaluados, los resultados no mostraron diferencias estadísticamente significativas entre médicos y enfermeros. El análisis de significación estadística nos muestra que, aunque las puntuaciones medias de los ítems varían ligeramente entre médicos y enfermeros, estas diferencias no son estadísticamente significativas (Tabla 36).

Tabla 36. Valoración media, DE y significación a cada ítem de por categoría profesional.

	Médico	Enfermero	Valor de p
	N=104 (86%)	N=17 (14%)	
Facilidad comprensión (P.7)	4.29 ± 0.878	4.18 ± 0.883	0.313
Bubble plots ayuda identificar resultados (P.8)	4.40 ± 0.830	4.24 ± 0.562	0.211
Representación gráfica mejora evaluación (P.9)	4.32 ± 0.851	4.35 ± 0.606	0.434
Acceso a datos mejora evaluación (P.10)	4.37 ± 0.837	4.35 ± 0.606	0.477
Interpretación más rápida de resultados (P.11)	4.37 ± 0.882	4.29 ± 0.849	0.378
Útil toma en toma de decisiones (P.12)	4.37 ± 0.893	4.35 ± 0.786	0.478
Facilita más la exploración de los datos (P.13)	4.35 ± 0.879	4.18 ± 0.809	0.385
Evolución Histórica es útil (P.14)	4.38 ± 0.906	4.18 ± 0.951	0.193
Requiere menos conocimientos estadísticos (P.15)	4.33 ± 0.950	4.18 ± 0.809	0.269
Satisfacción con metodología (P.16)	4.43 ± 0.879	4.35 ± 0.702	0.361
Promedio	4.36 ± 0.878	4.26 ± 0.756	

En resumen, ambas categorías profesionales evaluaron positivamente la utilidad de TU®, con medias de puntuación que reflejan una valoración alta en todos los ítems de la encuesta. La consistencia en las respuestas y la similitud en las puntuaciones medias y las DE entre ambos grupos profesionales sugieren una percepción compartida de la utilidad de la metodología propuesta por TU®.

4.2.3.2 Análisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU® según las dos categorías profesionales participantes (médico / enfermero)

Se realizó un análisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU®, diferenciando entre las dos categorías profesionales participantes (médicos y enfermeros). Los resultados incluyen datos sobre las frecuencias de valoración (1-5) otorgadas a cada ítem de forma individual y global, así como la identificación de los ítems mejor valorados por cada categoría profesional. Para consultar el detalle de este análisis, se remite al lector al **ANEXO 7**, donde se presentan tablas y figuras que resumen los resultados observados.

4.2.4 Análisis por género

4.2.4.1 Análisis de las puntuaciones medias, DE y significación de la encuesta de utilidad TU® según género (mujer / hombre)

Se presenta un desglose de las puntuaciones medias, DE y significación estadística de los 10 ítems de la encuesta de utilidad de TU®, categorizados por género: femenino y masculino (**Tabla 37**).

Tabla 37. Valoración media, DE y significación de cada ítem por género.

Pregunta	Mujer	Hombre	Valor de p
	N= 96 (89%)	N= 25 (21%)	
Facilidad compresión (P.7)	4.25 ± 0.854	4.38 ± 0.970	0.263
Bubble plots ayuda identificar resultados (P.8)	4.37 ± 0.754	4.42 ± 0.974	0.402
Representación gráfica mejora evaluación (P.9)	4.31 ± 0.769	4.38 ± 1.013	0.363
Acceso a datos mejora evaluación (P.10)	4.33 ± 0.774	4.50 ± 0.933	0.178
Interpretación más rápida de resultados (P.11)	4.32 ± 0.848	4.50 ± 0.978	0.184
Útil toma en toma de decisiones (P.12)	4.34 ± 0.853	4.46 ± 0.977	0.278
Facilita más la exploración de los datos (P.13)	4.32 ± 0.823	4.50 ± 0.978	0.178
Evolución Histórica es útil (P.14)	4.32 ± 0.884	4.50 ± 1.022	0.194
Requiere menos conocimientos estadísticos (P.15)	4.28 ± 0.944	4.42 ± 0.881	0.258
Satisfacción con metodología (P.16)	4.37 ± 0.846	4.63 ± 0.875	0.097
Promedio	4.31 ± 0.834	4.46 ± 0.960	

En la categoría de mujeres (N=96), el promedio de todas las respuestas de este grupo es de 4.31 ± 0.834, discretamente inferior a la valoración media global de 4.35 +/- 0.860. La media de las puntuaciones por ítem oscila entre 4.25 ± 0.854 para el ítem P. 7 (Facilidad comprensión) y 4.37 ± 0.846 para el ítem P. 16 (Satisfacción con metodología).

En la categoría de hombres (N=25), la media global de las respuestas en este grupo es ligeramente superior, con un valor de 4.46 ± 0.960, discretamente superior a la valoración media global de 4.35 +/- 0.860. Las puntuaciones medias varían entre 4.38 ± 0.970 para

el ítem P. 7 (Facilidad comprensión), y en el rango superior se encuentra el ítem P. 16 (Satisfacción con metodología) con una media y DE de 4.63 ± 0.875.

En ambas categorías, la pregunta que obtuvo la mayor puntuación fue la P. 16 (Satisfacción con metodología), mientras que la P. 7 (Facilidad comprensión) obtuvo la menor puntuación.

En resumen, tanto los hombres como las mujeres evaluaron positivamente la utilidad de TU®, con medias de puntuación que reflejan una valoración alta en todos los ítems de la encuesta. La consistencia en las respuestas y la similitud en las puntuaciones medias y las DE entre ambos grupos de género sugieren una percepción compartida de la utilidad de la metodología propuesta por TU®.

4.2.4.2 Análisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU® según género

Se llevó a cabo un análisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU®, según género (mujeres y hombres). Los resultados incluyen datos sobre las frecuencias de valoración (1-5) otorgadas a cada ítem de forma individual y global, así como la identificación de los ítems mejor valorados por cada género. Para consultar el detalle de este análisis, se remite al lector al **ANEXO 8**, donde se presentan tablas y figuras que resumen los resultados observados.

4.2.5 Análisis por grupos de edad

4.2.5.1 Análisis de las puntuaciones medias, DE y significación de la encuesta de utilidad TU® por grupos de edad

Se detalla la puntuación media de cada ítem y de forma global para el ítem de los participantes de todos los grupos de edad de la encuesta de utilidad de TU® recibió por rango de edad junto con DE y el valor de significación calculado mediante análisis de varianza (ANOVA) (**Tabla 38**).

Tabla 38. Puntuación media, DE y valor p de cada ítem de la encuesta de utilidad de TU® por grupos de edad.

Pregunta	<33 años	33-43 años	44-54 años	55-65 años	Total	Valor-p
	N= 68 (56%)	N= 26 (21%)	N= 11 (9%)	N= 16 (13%)	121 (100%)	
Facilita compresión (P. 7)	4.24 ± 0.979	4.04 ± 0.720	4.73 ± 0.467	4.50 ± 0.730	4.27 ± 0.876	0.109
Bubble plots (P. 8)	4.26 ± 0.908	4.38 ± 0.697	4.73 ± 0.467	4.63 ± 0.500	4.38 ± 0.799	0.171
Representación gráfica (P. 9)	4.24 ± 0.900	4.19 ± 0.749	4.73 ± 0.467	4.63 ± 0.619	4.32 ± 0.819	0.096
Acceso a datos (P. 10)	4.28 ± 0.928	4.27 ± 0.667	4.73 ± 0.467	4.63 ± 0.500	4.36 ± 0.806	0.170
Interpretación rápida (P. 11)	4.29 ± 0.978	4.27 ± 0.667	4.64 ± 0.674	4.56 ± 0.814	4.36 ± 0.874	0.458
Útil toma decisiones (P. 12)	4.31 ± 1.026	4.15 ± 0.613	4.73 ± 0.467	4.69 ± 0.602	4.36 ± 0.876	0.118
Facilidad explorar datos (P. 13)	4.26 ± 0.971	4.23 ± 0.652	4.82 ± 0.405	4.63 ± 0.719	4.36 ± 0.855	0.104
Evolución Histórica (P. 14)	4.34 ± 0.987	4.08 ± 0.891	4.82 ± 0.405	4.56 ± 0.727	4.36 ± 0.912	0.105
Menos conocimientos (P. 15)	4.18 ± 1.064	4.42 ± 0.758	4.64 ± 0.505	4.44 ± 0.727	4.31 ± 0.930	0.331
Satisfecho (P. 16)	4.32 ± 0.999	4.27 ± 0.667	4.91 ± 0.302	4.75 ± 0.447	4.42 ± 0.854	0.051
Promedio	4.27 ± 0.974	4.23 ± 0.708	4.74 ± 0.462	4.60 ± 0.638	4.35 ± 0.86	

El grupo de participantes menores de 33 años valoró más positivamente el ítem P. 14 (Evolución Histórica es útil), con una puntuación de 4.34 ± 0.987 , mientras que el ítem P. 15 (Se requieren menos conocimientos) obtuvo la puntuación más baja en este grupo, con 4.18 ± 1.064 . En promedio, este grupo puntuó la utilidad de TU® en 4.27, mostrándose como el tercer grupo de los cuatro en valoración comparado con las otras categorías de edad, y una mayor variabilidad en sus respuestas, como lo indica una DE promedio de 0.974, oscilando entre 0.900 y 1.064.

Para el grupo de edad entre 34 y 43 años, la pregunta mejor valorada fue la P. 8 (*Bubble plots* ayuda identificar resultados), con una puntuación de 4.38 ± 0.697 , y la menos valorada fue P. 14 (Evolución Histórica es útil), con 4.08 ± 0.891 . El promedio general de este grupo fue de 4.23, el más bajo de los cuatro grupos etarios, y con una variabilidad moderada en las respuestas, evidenciada por una DE promedio de 0.708, oscilando entre 0.613 y 0.891.

En el grupo de 44 a 54 años, el ítem P. 16 (Satisfacción con la metodología) obtuvo la puntuación más alta, con un 4.91 ± 0.302 . El ítem con menor puntuación en este grupo fue el ítem P. 11 (Interpretación más rápida de resultados), con 4.64 ± 0.674 . El promedio general de este grupo fue de 4.74, el más alto entre todos los grupos y de manera consistente, como lo muestra una DE promedio de 0.462, oscilando entre 0.302 y 0.674, la más baja de todos los grupos.

Finalmente, los participantes de 55 a 65 años también otorgaron la puntuación más alta al ítem P. 16 (Satisfacción con la metodología), con 4.75 ± 0.447 , mientras que el ítem P. 7 (Facilidad comprensión) fue la menos valorada, con 4.50 ± 0.730 , también este grupo tuvo un promedio general de 4.60, con una DE promedio de 0.638, oscilando entre 0.447 y 0.814.

El análisis global de los resultados muestra que el ítem mejor valorado en general fue el ítem P. 16 (Satisfacción con la metodología), con un promedio de puntuación de 4.42 ± 0.854 , mientras que el ítem P. 7 (Facilidad comprensión) obtuvo de promedio la puntuación más baja, con 4.27 ± 0.876 . Además, las DE indican que los grupos de mayor edad (44-54 y 55-65 años) presentaron menor variabilidad en sus respuestas con DE promedio de 0.462 y 0.638 respectivamente. Los grupos más jóvenes mostraron una mayor variabilidad con DE promedio de 0.974 y 0.708 respectivamente.

4.2.5.2 Análisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU® según edad

Se realizó un análisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU®, diferenciando por grupos de edad. Los resultados reflejan una percepción positiva uniforme en todos los grupos, con predominio de valoraciones altas (categorías 4 y 5) y una variabilidad mínima en las respuestas. Los detalles específicos sobre las frecuencias de valoración por ítem y grupo de edad, incluidas las tablas y figuras asociadas, se presentan en el **ANEXO 9**.

4.2.5.3 Subanálisis las puntuaciones medias, DE y significación de la encuesta de utilidad TU® por categorías participantes jóvenes (< 44 años) y participantes mayores (≥ 44 años)

Se ha realizado un subanálisis agrupando las edades en profesionales jóvenes, menores de 44 años (N=94; 77.68%), y profesionales mayores o iguales a 44 años (N=27; 22.31%) (**Tabla 39**).

En primer lugar, analizamos las diferencias en las puntuaciones medias junto con su DE y diferencia estadística por cada ítem de la encuesta comparando el grupo de los profesionales jóvenes frente a los mayores.

En el ítem sobre si la interfaz interactiva de la metodología facilita la comprensión de los resultados en comparación con las tablas estáticas de las URT (P. 7), los profesionales mayores obtuvieron una media de 4.59 (DE = 0.636) frente a una media de 4.18 (DE = 0.915) en los profesionales jóvenes, con una significación estadística de 0.015. Esto indica que los profesionales mayores perciben una mayor facilidad de comprensión de la interfaz interactiva.

En el ítem de la utilidad de los *bubble plots* para identificar rápidamente los resultados de interés (P. 8), la media para los mayores fue de 4.67 (DE = 0.48) comparada con 4.3 (DE = 0.853) para los jóvenes, con una significación de 0.003.

La representación gráfica de los resultados de interés (P. 9) y el acceso a datos detallados para mejorar la evaluación (P. 10) fueron también mejor valorados por los mayores con medias de 4.67 (DE = 0.555) y 4.67 (DE = 0.48), respectivamente, frente a medias de 4.22 (DE = 0.857) y 4.28 (DE = 0.86) en los jóvenes, con significaciones de 0.006 y 0.002, respectivamente.

En la percepción sobre si la metodología permite una interpretación más rápida de los datos (P. 11), los mayores tuvieron una media de 4,59 (DE = 0,747), ligeramente superior a la media de 4,29 (DE = 0,899) de los jóvenes, aunque esta diferencia no alcanzó significación estadística ($p = 0,055$).

En la utilidad de la información proporcionada para la toma de decisiones (P. 12), los mayores valoraron significativamente más esta característica con una media de 4.7 (DE = 0.542) en comparación con los jóvenes cuya media fue de 4.27 (DE = 0.93), con una significación de 0.001.

En la facilidad de uso de la interfaz para una mayor exploración de los datos (P. 13), los mayores obtuvieron una media de 4.7 (DE = 0.609) frente a una media de 4.26 (DE = 0.891) en los jóvenes, con una significación de 0.008.

La evolución histórica de la evidencia proporcionada por la metodología (P. 14) fue mejor valorada por los mayores con una media de 4.67 (DE = 0.62) frente a 4.27 (DE = 0.964) de los jóvenes, con una significación de 0.006. En cuanto a la percepción de que la metodología requiere menos conocimientos estadísticos para interpretar los resultados (P. 15), los

mayores también tuvieron una media superior de 4.52 (DE = 0.643) en comparación con 4.24 (DE = 0.991) en los jóvenes, con una significación de 0.046.

La satisfacción general con la metodología TU® (P. 16), los jóvenes mostraron una mayor media de 4.31 (DE = 0.916) frente a 4,18 (DE = 0.396) de los mayores, con una significación de 0,001.

Analizando el promedio global de las puntuaciones, los profesionales mayores presentaron una media de 4,596 (DE = 0.5708), mientras que los jóvenes tuvieron una media de 4.262 (DE = 0.9076). Esto indica una tendencia general de los profesionales mayores a valorar más positivamente la metodología TU®. Además, las desviaciones estándar más bajas en los grupos de mayores sugieren una menor variabilidad en sus respuestas, lo que podría indicar una percepción más consistente y uniforme de la metodología dentro de este grupo etario.

En conclusión, los profesionales mayores de 44 años tienden a valorar más positivamente varios aspectos de la metodología The Umbrella Collaboration® en comparación con los profesionales más jóvenes. Esto es evidente en las diferencias significativas en varios ítems de la encuesta, especialmente en la utilidad para la toma de decisiones (P. 12), la facilidad de exploración de datos (P. 13) y la necesidad de menos conocimientos estadísticos (P. 15). Sin embargo, en términos de satisfacción general (P. 16), los profesionales más jóvenes muestran una mayor puntuación media.

Tabla 39. Puntuaciones media, DE y valor-p por ítem de la encuesta por grupo de edad jóvenes (< 44 años) y mayores (> 44 años).

Pregunta (P. N.º)	Jóvenes (< 44 años)		Mayores (≥ 44 años)		Valor-p
	N=94 (77.68%)		N=27 (22.31%)		
	Media	DE	Media	DE	
TU Facilita comprensión (P. 7)	4.18	0.915	4.59	0.636	0.015
Bubble plots ayuda identificar resultados (P. 8)	4,3	0.853	4.67	0.48	0.003
Representación gráfica mejora evaluación (P. 9)	4.22	0.857	4.67	0.555	0.006
Acceso a datos mejora evaluación (P. 10)	4.28	0.86	4.67	0.48	0.002
Interpretación más rápida de resultados (P. 11)	4.29	0.899	4.59	0.747	0.055
Útil toma en toma de decisiones (P. 12)	4.27	0.93	4.7	0.542	0.001
Facilita más la exploración de los datos (P. 13)	4.26	0.891	4.7	0.609	0.008
Evolución Histórica es útil (P. 14)	4.27	0.964	4.67	0.62	0.006
Requiere menos conocimientos estadísticos (P. 15)	4.24	0.991	4.52	0.643	0.046
Satisfacción con metodología TU (P. 16)	4.31	0.916	4.18	0.396	0.001
Promedio	4,262	0.9076	4.596	0.5708	

4.2.5.4 Subanálisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU® por categorías participantes jóvenes (< 44 años) y participantes mayores (≥ 44 años)

Se llevó a cabo un subanálisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU®, diferenciando entre profesionales jóvenes (< 44 años) y mayores (≥ 44 años). Los resultados muestran que los profesionales mayores tienden a valorar más positivamente los ítems de la metodología TU® en comparación con los jóvenes, especialmente en ítems relacionados con la rapidez en la interpretación de datos, la facilidad para explorar los datos y la satisfacción general con la metodología, donde se observaron diferencias estadísticamente significativas. Las valoraciones bajas (categorías 1 y 2) fueron mínimas en ambos grupos, reflejando una percepción general positiva. Los detalles completos de este análisis, incluidas tablas y figuras, se presentan en el **ANEXO 10**.

4.2.6 Análisis por hospitales

4.2.6.1 Análisis de las puntuaciones medias, DE y valor p de la encuesta de utilidad TU® por Hospitales

Describimos la puntuación media junto con su DE y significación estadística a cada ítem de la encuesta para cada hospital que participó en la encuesta y todos los hospitales en forma global (**Tabla 40**).

En general, las puntuaciones medias totales por hospital fueron altas, situándose en un rango desde 4.107 (HU. Puerta de Hierro), 4.291 (HU. La Paz), 4.349 (HU. Ramón y Cajal), 4.364 (HU. Clínico), y 4.452 (HU. Cruz Roja). El promedio de medias totales para todas las preguntas fue de 4.35 ± 0.860, indicando una valoración generalmente positiva de la metodología. Las DE, que reflejan la consistencia de las respuestas, variaron entre los hospitales. De esa manera la DE más elevada corresponde al HU. Puerta de Hierro con un valor promedio de 1.06 (IC: 0.801 – 1.291), seguido por el HU. La Paz con 1.007 (IC: 0.902 – 1.065), el HU. Clínico 0.938 (IC: 0.86 – 1.063), el HU. Cruz Roja con 0.678 (IC: 0.636 – 0.911) y finalmente la DE más baja corresponde con el HU. Ramón y Cajal con 0,673 (IC: 0.596 – 0.739).

Tabla 40. Puntuación media, DE y valor p de cada ítem por Hospitales.

Pregunta (N.º)	HU. Puerta de Hierro	HU. Clínico	HU. La Paz	HU. Ramón y Cajal	HU. Cruz Roja	Valor-p
	N=13 (10.7%)	N=21 (17.4%)	N=25 (20.7%)	N=22 (18.2%)	N=40 (33.1%)	
P. 7	3.92 ± 1.188	4.40 ± 0.913	4.14 ± .014	4.23 ± 0.612	4.40 ± 0.123	0.415
P. 8	4.15 ± 0.801	4.36 ± 0.952	4.33 ± 1.017	4.32 ± 0.646	4.53 ± 0.640	0.637
P. 9	4.08 ± 0.862	4.40 ± 0.913	4.10 ± 1.044	4.23 ± 0.612	4.53 ± 0.679	0.218
P. 10	4.15 ± 0.801	4.24 ± 0.879	4.29 ± 1.056	4.45 ± 0.596	4.50 ± 0.716	0.549
P. 11	4.00 ± 1.291	4.36 ± 0.860	4.43 ± 0.978	4.41 ± 0.666	4.40 ± 0.778	0.654

Pregunta (N.º)	HU. Puerta de Hierro	HU. Clínico	HU. La Paz	HU. Ramón y Cajal	HU. Cruz Roja	Valor-p
	N=13 (10.7%)	N=21 (17.4%)	N=25 (20.7%)	N=22 (18.2%)	N=40 (33.1%)	
P. 12	4.31 ± 1.182	4.36 ± 0.907	4.33 ± 1.017	4.18 ± 0.733	4.50 ± 0.751	0.742
P. 13	4.31 ± 1.109	4.32 ± 0.945	4.38 ± 0.973	4.36 ± 0.658	4.38 ± 0.774	0.998
P. 14	4.15 ± 1.068	4.32 ± 1.069	4.33 ± 1.065	4.45 ± 0.739	4.40 ± 0.778	0.907
P. 15	4.08 ± 1.188	4.32 ± 1.030	4.29 ± 0.902	4.45 ± 0.739	4.30 ± 0.911	0.854
P. 16	3.92 ± 1.115	4.56 ± 0.917	4.29 ± 1.007	4.41 ± 0.734	4.58 ± 0.636	0,139
Total	4.107 ± 1.06	4.364 ± 0.938	4.291 ± 1.007	4.349 ± 0.673	4.452 ± 0.678	

El Hospital Universitario Puerta de Hierro obtuvo un promedio de puntuación de 4.107 (DE: 1.0605; IC: 0.801 - 1.291). Esto sugiere una cierta variabilidad en las respuestas, aunque las puntuaciones se mantuvieron en el rango superior. Los ítems con mayor puntuación media fueron P. 12 (Utilidad de la metodología para tomar decisiones) con 4.31 (DE: 1.182) y P. 13 (Facilidad explorar datos) con 4.31 (DE: 1.109), mientras que la menor fue P. 7 (Facilita comprensión) con 3,92 (DE: 1.188). Las DE fueron más altas en las preguntas P. 11 (1.291) y P. 12 (1.182), indicando menos consistencia en las respuestas para estos ítems.

El Hospital Universitario Clínico de San Carlos presentó un promedio de puntuación de 4.364 (DE: 0.9385; IC: 0.86 – 1.069). La mayor puntuación media se observó en el ítem P. 16 (Satisfacción) con 4.56 (DE: 0.917), mientras que la menor fue en el ítem P. 10 (Acceso a datos) con 4.24 (DE: 0.879).

En el Hospital La Paz, el promedio de puntuación fue de 4.291 (DE: 1.0073; IC: 0.902 - 1.056). El ítem mejor valorado fue P. 11 (Interpretación rápida) con 4.43 (DE: 0.978), y la menor valoración se dio en P. 9 (Representación gráfica) con 4.1 (DE: 1.044). Las DE fueron las segundas más altas por hospitales con un promedio de 1.073, lo que sugiere respuestas menos consistentes; la mayor variabilidad se encontró en P. 10 (1.056), y la menor en P. 15 (0.902).

El Hospital Ramón y Cajal alcanzó promedio de puntuación de 4.349, (DE: 0.6735; IC: 0.596 - 0.978). La puntuación más alta con 4.45 fue para los ítems P. 10 (Acceso a datos), P. 14 (Evolución histórica) y P. 15 (Menos conocimientos), mientras que la más baja fue para el ítem P. 12 (Utilidad de la metodología para tomar decisiones) con 4.18 (DE: 0.733). Las DE fueron relativamente bajas, lo que indica una alta consistencia en las respuestas; con un promedio de 0.6735, la mayor variabilidad se observó en el ítem P. 15 (0.739) y la más baja en el ítem P. 10 (0.596).

El Hospital Cruz Roja tuvo el promedio de puntuación más alta entre todos los hospitales, con 4.452. Las DE promedio de 0.7441, oscilando entre 0.636 y 0.911. La puntuación más alta se observó en el ítem P. 16 (Satisfacción) (4.58), y la más baja en el ítem P. 15 (Menos conocimientos) (4,3).

El análisis ANOVA se utilizó para determinar si existían diferencias estadísticamente significativas en las valoraciones de los distintos ítems de la encuesta sobre la metodología TU® entre los cinco hospitales participantes. Los resultados muestran que, para todos los ítems evaluados, no hubo diferencias estadísticamente significativas entre los hospitales en ninguna de las preguntas (**Tabla 40**).

4.2.6.2 Análisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU® por Hospitales

Se realizó un análisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU®, diferenciando entre hospitales participantes. Los resultados reflejan que las valoraciones más altas (categoría 5) fueron predominantes en todos los hospitales y para la mayoría de los ítems evaluados, aunque con ligeras variaciones en la distribución de frecuencias por ítem y hospital. Los detalles específicos del análisis, incluidas tablas y figuras que resumen las valoraciones para cada hospital, se presentan en el **ANEXO 11**.

4.2.7 Análisis de médicos especialistas en geriatría y resto de especialidades

En esta sección vamos a exponer y analizar los resultados obtenidos por la categoría de especialistas en geriatría que participaron en la encuesta, haciendo un análisis detallado de geriatras frente a médicos no geriatras.

4.2.7.1 Diferencias poblacionales entre Geriatras y no Geriatras

La participación de los médicos fue de 104, de los cuales 93 (89.43%) fueron geriatras y 11 (10.57%) no geriatras (**Tabla 41**). No hay claras diferencias significativas en cuanto a los grupos de edad con grupos de edad de tendencia más jóvenes en Geriatras con respecto a no Geriatras (mayores de 44 años en geriatría 15% vs 54.6% en no Geriatras, p=0.07).

Tabla 41. Características Demográficas de los Participantes según Especialidad (Geriatras vs. No Geriatras).

	Geriatras 93 (89.43%)	No Geriatras 11 (10.57%)	Valor-p
Género			
Mujer	71 (76.3%)	9 (81.8%)	0.74
Hombre	25 (23.7%)	2 (18.2%)	
Subgrupos de edad			
< 33 años	57 (61.3%)	5 (45.5%)	0.07
33-43 años	22 (23.7%)	-	
44-54 años	7 (7.5%)	2 (18.2%)	
55-65 años	7 (7.5%)	4 (36.4%)	

	Geriatras 93 (89.43%)	No Geriatras 11 (10.57%)	Valor-p
Género			
H U Puerta de Hierro-Majadahonda	7 (7.5%)	1 (9.1%)	0.44
H U La Paz	20 (21.5%)	2 (18.2%)	
H U Clínico San Carlos	21 (22.6%)	-	
H U Ramón y Cajal	16 (17.2%)	3 (27.3%)	
H U Central de la Cruz Roja "San José y Santa Adela"	29 (31.2%)	5 (45.5%)	

4.2.7.2 Puntuaciones medias, DE y valor p de la encuesta de utilidad TU® entre Geriatras y no Geriatras.

Se muestran las diferencias de puntuaciones medias a los ítems de la encuesta de utilidad de TU® entre los dos grupos a estudio (**Tabla 42**).

La puntuación para los médicos geriatras de promedio es de 4.34, oscilando entre 4.26 y 4.42, mientras que para los médicos no-geriatras el promedio es de 4.51 variando entre 4,36 y 4,64.

En general, las puntuaciones de los médicos no-geriatras son ligeramente superiores a las de los médicos geriatras en todos los ítems, pero sin significación estadística para ningún ítem. El ítem P. 9 presenta la mayor diferencia en la puntuación entre ambos grupos, con una media de 4.28 para los geriatras y 4.64 para los no-geriatras, mostrando una tendencia hacia una mayor satisfacción o acuerdo con la utilidad de la metodología entre los no-geriatras.

Las DE indican la dispersión de las respuestas. Para los geriatras, las DE varían entre 0.849 y 0.967, mientras que para los no-geriatras oscilan entre 0.522 y 0.822. Las menores DE en el grupo de los no-geriatras sugieren una mayor consistencia en sus respuestas en comparación con los geriatras.

Tabla 42. Puntuación media y DS de cada ítem entre Geriatras y no Geriatras.

Pregunta (N.º)	Geriatras (N=93)		No-Geriatras (N=11)		Valor-p
	Media	DE	Media	DE	
P. 7	4.26	0.908	4.55	0.522	0.153
P. 8	4.4	0.849	4.45	0.688	0.416
P. 9	4.28	0.864	4.64	0.674	0.095
P. 10	4.35	0.868	4.45	0.522	0.355
P. 11	4.34	0.915	4.55	0.522	0.238

P. 12	4.34	0.915	4.55	0.688	0.241
P. 13	4.34	0.891	4.36	0.809	0.472
P. 14	4.37	0.93	4.55	0.688	0.268
P. 15	4.31	0.967	4.45	0.82	0.32
P. 16	4.42	0.901	4.55	0.688	0.327
Promedio	4.341	0.9008	4.51	0.6621	

4.2.7.3 Análisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU® entre Geriatras y no Geriatras.

Se realizó un análisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU®, comparando geriatras y no-geriatras. Los resultados muestran que ambos grupos otorgaron principalmente valoraciones altas (categorías 4 y 5), representando el 85.9% de las valoraciones en los geriatras y el 91.81% en los no-geriatras. Aunque los no-geriatras tendieron a otorgar valoraciones máximas (categoría 5) con mayor frecuencia (59.09%) que los geriatras (54.40%), no se observaron diferencias estadísticamente significativas entre los dos grupos para ninguno de los ítems evaluados. Los detalles completos de este análisis, incluidas las frecuencias de valoración por ítem y los resultados estadísticos, se presentan en el **ANEXO 12**.

4.2.8 Análisis de médicos especialistas en geriatría

Dado que el grupo de Geriatras es el más abundante de la encuesta (93 (89.43%)), es oportuno profundizar en las respuestas de los diferentes ítems de dicho grupo.

4.2.8.1 Análisis de las puntuaciones medias y DE de la encuesta de utilidad TU® entre Geriatras.

El análisis de las puntuaciones medias y DE proporcionadas por los geriatras en la encuesta de utilidad de la metodología TU® ofrece una visión sobre la percepción general de esta herramienta para los especialistas en geriatría encuestados (**Tabla 43**).

Tabla 43. Puntuación media y DS de cada ítem entre Geriatras (N=93).

Pregunta	Media puntuación	Desviación Estándar
Facilita compresión (P. 7)	4.26	0.908
Bubble plots (P. 8)	4.4	0.849
Representación gráfica (P. 9)	4.28	0.864
Acceso a datos (P. 10)	4.35	0.868
Interpretación rápida (P. 11)	4.34	0.915
Útil toma decisiones (P. 12)	4.34	0.915
Facilidad explorar datos (P. 13)	4.34	0.891

Pregunta	Media puntuación	Desviación Estándar
Evolución Histórica (P. 14)	4.37	0.93
Menos conocimientos (P. 15)	4.31	0.967
Satisfacción (P. 16)	4.42	0.967
Promedio	4.341	0.9074

En términos generales, las puntuaciones medias de las respuestas a la encuesta obtienen un promedio de 4.341(DE: 0,9074) oscilando entre 4.26 y 4.42 en una escala de 1 a 5. Este rango sugiere una valoración positiva generalizada de los geriatras hacia la metodología. El ítem que recibió la puntuación media más alta fue la P. 16, relativa a la satisfacción general, con una media de 4.42 (DE: 0.967).

El ítem P. 8, que se refiere a los *"Bubble plots"*, también obtuvo una puntuación media elevada de 4.4 (DE: 0.849), indicando que esta representación gráfica es muy apreciada por los geriatras. Los ítems P. 14 y P. 10, relacionadas con la evolución histórica de los datos y el acceso a los mismos, también se destacaron con puntuaciones medias de 4.37 (DE: 0.93) y 4.35 (DE: 0.868), respectivamente.

Las DEs, que miden la variabilidad en las respuestas, varían de 0.849 a 0.967. El ítem P. 8 (*Bubble plots*), con una DE de 0.849, muestra la menor variabilidad en las respuestas, sugiriendo una opinión bastante homogénea entre los geriatras respecto a los *"Bubble plots"*. Por otro lado, los ítems P. 15 (Menos conocimientos) y P. 16 (Satisfacción general), con desviaciones estándar de 0.967, indican una mayor variabilidad en las opiniones sobre la utilidad de la metodología para quienes tienen menos conocimientos estadísticos previos y la satisfacción general, respectivamente.

4.2.8.2 Análisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU® en el grupo de Geriatras

Se realizó un análisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU® dentro del grupo de geriatras (N=93). Los resultados muestran una clara tendencia hacia las valoraciones altas, con un 54.40% de las respuestas en la categoría máxima (5) y un 31.50% en la categoría 4. Las valoraciones bajas (1 y 2) fueron mínimas, representando menos del 4% del total de respuestas. Los ítems mejor valorados incluyeron P. 14 (Evolución histórica) y P. 16 (Satisfacción general), ambos con un 60.21% de valoraciones máximas. Estos hallazgos reflejan una percepción muy positiva de la herramienta TU® entre los geriatras. Los detalles completos de este análisis, incluidas tablas y figuras, se presentan en el **ANEXO 13**.

4.2.8.3 Análisis de las puntuaciones medias, DE y diferencia estadística de la encuesta de utilidad TU® entre Geriatras por grupos de edad

El análisis de las puntuaciones medias por ítem en la encuesta de utilidad de la metodología TU® por parte de los Geriatras revela variaciones entre los cuatro grupos de edad de geriatras participantes (**Tabla 44**).

Tabla 44. Puntuación media, DE y valor p de cada ítem por grupos de edad de geriatras.

Pregunta	< 33 a		33 - 43 a		44 - 54 a		55 - 65 a		Total		ANOVA valor-p
	N=57; 61.3%		N=22; 23.7%		N=7; 7.5%		N=7; 7.5%		N=93; 100%		
	Media	DE	Media	DE	Media	DE	Media	DE	Media	DE	
P. 7	4.23	1.018	4.05	0.722	4.57	0.535	4.86	0.378	4.26	0.908	0.161
P. 8	4.26	0.955	4.45	0.671	4.86	0.378	4.86	0.378	4.4	0.849	0.130
P. 9	4.21	0.940	4.18	0.795	4.57	0.535	4,86	0.378	4.28	0.864	0.207
P. 10	4.28	0.996	4.27	0.631	4.71	0.488	4.86	0.378	4.35	0.868	0.247
P. 11	4.28	1.031	4.23	0.685	4.57	0.787	5	0.000	4.34	0.915	0.199
P. 12	4.3	1.068	4.18	0.588	4.71	0.488	4.86	0.378	4.34	0.915	0.244
P. 13	4.26	1.009	4.18	0.664	4.86	0.378	5	0.000	4.34	0.891	0.060
P. 14	4.37	0.993	4.05	0.899	4.71	0.488	5	0.000	4.34	0.930	0.074
P. 15	4.23	1.086	4.36	0.790	4.57	0.535	4.57	0.787	4.31	0.967	0.689
P. 16	4.37	1.029	4.23	0.685	4.86	0.378	5	0.000	4.42	0.901	0.123
Promedio	4.279	1.0125	4.218	0.713	4.699	0.499	4.887	0.2677	4.338	0.9008	

En términos generales, el grupo de edad más joven (<33 años) mostró puntuaciones medias que oscilaron entre 4.21 y 4.7, con DE relativamente altas, indicando una mayor variabilidad en las respuestas. El ítem P. 14 (Evolución histórica) obtuvo la puntuación más alta (4.37; DE: 0.993), mientras que el ítem P. 9 (Representación gráfica) obtuvo la puntuación más baja (4.21; DE: 0.940). Las DE en este grupo variaron entre 0.940 y 1.086, reflejando una dispersión considerable en las opiniones.

El grupo de edad entre 33 y 43 años presentó puntuaciones que oscilan entre 4.05 y 4.45, con la puntuación más alta en el ítem P. 8 (*Bubble plots*) con 4.45 (DE: 0,671) y la más baja en el ítem P. 7 (Facilita comprensión) con 4.05 (DE: 0.722). Las DE en este grupo fueron menores, variando de 0.588 a 0.899, lo que indica una mayor consistencia en las respuestas comparado con el grupo más joven.

El grupo de edad entre 44 y 54 años mostraron puntuaciones consistentemente altas, con un rango de 4.57 a 4.86. Los ítems P. 8 (*Bubble plots*) y P. 13 (Facilidad para explorar datos) obtuvieron las puntuaciones más altas de 4.86 (DE: 0,378/0,378), mientras que el ítem P. 7 (Facilita comprensión) obtuvo la puntuación más baja dentro de este grupo con 4.57 (DE: 0.535). Las DE en este grupo fueron las más bajas, oscilando entre 0.378 y 0.535, lo que refleja una alta homogeneidad en las respuestas.

El grupo de edad más avanzado, entre 55 y 65 años, presentó las puntuaciones más altas entre todos los grupos, con un rango de 4.86 a 5.00. Los ítems P. 11 (Interpretación rápida), P. 13 (Facilidad para explorar datos), P. 14 (Evolución histórica) y P. 16 (Satisfacción) obtuvieron la puntuación máxima de 5.00 (DE: 0.000), reflejando una percepción extremadamente positiva de estos aspectos de la metodología. Las DE en este grupo fueron las más bajas

en general, con varias preguntas obteniendo una DE de 0,000, indicando una uniformidad absoluta en las respuestas.

El promedio total de las puntuaciones para todos los grupos fue de 4.338 con una DE de 0.9008. Comparando los promedios de cada grupo, se observa que los grupos de mayor edad (44-54 y 55-65 años) tienden a valorar más positivamente la metodología, con promedios de 4.699 y 4.887 respectivamente, y menores DE, lo que indica respuestas más consistentes. Los grupos más jóvenes (<33 y 33-43 años) presentan promedios ligeramente inferiores (4.279 y 4.218) y mayores DE, sugiriendo una mayor variabilidad en las opiniones.

El análisis de varianza (ANOVA) se utilizó para determinar si existían asociaciones entre las puntuaciones de los diferentes grupos de edad para cada ítem de la encuesta, no encontrándose ninguna diferencia estadísticamente significativa.

4.2.8.4 Análisis descriptivo de la frecuencia de valoración de valoración por ítem en la encuesta de utilidad de TU® por grupos de edad de los Geriatras

Se realizó un análisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU®, desglosando los resultados por grupos de edad dentro del colectivo de geriatras. Los resultados muestran que las valoraciones más altas (categoría 5) fueron predominantes en todos los grupos de edad, con una tendencia general hacia una percepción positiva de los ítems evaluados. Los geriatras menores de 33 años (N=57) otorgaron valoraciones 5 especialmente a los ítems P. 16 (Satisfacción) y P. 14 (Evolución histórica), con un 61.40% cada uno. En el grupo de 33 a 44 años (N=22), los ítems P. 8 (*Bubble plots*) y P. 15 (Menos conocimientos) destacaron con un 54.54% de valoraciones 5.

Entre los geriatras de 44 a 54 años (N=7), los ítems P. 8 (*Bubble plots*) y P. 13 (Facilidad para explorar datos) recibieron un 85.71% de valoraciones 5. Finalmente, en el grupo de 55 a 65 años (N=7), los ítems P. 11 (Interpretación rápida), P. 13 (Facilidad para explorar datos), P. 14 (Evolución histórica) y P. 16 (Satisfacción) lograron un 100% de valoraciones 5. Las valoraciones bajas (categorías 1 y 2) fueron mínimas o inexistentes en todos los grupos de edad.

Los detalles completos de este análisis, incluidas tablas y figuras, se presentan en el **ANEXO 14**.

4.2.8.5 Análisis de las puntuaciones medias, DE y significación estadística de la encuesta de utilidad TU® entre Geriatras jóvenes (< 44 años) y Geriatras mayores (≥ 44 años)

Se ha realizado al igual que en el estudio global de la encuesta, un subestudio dentro del grupo de geriatras agrupando las edades en geriatras jóvenes, menores de 44 años (N:79) y geriatras mayores o iguales a 44 años (N: 14) (**Tabla 45**).

Tabla 45. Puntuación media y DE por ítem por grupo de geriatras de edad jóvenes (< 44 años) y mayores (> 44 años).

Pregunta	Geriatras Jóvenes (< 44 años)		Geriatras Mayores (≥ 44 años)		ANOVA valor p
	N=79 (84.94%)		N=14 (15.05%)		
	Media	DE	Media	DE	
P. 7	4.18	0.944	4.71	0.469	0.02
P. 8	4.32	0.885	4.86	0.363	0.001
P. 9	4.2	0.897	4.71	0.469	0.02
P. 10	4.28	0.905	4.79	0.426	0.001
P. 11	4.27	0.943	4.79	0.579	0.005
P. 12	4.27	0.957	4.79	0.426	0.001
P. 13	4.24	0.923	4.93	0.267	0.001
P. 14	4.28	0.973	4.86	0.363	0.001
P. 15	4.27	1.009	4.57	0.646	0.139
P. 16	4.33	0.944	4.93	0.267	0.001
Promedio	4.264	0.938	4.794	0.4275	

El análisis de las puntuaciones medias por ítem en la encuesta de utilidad de la metodología TU® entre los geriatras revela variaciones significativas entre los dos grupos de edad de geriatras participantes: jóvenes (menores de 44 años) y mayores (mayores o iguales a 44 años).

En términos generales, los geriatras jóvenes (< 44 años) mostraron un promedio de puntuación de 4.264 con las puntuaciones medias que oscilan entre 4.18 y 4.33, con DE de 0.938 (IC: 0.885 – 1.009). El ítem P. 16 (Satisfacción con la metodología) obtuvo la puntuación más alta (4.33; DE: 0.944), mientras que el ítem P. 7 (Facilita la comprensión) obtuvo la puntuación más baja (4.18; DE: 0.944). Las DE en este grupo variaron entre 0.885 y 1.009, reflejando una dispersión considerable en las opiniones.

Por otro lado, los geriatras mayores (≥ 44 años) presentaron una puntuación media global de 4.794 con las puntuaciones que oscilan entre 4.57 y 4.93, con la puntuación más alta en el ítem P. 16 (Satisfacción con la metodología) con 4.93 (DE: 0.267) y la más baja en el ítem P. 15 (Requiere menos conocimientos estadísticos) con 4.57 (DE: 0.646). Las DE en este grupo fueron menores con un promedio de 0.4275, variando de 0.267 a 0.646, lo que indica una mayor consistencia en las respuestas comparado con los geriatras jóvenes.

En promedio, los geriatras mayores valoraron la metodología TU® más positivamente (media puntuación: 4.794; DE: 0.4275) en comparación con los geriatras jóvenes (media puntuación: 4.264; DE: 0.938).

El análisis de varianza (ANOVA) se utilizó para determinar si existían asociaciones entre las puntuaciones de geriatras según la categoría de jóvenes frente a mayores para cada ítem de la encuesta, encontrándose diferencias estadísticamente significativas para todos los ítems exceptuando el ítem P. 15 sobre la necesidad de conocimientos estadísticos.

Las diferencias estadísticamente significativas observadas sugieren que la edad y, posiblemente, la experiencia profesional influye en cómo los geriatras interpretan y valoran la utilidad de la herramienta. Los geriatras mayores tienden a otorgar puntuaciones más altas y a mostrar una percepción más consistente de los beneficios de TU®, lo que podría indicar una apreciación específica de las características de la herramienta que facilitan la interpretación y el uso práctico de la evidencia en su toma de decisiones clínicas. Este hallazgo sugiere que los geriatras de mayor edad, con una trayectoria probablemente más extensa en la evaluación de evidencia, encuentran en TU® un recurso especialmente valioso para organizar y sintetizar información de manera clara y accesible. La menor variabilidad en sus respuestas apunta a una valoración más uniforme, lo cual podría reflejar una confianza consolidada en el potencial de TU® para mejorar su práctica clínica en el ámbito geriátrico. Por otro lado, la mayor dispersión en las respuestas de los geriatras jóvenes sugiere una diversidad de perspectivas, que podría deberse a distintas fases de adaptación o a un proceso de aprendizaje en la interpretación de herramientas de síntesis de evidencia. Esta variabilidad no implica una percepción negativa, sino que puede reflejar expectativas diversas y una apertura a explorar diferentes maneras de usar TU® en el contexto clínico.

4.2.8.6 Subanálisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU® por categorías geriatras jóvenes (< 44 años) y geriatras mayores (≥ 44 años)

Se llevó a cabo un subanálisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU®, diferenciando entre geriatras jóvenes (< 44 años) y mayores (≥ 44 años). Los resultados muestran que los geriatras mayores tienden a valorar más positivamente los ítems de la metodología TU® en comparación con los jóvenes, especialmente en ítems relacionados con la rapidez en la interpretación de datos, la facilidad para explorar los datos y la satisfacción general con la metodología. Las valoraciones bajas (categorías 1 y 2) fueron mínimas en ambos grupos, reflejando una percepción general positiva. Los detalles completos de este análisis, incluidas tablas y figuras, se presentan en el ANEXO 15.

4.2.8.7 Análisis de las puntuaciones medias, DE y significación estadística de la encuesta de utilidad TU® entre Geriatras por hospitales

El promedio total de puntuaciones para los geriatras de todos los hospitales fue de 4,341 con una DE de 0.9008 (Tabla 46).

Los geriatras del Hospital Puerta de Hierro tienden a valorar la metodología de manera menos favorable, con una media de 3.928 y una DE de 1.2794, la más alta entre todos los hospitales, indicando una mayor variabilidad en las respuestas. Este hospital frecuentemente se encuentra por debajo del promedio total, lo que sugiere una percepción menos positiva en comparación con otros hospitales.

Los geriatras del Hospital Clínico presentan una media de puntuación de 4.355 y una DE de 1.015. A pesar de estar por encima del promedio total, la DE relativamente alta refleja una considerable variabilidad en las respuestas.

Los geriatras del Hospital La Paz tienen una media de 4.291 y una DE de 1.007, ligeramente por debajo del promedio total, con una alta variabilidad en las opiniones.

Los geriatras del Hospital Ramón y Cajal muestran una media de puntuación de 4.395 y una DE de 0.6324, destacándose por estar por encima del promedio total y por tener una de las DE más bajas, lo que indica una mayor consistencia en las respuestas.

Los geriatras del Hospital Cruz Roja presentan media de puntuación más alta de 4.44 y una DE de 0.7533.

Tabla 46. Puntuación Media, DE y significación (ANOVA) por geriatras distintos hospitales.

Pregunta	HU Puerta de Hierro (N=7; 7.52%)	HU Clínico (N=20; 21.5%)	HU La Paz (N=21; 22.58%)	HU Ramón y Cajal (N=16; 17.2%)	HU Cruz Roja (N=29; 31.18%)	Total (N=93)	ANOVA valor p
P. 7	3.71+/-1.49	4.35+/-0.98	4.14+/-1.01	4.25+/-0.57	4.41+/-0.73	4.26+/-0.90	0.425
P. 8	4.14+/-1.06	4.35+/-1.04	4.33+/-1.01	4.38+/-0.61	4.55+/-0.63	4.4+/-0.84	0.783
P. 9	3.86+/-1.06	4.35+/-0.98	4.1+/-1.04	4.25+/-0.57	4.48+/-0.68	4.28+/-0.86	0.362
P. 10	4+/-1	4.25+/-0.96	4.29+/-1.05	4.44+/-0.62	4.52+/-0.73	4.35+/-0.86	0.608
P. 11	3.57+/-1.51	4.35+/-0.93	4.43+/-0.97	4.44+/-0.62	4.41+/-0.78	4.34+/-0.91	0.24
P. 12	3.86+/-1.46	4.4+/-0.99	4.33+/-1.01	4.25+/-0.68	4.48+/-0.73	4.34+/-0.91	0.584
P. 13	4+/-1.41	4.35+/-0.98	4.38+/-0.97	4.38+/-0.61	4.38+/-0.77	4.34+/-0.89	0.892
P. 14	4.43+/-1.41	4.25+/-1.16	4.33+/-1.06	4.5+/-0.73	4.38+/-0.77	4.37+/-0.93	0.953
P. 15	4.08 +/-1.18	4.35+/-1.08	4.29+/-0.90	4.56+/-0.62	4.24+/-0.98	4.31+/-0.96	0.743
P. 16	3.71+/-1.38	4.55+/-0.99	4.29+/-1	4.5+/-0.63	4.55+/-0.68	4.42+/-0.90	0.205
Promedio	3.92+/-1.21	4.35 +/-1.01	4.29+/-1	4.39+/-0.63	4.44+/-0.75	4.34+/-0.90	

- Al analizar los ítems específicos de la encuesta, se observan algunas tendencias:
- En el ítem P.7, los geriatras del Hospital Puerta de Hierro tienen la media más baja (3.71) y la mayor DE (1.496), situándose por debajo del promedio total (4.26). En contraste, el Hospital Cruz Roja presenta la puntuación más alta en este ítem (4.41), superando el promedio con una DE de 0.733.
 - En el ítem P.8, los geriatras de Puerta de Hierro valoran más bajo (4.14) y los de Cruz Roja más alto (4.55).
 - En el ítem P.9, la puntuación más baja es para Puerta de Hierro (3.86), mientras que Cruz Roja alcanza 4.48.

- El ítem P.10, muestra una media más baja en Puerta de Hierro (4.00), y la más alta en Cruz Roja (4.52). Este patrón de Puerta de Hierro con puntuaciones más bajas y Cruz Roja con puntuaciones más altas se repite en varios ítems, como P.11 (interpretación rápida) y P.12 (utilidad en la toma de decisiones).
- En el ítem P.13, los geriatras del HU Puerta de Hierro vuelve a tener una media de 4.00, la más baja, mientras que los hospitales restantes presentan medias más consistentes, todas cercanas a 4.35.
- Para P.14, los geriatras del HU Clínico San Carlos muestra una ligera desviación por debajo del promedio (4.25), mientras que Ramón y Cajal se destaca con la puntuación más alta (4.50).
- El ítem P.15, los geriatras del HU Puerta de Hierro presentan la media más baja (4.00), y Ramón y Cajal con la más alta (4.56).
- El ítem P.16, los geriatras del HU Puerta de Hierro está en la parte inferior (3.71), mientras que HU Clínico San Carlos y HU Cruz Roja están en la parte superior (4.55).

El análisis de varianza (ANOVA) se utilizó para determinar si existían asociaciones entre las puntuaciones de los diferentes grupos de edad para cada ítem de la encuesta, no encontrándose ninguna diferencia estadísticamente significativa.

4.2.8.8 Análisis descriptivo de la frecuencia de valoración de valoración por ítem en la encuesta de utilidad de TU® entre Geriatras por hospitales

Se realizó un análisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU®, desglosando las respuestas de los geriatras según los cinco hospitales participantes. Los resultados muestran que las valoraciones más altas (categoría 5) fueron predominantes en todos los hospitales, destacando el Hospital Cruz Roja con el mayor porcentaje de valoraciones 5 (59.31%) y el Hospital Clínico San Carlos con el 58.00%. Los ítems mejor valorados variaron entre hospitales: el ítem P. 14 (Evolución histórica) fue el más destacado en el Hospital Puerta de Hierro, mientras que el ítem P. 16 (Satisfacción) predominó en el Clínico San Carlos, La Paz y la Cruz Roja.

Las valoraciones bajas (1 y 2) fueron mínimas en todos los hospitales, siendo más frecuentes en el Puerta de Hierro (12.85%) en comparación con el Clínico San Carlos (5.50%), La Paz (4.75%), Cruz Roja (0.68%) y Ramón y Cajal (0.00%). Un análisis estadístico mediante la prueba de chi-cuadrado no encontró diferencias estadísticamente significativas entre las respuestas de los geriatras de los diferentes hospitales.

Los detalles completos de este análisis, incluidas tablas y figuras, se presentan en el **ANEXO 16**.

5

DISCUSIÓN

5.1 DISCUSION GLOBAL

The Umbrella Collaboration® (TU®), por los resultados obtenidos en este trabajo de investigación, se presenta como una metodología válida para la síntesis terciaria de evidencia, mostrando su capacidad para alcanzar resultados comparables a las revisiones *umbrella* tradicionales (URT). TU® replica la mayoría de los resultados de interés (RdI) identificados por las URT con un alto nivel de concordancia y amplía el alcance al identificar un número adicional de RdI. Su métrica experimental, R_{TU} , demuestra una fuerte correlación con medidas estadísticas consolidadas de tamaño de efecto, como la d de Cohen, lo que refuerza su validez como herramienta para evaluar la magnitud de intervenciones o exposiciones a riesgos. Además, la valoración de certeza de TU®, basada en un análisis de sentimiento automatizado, presenta una concordancia aceptable con el sistema GRADE, un estándar ampliamente reconocido. También se observó un alto grado de concordancia en la dirección y la significación estadística de los RdI entre ambas metodologías, lo que refuerza la robustez de los resultados de TU®.

Entre las principales ventajas de TU® destaca su rapidez en la ejecución, completando proyectos en horas frente a los meses requeridos por las URT. Esta eficiencia, junto con su capacidad para ofrecer resultados consistentes, la posiciona como una herramienta valiosa para contextos en los que la síntesis de evidencia rápida es crucial o los medios son limitados. La posibilidad de actualización continua de los proyectos cada vez que surgen nuevos RS/MA añade un valor adicional que las URT no ofrecen.

En cuanto a su aceptación, los resultados del estudio sobre la percepción de los profesionales sanitarios de geriatría muestran una evaluación favorable de TU®, destacando su utilidad y claridad. Aunque algunos encuestados expresaron desconocimiento inicial sobre la metodología, la mayoría reconoció su potencial para facilitar la integración de evidencia en la práctica clínica. Este resultado es alentador y sugiere que TU® podría ganar mayor aceptación a medida que se familiaricen más con sus capacidades.

Sin embargo, todos estos hallazgos deben interpretarse con cautela. La limitada cantidad de datos analizados, tanto en la comparación de metodologías como en la encuesta de percepción, refleja las restricciones inherentes a un estudio exploratorio y pionero en este campo. Este trabajo representa un primer paso en la validación de TU®, y será necesario ampliar estas investigaciones con muestras más grandes y en contextos variados para consolidar su utilidad y aplicabilidad. Aun así, los resultados preliminares son prometedores y establecen una base sólida para el desarrollo continuo y la integración de TU® en las prácticas de síntesis de evidencia.

5.2 DISCUSIÓN ESTUDIO I

La discusión de este trabajo tiene como objetivo evaluar los hallazgos obtenidos en relación con las hipótesis planteadas, situándolos en el contexto de la literatura científica existente.

Inicialmente, se evalúa la capacidad de TU® para alcanzar conclusiones comparables al patrón oro que suponen las URT, centrándose en aspectos clave como la identificación y en las propiedades de los RdI y la eficiencia en los tiempos de ejecución. Este análisis busca proporcionar una visión de cómo TU® se podría posicionar como una herramienta innovadora en la síntesis de evidencia terciaria, destacando su capacidad para enfrentar las limitaciones de las URT y ofrecer soluciones adaptadas a las necesidades del ámbito clínico.

5.2.1 Identificación de URT en el área de la geriatría para la comparación metodológica

Para evaluar la validez de TU® en comparación con el enfoque tradicional, fue necesario identificar un conjunto de URT que sirviera como referencia. Este proceso de selección fue realizado de manera selectiva, ajustándose a los objetivos específicos de este estudio.

Se realizó una búsqueda focalizada en PubMed el 5 de marzo de 2023, utilizando los términos "*Umbrella*" y "*Geriatric*", con el propósito de identificar revisiones representativas en el área de la geriatría. Este enfoque respondió a la necesidad de encontrar un grupo manejable y representativo de URT para habilitar una comparación metodológica eficiente. De las 111 entradas iniciales, 36 cumplieron los criterios de inclusión, y finalmente se incluyeron 8 de manera aleatoria para el análisis comparativo con TU®.

Es importante subrayar que este procedimiento fue diseñado intencionalmente para satisfacer los requisitos de la comparación metodológica sin pretender un análisis exhaustivo del espectro completo de la literatura disponible. Aunque se identificaron 36 URT en la búsqueda inicial, se seleccionaron de forma aleatoria solo 8 para esta comparación. Esta decisión responde a que este estudio representa una primera valoración de una metodología experimental como TU®, careciendo de experiencias o referencias previas frente a la metodología tradicional. En un contexto donde las tecnologías de automatización avanzan rápidamente, era prioritario presentar resultados iniciales del comportamiento de una herramienta capaz de asistir en todas las etapas de la síntesis terciaria de evidencia.

Por otra parte, el desarrollo de TU® responde al creciente interés en herramientas de automatización para la síntesis de evidencia, un área que sin duda marcará el futuro cercano de la investigación científica. Sin embargo, la introducción de nuevas metodologías siempre plantea desafíos, y TU® no ha sido la excepción. Durante las presentaciones del proyecto, se ha observado cierto desconocimiento generalizado respecto al concepto de síntesis terciaria automatizada, lo que refleja la necesidad de un mayor esfuerzo en la difusión y comprensión de estas herramientas.

Es crucial enfatizar que TU® no busca reemplazar a las URT, sino complementarlas con un enfoque alternativo que aprovecha la tecnología para circunstancias en las que el abordaje tradicional no sea oportuna. Este modelo representa una evolución metodológica, no una ruptura, y tiene el potencial de coexistir con los métodos convencionales, ofreciendo una herramienta adicional para enfrentar los retos actuales de la síntesis de evidencia. A

medida que crezca la familiaridad y confianza en este tipo de innovaciones, es probable que el interés por TU® y otras herramientas similares aumente, consolidando su lugar en el ecosistema de la investigación científica.

5.2.2 Identificación de Resultados de Interés (RdI)

La identificación de RdI constituye un eje esencial en la comparación entre TU® y las URT. En el contexto de la síntesis terciaria, un RdI representa un hallazgo específico evaluado previamente en una o más RS/MA, que se integra como resultado de la síntesis para responder a una pregunta de investigación. Este concepto no solo subraya la importancia de agregar datos relevantes, sino también de presentar un resultado final que represente de manera fiel y precisa las conclusiones globales derivadas de los estudios analizados (RS/MA).

Dado que la literatura sobre síntesis terciaria es limitada, esta comparación recurre a principios y hallazgos de síntesis secundaria para contextualizar y analizar los factores metodológicos en juego. Muchas de las limitaciones y fortalezas identificadas en síntesis secundaria se comparten con la síntesis terciaria, lo que permite extrapolar ciertos aspectos y aplicar estrategias similares.

El análisis comparativo entre las metodologías de URT y TU® reveló diferencias sustanciales en la cantidad y diversidad de RdI identificados. En los ocho proyectos analizados, las URT identificaron un total de 86 RdI, mientras que TU® detectó 410, lo que supone un incremento de casi cinco veces (factor multiplicador de 4.77). Además de coincidir en el 84.88% de los RdI identificados por las URT, TU® aportó 337 nuevos RdI no capturados por las revisiones tradicionales. Esta capacidad de TU® para identificar una mayor cantidad de RdI no parece explicarse únicamente por las estrategias de búsqueda o los criterios de inclusión. Según el Handbook de Cochrane (108), las metodologías *umbrella* permiten dos variantes principales en la formulación de la pregunta de investigación: preguntas amplias, que buscan describir el cuerpo completo de evidencia, o preguntas específicas, que se centran en un área concreta (133). Si el objetivo es describir exhaustivamente el cuerpo de evidencia, la pregunta debe ser amplia y evitar restricciones excesivas en la estrategia de búsqueda, ya que limitar el campo puede dificultar el descubrimiento de nuevas áreas. Aunque TU® permite realizar ambos enfoques, al igual que una URT tradicional, en este estudio TU® empleó criterios de búsqueda y criterios de inclusión amplios, mientras se observó que algunos proyectos URT, como el proyecto 6, aplicaron restricciones tanto en la estrategia de búsqueda como en los criterios de inclusión, y los proyectos 3 y 4 limitaron los criterios de inclusión. Sin embargo, incluso al considerar únicamente los proyectos URT que no aplicaron restricciones, TU® mantiene un factor multiplicador cercano a 5 (4.93). Esto sugiere que alguna propiedad inherente de TU® permite identificar una mayor cantidad de RdI. En primer lugar, el algoritmo de TU® opera de manera sistemática y equitativa, minimizando posibles sesgos humanos en la selección de resultados. Este enfoque automatizado garantiza que cualquier RdI presente en el *abstract* de una RS/MA sea identificado y registrado, incluso si luego se considera menos relevante por un revisor

humano. Esto contrasta con los métodos manuales tradicionales, donde los investigadores tienden a priorizar los RdI más relevantes, lo que podría derivar en la exclusión de algunos resultados potencialmente significativos. Otra propiedad clave de TU® es su capacidad para manejar grandes volúmenes de información y datos sin la limitación de la sobrecarga de trabajo, un desafío común en metodologías tradicionales. Esto permite incluir múltiples RdI sin comprometer la calidad o el alcance del análisis, un factor decisivo en la identificación de resultados adicionales que de otro modo podrían quedar fuera.

Otro factor respecto a la capacidad de identificación de RdI en la comparación entre TU® y las URT es la evaluación de la calidad de los estudios primarios incluidos. Las URT tienden a priorizar RS/MA de alta calidad, lo que limita el número de estudios analizados, pero favorece una mayor precisión y rigor en los resultados obtenidos. Por el contrario, TU®, al basarse en *abstracts*, no realiza valoraciones explícitas de calidad, lo que permite incluir un mayor número de estudio, aumentando los RdI identificados. Este enfoque, sin embargo, no debe considerarse una limitación insalvable, sino una característica inherente al diseño metodológico de TU®, que prioriza la exhaustividad en la identificación de RdI. La literatura respalda la importancia de la calidad en la síntesis de evidencia, pero no se incumple el análisis de calidad exclusivamente en TU® (199). La valoración de la calidad en síntesis secundaria no es cumplimentada por todas las RS/MA, como documenta Seehra et al. (200) hasta un 30% de las RS/MA no evaluaron explícitamente la calidad de los estudios incluidos, y únicamente el 12.9% establecieron umbrales de calidad como criterio de inclusión para metaanálisis. Esto pone de manifiesto que incluso en revisiones sistemáticas tradicionales, no siempre se logra un control estricto de calidad.

Aunque TU® no evalúa la calidad de las RS/MA incluidas de manera directa, el uso de herramientas como el listado de verificación para *abstracts* de PRISMA (40) podría implementarse en el futuro para abordar esta limitación. Sin embargo, es importante señalar que, según estudios previos, los *abstracts* tienden a recibir puntuaciones bajas en términos de calidad según los criterios de PRISMA (201) (202) (203) (204). Lo que suscita mayor interés en los resultados, y a pesar de que los *abstracts* no suelen cumplir todos los criterios PRISMA, TU® es capaz de identificar solo en los *abstracts*, más RdI que los URT.

Otro posible factor que influye en las diferencias en la capacidad de identificación de RdI entre TU® y las URT es el inevitable sesgo cognitivo inherente a toda actividad humana (205). En las URT, los autores pueden tender a seleccionar RdI que estén alineados con su experiencia, familiaridad o con una hipótesis específica, lo que limita la variabilidad de los resultados reportados. Este aspecto no queda del todo resuelto en TU®, que precisa de cierta supervisión por parte de un revisor humano en la selección y aceptación final de los resultados, pero gran parte de la metodología está estructurada en procesos realizados por un programa informático que pretende reducir los sesgos cognitivos comúnmente realizados por humanos, evaluar todos los estudios bajo los mismos parámetros y una neutralidad en la interpretación. Esta tendencia se observa claramente en la dirección de los efectos presentados en los RdI analizados en las URT, que solo muestran efectos en

una sola dirección. En contraste, TU®, en el análisis de todos los Rdl identificados, no solo los coincidentes con URT, refleja efectos en ambas direcciones, positivos y negativos. Este enfoque de TU® está en línea con las recomendaciones establecidas por organismos como el Joanna Briggs Institute y la Cochrane Collaboration, que enfatizan la importancia de presentar un balance entre los beneficios y los riesgos para proporcionar una visión integral de los efectos observados. Según estas directrices, una *umbrella review* debería incluir tanto resultados positivos como adversos para facilitar una toma de decisiones más informada (108,141).

Finalmente, cabe considerar la posibilidad de que TU® pueda identificar Rdl que sean triviales o de escasa relevancia clínica. Sin embargo, este escenario parece poco probable, ya que todos los Rdl detectados por TU® provienen de RS/MA previamente publicados, los cuales, en teoría, deberían cumplir con estándares mínimos de calidad y relevancia para ser publicados e indexados en la base de datos de MEDLINE. Según el Handbook de Cochrane, sus revisiones deben incluir únicamente resultados significativos y evitar aquellos triviales, limitándose a un número reducido de resultados críticos e importantes (133). Este principio metodológico sugiere que los Rdl identificados en RS/MA publicados deberían ser clínicamente significativos en su mayoría. No obstante, para reforzar la aplicabilidad práctica de TU®, sería valioso evaluar en el futuro la relevancia clínica de los Rdl adicionales encontrados.

TU® demuestra ser una herramienta prometedora en la síntesis de evidencia terciaria, capaz de identificar Rdl de manera similar a los URT, pero además de ampliar considerablemente el rango de estos. No obstante, la relevancia clínica de los resultados adicionales debe ser evaluada en estudios futuros para garantizar su aplicabilidad práctica.

5.2.3 Discusión Tiempos de Ejecución Proyectos The Umbrella Collaboration®

El tiempo requerido para completar un proyecto de síntesis terciaria es un aspecto relevante en la evaluación de la eficiencia de este tipo de estudios, aunque rara vez ha sido estudiado o documentado específicamente. En el caso de las URT, no suele llevarse un registro preciso de los tiempos de realización, y los URT seleccionados para este estudio no fueron una excepción. En la metodología experimental TU®, el tiempo requerido para completar el proyecto de síntesis terciaria se registra de forma precisa. Este cronometraje incluye todo el proceso, desde el inicio del proyecto con la introducción del primer término de búsqueda hasta la finalización de su primera versión, sin contabilizar posibles actualizaciones posteriores que añadirían tiempo adicional. Esta capacidad de medir con exactitud el tiempo de ejecución en horas y minutos representa una novedad en la realización de proyectos de síntesis (ya sea secundaria o terciaria), un aspecto que nunca había sido registrado con tal nivel de detalle. Esta innovación no solo permite una evaluación precisa de la eficiencia en TU®, sino que además anticipa una tendencia en la que, con la incorporación de herramientas de automatización, la medición exacta de los tiempos de realización será cada vez más frecuente. Con ello, las estimaciones en términos de meses o años, habituales en la metodología tradicional, podrían pronto quedar obsoletas.

La ausencia de datos concretos de tiempos de ejecución de los URT nos ha impedido realizar un análisis comparativo directo de las diferencias en los tiempos de ejecución entre ambas metodologías. Sin embargo, los resultados obtenidos con TU® son lo suficientemente reveladores y destacables como para evidenciar las ventajas que ofrece esta metodología. Los tiempos de ejecución de los proyectos con la metodología TU®, en rangos de horas, contrastan con la creciente demanda de procesos ágiles y resultados rápidos, especialmente en contextos donde la toma de decisiones basada en evidencia requiere un acceso rápido a información de alto nivel.

En los proyectos TU® estudiados, el tiempo requerido para completar cada proyecto de síntesis terciaria osciló entre 1,5 y 10 horas, lo cual está por debajo de los tiempos generalmente asumidos para las URT, entre 6 – 12 meses, incluso contrasta con los tiempos de ejecución de RS/MA que se estima entre 1 – 2 años (30), y de sus adaptaciones aceleradas como las RRP entre 5 – 12 semanas (63) o incluso los 11 días laborales del método Two-Week SR (34).

La rapidez de la metodología experimental TU® representa una característica altamente favorable, especialmente en el contexto de la creciente demanda internacional de métodos más ágiles para la síntesis de evidencia. Organismos como la OMS (206), la Cochrane Collaboration (31) y la Joanna Briggs Institute (65) han enfatizado la importancia de disponer de metodologías capaces de ofrecer resultados en plazos reducidos sin comprometer la validez de los mismos, un principio clave que la metodología TU® parece cumplir al abordar síntesis terciarias en tiempos excepcionalmente bajos.

En este estudio, el tiempo de ejecución medido corresponde exclusivamente a la primera versión de los proyectos TU®, sin considerar actualizaciones posteriores. Sin embargo, la metodología The Umbrella Collaboration® presenta un enfoque innovador para la actualización de proyectos de síntesis terciaria, permitiendo que los proyectos permanezcan activos tras su creación. El sistema verifica automáticamente cada 24 horas si se han publicado nuevos estudios que cumplan con los criterios de búsqueda definidos. En caso de identificar nuevas RS/MA relevantes, se notifica al revisor humano, quien evalúa la idoneidad del nuevo estudio para su inclusión. Si se aprueba, los resultados son revisados, corregidos y confirmados, integrándose automáticamente al conjunto de resultados previos. Este nivel de actualización no tiene equivalentes en la síntesis terciaria, pero es reconocida la importancia de la actualización para mantener la relevancia y precisión de los resultados. En RS/MA se recomienda actualizar los resultados cada dos años o cuando surjan datos relevantes, aunque estudios han señalado importantes déficits en su implementación (38–41). La capacidad de TU® para integrar un sistema automatizado de actualización representa una ventaja significativa en la síntesis de evidencia, permitiendo mitigar la obsolescencia de los resultados y asegurar la relevancia continua en áreas de alto recambio de evidencia, que tendrá que ser evaluado en futuros estudios.

Los resultados indican que, en TU®, los proyectos con un mayor número de referencias incluidas tienden a requerir tiempos de ejecución algo más largos, lo cual es esperable

dado que una mayor cantidad de RS/MA incluidos, implica una evaluación más exhaustiva y consumidora de tiempo. Este hallazgo sugiere que, aunque TU[®] optimiza el proceso de síntesis terciaria, la complejidad inherente a la cantidad de información sigue siendo un factor que incide en el tiempo de ejecución total del proyecto, especialmente cuando se depende de un humano para realizar esta tarea. En un futuro quizá no lejano, las herramientas de automatización podrían mitigar este factor limitante. De hecho, en el proyecto The Umbrella Collaboration[®], ya se contempla la posibilidad de que el programa realice el proceso de análisis de los datos de las RS/MA de forma completa, eliminando la necesidad de intervención humana. No obstante, esta automatización aún requiere un proceso de aprendizaje basado en datos de proyectos realizados de forma semiautomática por revisores humanos, que permita mejorar los algoritmos con *machine learning* y alcanzar una autonomía suficiente en el análisis de grandes volúmenes de información sin la intervención humana.

Es destacable que el número de términos de búsqueda no parece estar directamente relacionado con el tiempo de ejecución (datos no publicados), lo que refleja la capacidad de los algoritmos de TU[®] para gestionar búsquedas de diferente complejidad sin impactar significativamente en el tiempo de ejecución del proyecto. Este patrón podría deberse a que, en TU[®], la fase de búsqueda está diseñada para procesar de manera simple y eficiente términos de búsqueda sin repercutir en el tiempo de ejecución de los proyectos. En las metodologías tradicionales de síntesis de evidencia, la fase de búsqueda suele ser altamente especializada y compleja, frecuentemente dependiente de documentalistas especializados, que emplean hilos de búsqueda detallados y consultas avanzadas en múltiples bases de datos, siguiendo criterios estrictos para garantizar la exhaustividad de la búsqueda (207–209). En contraste, la metodología TU[®] simplifica el proceso, y emplea múltiples términos de búsqueda básicos en PubMed, generados y sugeridos por IA generativa. Este enfoque simplificado y más automatizado presenta un cambio en el modo de realizar búsquedas, y plantea una posible línea de estudio para investigar la eficiencia y adecuación de métodos de búsqueda menos complejos en síntesis de evidencia. Esto podría sugerir que, en ciertos contextos, los modelos simplificados de búsqueda asistida por IA generativa podrían cumplir los objetivos con un grado de eficacia comparable, abriendo la puerta a una reevaluación de los requisitos y protocolos considerados óptimos hasta la fecha (11,88,96). La simplicidad y efectividad del método de búsqueda de TU[®], que utiliza términos de búsqueda básicos sugeridos por IA generativa y no requiere pasos complejos ni la intervención de documentalistas especializados, contribuye a simplificar y agilizar la fase de búsqueda, y contribuye a realizar proyectos de síntesis terciaria en tiempos reducidos.

Los tiempos de ejecución significativamente bajos registrados en TU[®] no solo optimizan el proceso de síntesis terciaria, sino que también hacen más viable la obtención rápida de conocimiento de alto nivel. Esta ventaja temporal podría, en ciertos contextos, compensar posibles imprecisiones que se observan en otros aspectos de la metodología TU[®], ofreciendo una alternativa ágil y funcional cuando la rapidez es prioritaria. Esto podría tener un impacto positivo en la toma de decisiones en salud pública y gestión sanitaria,

donde el acceso rápido a una síntesis actualizada es crucial para implementar políticas y estrategias efectivas. Así, TU[®] se perfila como una herramienta prometedora en contextos donde la agilidad es esencial, permitiendo una síntesis de evidencia que, aunque aún debe ser validada exhaustivamente frente a los estándares tradicionales, ofrece un equilibrio valioso entre rapidez y profundidad analítica.

5.2.4 Comparativa de las propiedades de los Rdl entre ambas metodologías

5.2.4.1 Discusión sobre la Concordancia y Correlación de los Tamaños de Efecto entre las Metodologías Comparativas URT y TU[®]

En esta sección, hemos evaluado la concordancia y correlación entre las métricas de los TE obtenidos mediante las metodologías URT y TU[®], a través de tres niveles de análisis progresivos (Niveles O, A y B). Dentro de la metodología The Umbrella Collaboration[®] se desarrolló una métrica, R_{TU} , que permite uniformizar las métricas de TE de las RS/MA en una métrica común para la agregación de los datos. Esta métrica propia fue comparada frente a las métricas habituales, d de Cohen y SMD. A diferencia de la d de Cohen y la SMD, que han sido ampliamente utilizadas y reconocidas en la literatura, R_{TU} representa una nueva aproximación diseñada para capturar TE en un contexto de síntesis terciaria de evidencia, que precisa de una validación, en la que este estudio ha sido la primera ocasión.

En el Nivel O, donde se transformaron todas las métricas de TE de las URT a la d de Cohen, se observó que la transformación de ciertas métricas, especialmente los TE expresados como SMD, introducía inconsistencias significativas en los datos. Las magnitudes correspondientes a Rdl medidos con SMD, al ser transformados a la d de Cohen modificaban de forma alejada a la expresada en el URT el TE original, motivo por el cual se decidió no transformar la magnitud de SMD y utilizarla directamente en la comparación en los niveles sucesivos. Por otro lado, la transformación de las MD y WMD a SMD (para luego su transformación a la d de Cohen) que se empleó, al no disponer de los tamaños muestrales, fue el error estándar (SE) como estimador del desviación estándar (SD), un enfoque que sabemos que no es óptimo y que podía distorsionar la equivalencia entre las métricas de URT y TU[®], motivo por el cual se decidió eliminar los datos que tuviesen como métrica el MD o el WMD en el nivel B, conscientes de la pérdida considerable de datos podría afectar en parte algunos resultados (210,211).

En la evaluación en el Nivel O, observamos 15 valores que presentaban discordancia de más de un nivel de categoría en la categorización cualitativa de TE entre TU[®] y URT. De estos, tras un análisis de las RS/MA que fueron incluidos en los URT, comprobamos que dos valores resultaron ser clasificaciones no ajustadas en las propias URT. Estos errores subrayan la vulnerabilidad de las URT, las cuales hemos considerado "patrón oro" para la comparación con el método experimental TU[®]. Estos errores son causados principalmente por errores humanos a la hora de manejar grandes cantidades de datos en las distintas fases del proceso de síntesis de evidencia. La presencia de errores en el manejo de datos ha sido objetivada en estudios previos, incluso con una frecuencia alta en RS/MA, los

cuales pueden comprometer la calidad de la evidencia sintetizada. Xu et al., por ejemplo, encontraron que hasta el 85% de las RS/MA contenían errores de extracción de datos en al menos un metaanálisis (212). Asimismo, Mathes et al. destacaron que estos errores de extracción pueden introducir sesgos significativos en las estimaciones del efecto (213), mientras que el estudio de Kovacs et al. señala que estos errores suelen derivarse de problemas en la gestión de proyectos y preparación, afectando la eficiencia y la validez de las conclusiones en estudios complejos con grandes volúmenes de datos (214). Estos hallazgos destacan el potencial de la automatización, como proponemos en el proyecto The Umbrella Collaboration®, para reducir la incidencia de errores humanos, mejorando la consistencia y la precisión en la extracción y categorización de datos, lo cual resulta especialmente relevante en investigaciones de síntesis de evidencia complejas, como las URT.

La literatura sugiere que la validación de nuevas métricas implica no solo probar su precisión y especificidad, sino también identificar y corregir posibles fuentes de error que podrían afectar la interpretación de los resultados. Es posible que se cuestione si las modificaciones en los Niveles A y B fueron aplicadas para mejorar artificialmente los resultados de concordancia y correlación entre las metodologías URT y TU®. No obstante, resulta necesario aclarar que las decisiones de ajuste no se implementaron para beneficiar los resultados, sino como una respuesta a problemas metodológicos identificados en el análisis del Nivel O, y que se alinean con las mejores prácticas de investigación, que recomiendan ajustes para reducir el sesgo potencial y evitar interpretaciones incorrectas cuando se detectan inconsistencias en los métodos de transformación de datos (215).

Uno de los principales hallazgos en la comparación de concordancia de TE categóricos entre las metodologías en estudio, es la mejora progresiva en el grado de concordancia plena y consistente (suma de concordancia plena y concordancia parcial) a medida que se aplicaban los ajustes descritos. En el Nivel O, la concordancia plena fue del 28.8%, y la concordancia consistente alcanzó el 77.3%. En el Nivel A aumentó la concordancia plena hasta el 40.0% y en la concordancia consistente hasta el 89.2%. En el Nivel B aumentó la concordancia plena al 50.0% y la concordancia consistente al 93.8%. Este incremento en la concordancia sugiere que los ajustes introducidos fueron efectivos para reducir las distorsiones provocadas por las transformaciones de las métricas en el Nivel O, y que contribuyeron a una mayor alineación en la categorización cualitativa de los TE entre ambas metodologías. La reducción de la discordancia a lo largo de los niveles (del 22.7% en el Nivel O al 6.3% en el Nivel B) también apoya esta interpretación, indicando que los ajustes implementados mejoraron la precisión de la comparación.

Es especialmente positivo que, a pesar de su carácter experimental, la métrica R_{TU} se aproxima de manera notable a métricas establecidas como la d de Cohen y el SMD. Este hallazgo refuerza la idea de que R_{TU} tiene potencial para ofrecer resultados coherentes con los de métricas tradicionales, a pesar de las diferencias en su diseño y aplicación.

En la evaluación de concordancia entre metodologías sobre la clasificación categórica de

los TE del Nivel B, observamos una pérdida de significación estadística de la asociación entre las variables, en comparación con los Niveles O y A, lo cual podría interpretarse superficialmente como una disminución en la validez de los ajustes realizados. Sin embargo, esta interpretación sería incorrecta, pues la significación estadística (valor p) puede ser sensible al tamaño de muestra y, por lo tanto, su pérdida no implica necesariamente una disminución en la relevancia práctica de los resultados obtenidos. La reducción del tamaño de muestra en el Nivel B ha limitado de forma natural la potencia estadística de la prueba de χ^2 . Es importante destacar que esto no es un indicio de menor validez o representatividad en los hallazgos, sino una consecuencia esperada al aplicar ajustes necesarios para mejorar la precisión de la comparación entre ambas metodologías.

Cada vez más, en la literatura científica se cuestiona la dependencia en el valor p para la toma de decisiones en investigación, debido a su sensibilidad al tamaño de la muestra y a que no refleja necesariamente la relevancia clínica o práctica de los hallazgos. En nuestro estudio, esta dependencia es evidente. Ya que la reducción en el tamaño muestral del Nivel B ha contribuido a esta pérdida de significación, pero pensamos que no afecta el valor práctico de los hallazgos. Como han señalado diversos autores, el valor p es útil para determinar si un efecto existe, pero es el tamaño del efecto, el intervalo de confianza y la magnitud de la concordancia los que ofrecen una visión más completa y práctica sobre el impacto real de los resultados (216–219). En el mismo sentido es importante reconocer que la presencia de celdas vacías y de frecuencias bajas en la tabla de contingencia del Nivel B también resta potencia a la prueba de contraste (220).

La fuerza de la asociación entre las categorías cualitativas de los TE asignadas por ambas metodologías, medido a través de la V de Cramer, proporciona una cuantificación de dicha asociación. Aunque la V de Cramer no refleja la magnitud clínica del efecto, nos ofrece una comprensión de la interdependencia entre ambas clasificaciones categóricas de los TE de los Rdl, los valores de V de Cramer entre 0.437 (Nivel O), 0.380 (Nivel A) y 0.339 (Nivel B) indican que ambas metodologías tienden a clasificar la fuerza de asociación de manera similar, y la magnitud de esta fuerza sugiere que TU® está capturando patrones de clasificación de manera cercana a la URT. Incluso en el Nivel B, donde la prueba de contraste no es significativa, la V de Cramer sigue siendo una medida válida para interpretar la fuerza de la asociación observada en la muestra, donde la V de Cramer de 0.339 sugiere que hay una relación moderada entre las clasificaciones de TE en URT y TU® en la muestra observada. Esto puede tener relevancia práctica, especialmente si estamos interesados en ver si TU® logra capturar patrones similares a URT en la clasificación categórica de TE, mostrando una consistencia en las clasificaciones entre ambos métodos. En investigaciones aplicadas, la fuerza de la asociación puede ser más importante que la significación estadística cuando se evalúa la utilidad práctica de un método experimental (221). En el mismo sentido de la fuerza de la asociación, el principal valor de este análisis radica en la alta concordancia consistente observada en cada nivel: 77.3% en el Nivel O, 89.2% en el Nivel A, y 93.8% en el Nivel B. Este nivel de concordancia refleja la capacidad de TU® para aproximarse a la categorización de TE establecida por URT, lo cual es una evidencia

significativa de su validez práctica en el contexto de la síntesis terciaria de la evidencia. Los resultados de concordancia obtenidos confirman que TU[®] se aproxima de manera notable a métricas tradicionales como la *d* de Cohen y el SMD, y que los ajustes realizados a lo largo de los niveles O, A y B han fortalecido la validez de esta comparación sin introducir sesgos en favor de TU[®]. Este hallazgo respalda la idea de que, aunque el valor *p* sigue siendo una herramienta útil, la relevancia práctica y la concordancia entre metodologías proporcionan una interpretación más completa y fiable de los resultados. En última instancia, estos datos sugieren que TU[®] tiene un alto potencial para ser una herramienta fiable en estudios de síntesis de evidencia y en la categorización de TE.

En cuanto a la correlación de los valores cuantitativos de TE entre ambas metodologías, el coeficiente de correlación de Spearman mostró una correlación positiva y estadísticamente significativa en todos los niveles, aunque con una disminución en su magnitud a medida que se introducían ajustes. De forma positiva la pendiente de la línea de regresión, que comenzó con un valor de 2.72 en el Nivel O, se redujo a 2.11 en el Nivel A y a 0.8 en el Nivel B, acercándose al valor ideal de 1. Esto indica una relación más equilibrada entre los TE de ambas metodologías en el Nivel B, lo que sugiere que los ajustes han logrado reducir la amplificación excesiva observada en los Niveles O y A en URT, de lo cual podemos pensar que los datos procedentes de métricas de MD o WMD fueran las responsables del nivel pronunciado de la pendiente. Sin embargo, el coeficiente de determinación (R^2) fue bajo en todos los niveles (de 0.20 en el Nivel O a 0.11 en el Nivel B), lo cual implica que una parte importante de la variabilidad en los TE de URT no es explicada por los TE de TU[®]. Este resultado podría indicar que factores adicionales influyen en la relación entre ambas métricas y que el modelo lineal no es completamente adecuado para capturar esta relación.

Para consolidar la validez de R_{TU} , serán necesarios estudios adicionales que exploren su especificidad, sensibilidad y rendimiento en distintos contextos de investigación, así como una validación formal que evalúe su precisión en comparación con otros métodos. Estos resultados, en el complejo área de las métricas de TE, representan un primer paso en la evaluación de la métrica R_{TU} y de la metodología TU[®], como una herramienta potencialmente capaz de ofrecer una alternativa eficaz y efectiva en la síntesis terciaria de evidencia.

5.2.4.2 Dirección y Significación Estadística de los Resultados de Interés (Rdl)

En los trabajos de síntesis terciaria, la dirección y la significación estadística de los Rdl son aspectos fundamentales para interpretar el impacto de las intervenciones en salud y las exposiciones a riesgos. En este contexto, la dirección del efecto refleja si una intervención favorece o perjudica un resultado de salud, o si una exposición incrementa o disminuye un riesgo. Por su parte, la significación estadística evalúa si el efecto observado es robusto frente a la variabilidad estadística, considerando el intervalo de confianza de las métricas analizadas.

En este estudio, la comparación entre TU[®] y las URT mostró resultados consistentes en términos de dirección y significación estadística en el 82.19% de los casos en su análisis

inicial. Este porcentaje aumentó al 87.67% tras revisar y ajustar discrepancias entre los datos en las URT y como estaban reflejadas en las RS/MA incluidas en los URT, lo que subraya la importancia de una evaluación minuciosa de las fuentes para garantizar la calidad de los datos comparativos. Los ajustes realizados identificaron cuatro clasificaciones incorrectas en los URT de las 13 posibles, lo que refuerza la necesidad de herramientas automatizadas o semiautomatizadas, como The Umbrella Collaboration[®], para minimizar errores humanos y estructurales inherentes a los métodos tradicionales.

Un hallazgo notable fue la asociación altamente significativa entre las clasificaciones de dirección y significación estadística entre ambas metodologías ($\chi^2 = 25.28$, $p < 0.001$), con una concordancia medida con el índice Kappa de Cohen mostró niveles de concordancia moderada ($\kappa = 0.514$) y una fuerza de asociación considerada grande según la *V* de Cramer ($V = 0.588$). Estos resultados sugieren que TU[®] logra capturar patrones consistentes con los de las URT, a pesar de las diferencias en el diseño metodológico.

Finalmente, la concordancia observada en la definición de la dirección y significación estadística de los TE entre TU[®] y el estándar de oro representado por las URT consolida la posición de TU[®] como una metodología válida para la síntesis terciaria de evidencia. Este resultado no solo refuerza la validez de TU[®] demostrada en otros aspectos evaluados en este trabajo, sino que también ofrece confianza a futuros usuarios, tanto para quienes deseen emplear TU[®] en proyectos de síntesis terciaria como para quienes consulten sus resultados.

5.2.4.3 Correlación entre las Certezas de los Rdl de las Metodologías

En este estudio, la comparación de certezas se realizó exclusivamente con los URT que utilizaron GRADE para valorar la certeza, debido a su enfoque consolidado y ampliamente aceptado en la evaluación de la calidad de la evidencia (222), alineado con los conceptos utilizados en The Umbrella Collaboration[®]. En los URT se objetivan otras valoraciones de calidad, como los criterios bioestadísticos propuestos por Papatheodorou (115) que ofrecen un método estructurado para clasificar la certeza en revisiones *umbrella*, pero han recibido críticas significativas, como las señaladas por Schlesinger et al. (223), quienes cuestionaron la dependencia de valores estadísticos arbitrarios y su capacidad limitada para capturar la relevancia clínica y el riesgo de sesgo. Y por otro lado, la valoración con AMSTAR-2 que se centra en la calidad metodológica de las RS/MA (110), mientras que la certeza evaluada por TU[®] está más directamente relacionada con los conceptos de calidad utilizados en GRADE, lo que refuerza el sentido de esta comparación limitada a las valoraciones de las URT que fuesen realizadas con esta herramienta, aunque TU[®] propone un enfoque innovador que incluye un sistema de valoración de certeza basado en el análisis de sentimiento del contenido textual de los *abstracts*.

Inicialmente, el sistema GRADE hacía referencia a la "calidad de la evidencia", que describía el grado de confianza en que los resultados de los estudios eran válidos para apoyar decisiones clínicas. Con el tiempo, este término evolucionó a "confianza en las estimaciones", subrayando la precisión con la que se podían realizar recomendaciones

específicas basadas en la evidencia disponible (148). Actualmente, GRADE ha adoptado el término "certeza de la evidencia", que ha sido reconocido como una mejor representación de este concepto, ya que engloba tanto la confianza en la magnitud del efecto estimado como en su relevancia para la toma de decisiones (149).

La comparación entre las certezas, transformadas a una escala común, de los Rdl de la metodología experimental de TU® y los Rdl de los URT que utilizan GRADE, como estándar consolidado para valorar la certeza, proporciona información clave sobre la validez del enfoque automatizado de TU®. Los resultados de este análisis indican una correlación positiva y estadísticamente significativa entre ambas metodologías, lo que sugiere que TU® es capaz de aproximarse a los criterios rigurosos y estandarizados de GRADE. Sin embargo, estos hallazgos deben interpretarse con precaución debido al bajo número de valores analizados, lo que limita la generalización de los resultados y subraya la necesidad de investigaciones futuras para validar y consolidar esta relación en diferentes contextos y con muestras más amplias.

Es importante destacar que esta comparación también refuerza la confianza en el uso de TU®, tanto para investigadores que deseen aplicar esta herramienta en futuros proyectos como para quienes consulten sus resultados. El alineamiento con GRADE, aunque no perfecto, consolida a TU® como una metodología innovadora que no busca sustituir, sino complementar las prácticas tradicionales, aportando rapidez y automatización sin comprometer significativamente la calidad de las evaluaciones.

Este hallazgo, además, invita a reflexionar sobre el potencial de integrar análisis automatizados con criterios establecidos como GRADE, buscando optimizar tanto la eficiencia como la precisión en la síntesis de evidencia. A medida que estas metodologías evolucionan, la validación continua será clave para consolidar su papel en la práctica científica.

5.2.5 Contextualización de resultados con estudios previos

Nuestros hallazgos se inscriben en la creciente búsqueda de métodos que equilibren rapidez y rigor para la toma de decisiones en salud, especialmente en contextos con información limitada, recursos escasos y necesidad de información actualizada en tiempos recortados. En este sentido, el marco de síntesis rápida de evidencia (RES, por sus siglas en inglés) desarrollado por el *Greater Manchester Applied Research Collaboration* (ARC-GM), en colaboración con la *Greater Manchester Academic Health Science Network* (AHSN), proporciona un referente clave (224). El RES ha sido reconocido como una herramienta oportuna, flexible y rigurosa, que permite identificar, evaluar y contextualizar evidencia relevante de manera rápida y transparente. Este enfoque está profundamente integrado en el ecosistema de investigación y atención sanitaria de Greater Manchester, utilizando metodologías basadas en el marco GRADE para realizar evaluaciones rápidas y sistemáticas de la evidencia disponible, evaluando su relevancia para problemas específicos, de manera similar a operar que plantea TU®.

Nuestros resultados, así mismo, refuerzan la creciente validación de herramientas de automatización y de IA como apoyo para la síntesis de evidencia en investigación científica. En este sentido, el trabajo de FutureHouse (225) sobre el modelo PaperQA2 destaca el potencial de los LLM para superar las limitaciones tradicionales en revisiones de evidencia. Los datos de FutureHouse muestran que PaperQA2 iguala o supera el desempeño de expertos humanos en tareas de investigación científica. En nuestro estudio sobre la validación de TU® frente al patrón oro de las URT, hemos observado resultados alineados con esta tendencia. En particular, TU® ha demostrado una capacidad para identificar Rdl de manera aumentada con respecto a URT, con una concordancia aceptable con los resultados de URT y con aproximaciones en áreas relevantes como el TE, y la certeza en los resultados, además de que los tiempos de ejecución de los proyectos con la metodología TU® son extraordinariamente recortados. Estos hallazgos corroboran la utilidad de herramientas basadas en desarrollos informáticos e IA como apoyo en la síntesis y evaluación de evidencia en salud.

En conjunto, tanto RES, PaperQA2 como TU® destacan la importancia de combinar precisión y velocidad en el análisis de información, mostrando que los algoritmos informáticos y la IA puede no solo complementar, sino en algunos casos superar, las metodologías tradicionales en tareas científicas complejas. Esto subraya el valor de estas herramientas en la toma de decisiones fundamentadas, particularmente en campos como la salud, donde la calidad y rapidez en la evaluación de evidencia son cruciales.

En resumen, los resultados indican que The Umbrella Collaboration® tiene la capacidad de realizar proyectos de síntesis terciaria en tiempos significativamente más cortos que las revisiones URT. TU® no solo identifica la mayoría de los Rdl que detectan las URT con un alto porcentaje de concordancia, sino que también aporta nuevos Rdl adicionales. Además, la métrica experimental R_{TU} de magnitud de los TE muestra una correlación sólida con métricas consolidadas como la d de Cohen, y su valoración de certeza, basada en análisis de sentimiento, presenta una concordancia aceptable con GRADE, un sistema reconocido internacionalmente. Del mismo modo, los análisis de dirección y significación estadística de los Rdl también muestran una correlación positiva con los resultados de la metodología tradicional. Sin embargo, estos resultados deben interpretarse con cautela debido al bajo número de valores analizados, lo que refleja las limitaciones propias de una metodología en desarrollo y la falta de comparaciones previas similares. Aunque estos hallazgos son alentadores, serán necesarias investigaciones futuras más amplias para validar estas observaciones y consolidar la posición de TU® como herramienta válida para la síntesis terciaria de evidencia.

5.3 DISCUSIÓN DEL ESTUDIO II

El E-II se centra en la percepción de los profesionales sanitarios de servicios de geriatría de hospitales universitarios de la Comunidad de Madrid, sobre la utilidad de TU®. TU® se

propone en determinadas circunstancias, como una alternativa a las URT, las cuales, si bien ofrecen un nivel elevado y reconocido de síntesis de evidencia al consolidar resultados de RS/MA, enfrentan limitaciones en cuanto a la facilidad de interpretación y utilidad clínica rápida en áreas complejas y crecientes como la geriatría (16,114,226–228). Planteamos la encuesta al grupo profesional de especialistas en Geriatría que, por su experiencia en la atención de pacientes complejos, pueden aportar una valoración completa y bien fundamentada de la herramienta experimental TU®.

Las URT, como se detalla en guías metodológicas (126,141) sintetizan múltiples revisiones sistemáticas, proporcionando una “visión panorámica” de la evidencia acumulada en un campo específico (123). Sin embargo, estas revisiones suelen depender de formatos estáticos y pueden resultar poco accesibles para la interpretación ágil, una barrera significativa para los clínicos, que requieren en ocasiones información rápida y directamente aplicable en la toma de decisiones. TU®, con su enfoque interactivo apoyado en la automatización del proceso de síntesis con un programa informático apoyado en IA, busca abordar estas deficiencias proporcionando una presentación visual más intuitiva y funcionalidades que facilitan la consulta de datos actualizados en tiempo real.

La implementación de nuevos formatos de síntesis de evidencia requiere una adaptación a las necesidades y expectativas de los profesionales de la salud para asegurar que la información sea accesible y fiable. En este sentido, resulta esencial comparar formatos novedosos, como TU®, con los métodos tradicionales, permitiendo evaluar si cumplen con los estándares de claridad y utilidad práctica que los profesionales necesitan en entornos clínicos complejos (229).

Los resultados globales obtenidos sobre la valoración global de TU® indican que la mayoría de los profesionales sanitarios encuestados consideran útil esta herramienta en su contexto clínico. La encuesta de 10 ítems utilizó una escala tipo Likert, con puntuaciones de 1 a 5, para la valoración de los ítems de la encuesta. Donde 1 representaba estar “muy en desacuerdo” con la afirmación y 5 “muy de acuerdo” (230). La puntuación media global de 4.35 sobre 5 refleja una percepción positiva respecto a la utilidad clínica de TU® por parte de la población encuestada. La mayoría de los participantes coincidieron en que la herramienta es útil en términos generales, reflejándose esta percepción en un 86.19% de respuestas positivas, donde los profesionales seleccionaron las opciones de “de acuerdo” o “muy de acuerdo” en la mayoría de los ítems evaluados. Esta alta proporción de valoraciones favorables subraya la aceptación general de TU® como un recurso útil en la práctica clínica.

Al comparar la percepción de utilidad de TU® con resultados de otros estudios sobre herramientas de automatización de decisiones en salud, encontramos una limitación inicial en la falta de estudios que analicen directamente metodologías de síntesis automatizada frente a metodologías tradicionales. Sin embargo, existen estudios que exploran la percepción de utilidad de sistemas de soporte digital de decisiones clínicas (CDSS) y herramientas de IA en diversos contextos, aunque estos se centran en decisiones clínicas, manejo de pacientes y diagnósticos, mientras que TU® se enfoca en la síntesis terciaria de

evidencia. Aun con esta diferencia, nuestros resultados de aceptación son comparables con los de dichos estudios. Por ejemplo, en el estudio de simulación de Benrimoh et al. (231) con el sistema de soporte de decisiones AIFRED, el 70% de los participantes consideraron que el modelo de IA mejoraba la comprensión del tratamiento por parte de los pacientes, mientras que el 65% percibió un aumento en la confianza en el tratamiento. Además, el 50% de los médicos valoraron la herramienta como un recurso enriquecido para discutir opciones de tratamiento con los pacientes. En otro estudio, Tanguay-Sela et al. (232), el 60% de los médicos encontró útil un CDSS para seleccionar tratamientos en pacientes con trastorno depresivo mayor, siendo particularmente valorado en el contexto de toma de decisiones compartida. Finalmente, el trabajo de Helenason et al. (233) sobre un sistema de soporte de decisiones para el diagnóstico de melanoma muestra que los médicos de atención primaria valoraron positivamente la herramienta, con una puntuación promedio de 84.8 en la Escala de Usabilidad del Sistema (234), lo que se considera una usabilidad ‘buena’.

La valoración global positiva de TU® entre los profesionales de geriatría resalta la importancia de su utilidad para la práctica clínica, aspecto que diversos estudios señalan como determinante para la adopción de nuevas herramientas en salud (235–237). El grado de aceptación de TU® entre los profesionales de nuestra encuesta y la coincidencia con factores clave identificados en la literatura apoyan la viabilidad de su implementación futura (238).

Se observa que la percepción favorable hacia TU® es uniforme en los diversos grupos profesionales, lo cual sugiere que la herramienta responde a necesidades transversales en los equipos clínicos multidisciplinares. Tanto médicos como enfermeros expresaron una satisfacción general con TU®, y sus valoraciones positivas indican que la herramienta es capaz de adaptarse a diferentes perfiles de formación y roles dentro del ámbito clínico. Las diferencias en las puntuaciones entre ambos grupos no fueron estadísticamente significativas, lo cual es un hallazgo positivo en sí mismo, ya que refuerza la versatilidad de TU® en un equipo multidisciplinario. Sin embargo, es notable que los médicos mostraron una ligera dispersión en sus respuestas, lo que podría sugerir variabilidad en las expectativas o necesidades individuales o niveles de familiaridad con la interpretación de evidencia y las metodologías de síntesis, que podría ser interesante explorar en estudios futuros.

En el análisis por género, no se identificaron diferencias significativas, lo cual también respalda la aplicabilidad de TU® en un entorno clínico diverso, con mujeres y hombres valorando de manera similar la percepción y utilidad de la herramienta. A pesar de esta uniformidad, se observa una leve tendencia en los hombres a valorar TU® con puntuaciones más altas de forma global, con una puntuación media de 4.46, discretamente por encima de la media global de toda la población, y especialmente en ítems relacionados con la satisfacción general y la evolución histórica de los datos de síntesis, lo cual puede indicar diferencias en la percepción de estos aspectos que, aunque no significativas, sugieren variabilidad en cómo se interpreta la información presentada por género.

Las diferencias más significativas se observaron en el análisis por grupos de edad, donde se identificó una tendencia clara hacia valoraciones más altas en los profesionales de mayor edad (44-65 años) en comparación con los más jóvenes (< 44 años), con una puntuación media del grupo de jóvenes es de 4.26, mientras que el de mayores es de 4.59. Los profesionales de mayor edad tendieron a valorar significativamente más alto casi todos los ítems de la encuesta. Este hallazgo puede reflejar el valor que los profesionales con mayor experiencia otorgan a una herramienta como TU®, que simplifica y organiza la evidencia de una manera práctica, buscando facilitar su aplicación en la toma de decisiones clínicas. Al mismo tiempo, la variabilidad en las respuestas de los profesionales más jóvenes podría reflejar diferentes niveles de expectativas o familiaridad con herramientas y resultados de síntesis de evidencia. Esto sugiere que este grupo podría beneficiarse de una formación más profunda en el uso de TU® y en el valor que aporta la síntesis de evidencia en la práctica clínica. Mientras que los profesionales de mayor experiencia pueden apreciar la simplicidad y claridad en la exposición de resultados de síntesis de evidencia en un campo tan complejo como la geriatría, los más jóvenes podrían mostrar mayor reticencia hacia herramientas que simplifican procesos, ya sea por una menor familiaridad o por diferencias en su percepción de la importancia y valor de la síntesis de la evidencia a nivel terciario y de las metodologías automatizadas de síntesis.

Los resultados del E-II ofrecen una primera aproximación a la percepción de TU® entre profesionales sanitarios en el ámbito de la geriatría e indican que TU® es percibido positivamente por la mayoría de los profesionales sanitarios encuestados, que, aunque la muestra es modesta y los datos provienen de un número limitado de hospitales, los datos apuntan hacia la utilidad y claridad de la herramienta experimental para sintetizar información de síntesis terciaria. La uniformidad en la valoración entre géneros y categorías profesionales sugiere que TU® se adapta bien a un entorno clínico diverso, mientras que las diferencias significativas en función de la edad refuerzan el valor de la herramienta entre los profesionales más experimentados. Las tendencias observadas en otros grupos y subgrupos, sin llegar a presentar diferencias significativas, sugieren áreas de oportunidad para futuras investigaciones.

6

LIMITACIONES, FORTALEZAS Y LÍNEAS FUTURAS DE INVESTIGACIÓN

El presente trabajo, aunque con resultados positivos hacia la metodología experimental The Umbrella Collaboration®, no está exento de limitaciones que deben ser reconocidas para interpretar adecuadamente los resultados y para guiar futuros esfuerzos de investigación. Las limitaciones pueden analizarse en dos grandes bloques: las relacionadas con el estudio comparativo de metodologías entre TU® y las URT, y aquellas derivadas de la encuesta sobre la percepción de TU® realizada a profesionales sanitarios de servicios de geriatría.

6.1 LIMITACIONES DE LA INVESTIGACIÓN

6.1.1 Limitaciones del Estudio Comparativo entre TU® y URT

Una de las principales limitaciones del estudio comparativo entre metodologías radica en el número reducido de proyectos comparativos incluidas en el análisis. La comparación se realizó utilizando únicamente ocho URT, seleccionadas aleatoriamente de un total de 36 revisiones identificadas inicialmente. Aunque esta selección buscaba obtener una muestra representativa de URT dentro del área de la geriatría, no se documentó el método específico de aleatorización, lo que introduce la posibilidad de un sesgo de selección. Además, este número bajo de URT no permite captar la diversidad completa de enfoques y variaciones metodológicas presentes en las URT, ni asegura una representación de las diversas áreas temáticas dentro de la geriatría. Sin embargo, es importante destacar que las URT seleccionadas correspondían a publicaciones recientes, todas publicadas entre 2021 y 2023. Este aspecto supone una ventaja significativa, dado que, en el campo de la síntesis terciaria, que es relativamente reciente, los estudios más antiguos tienden a presentar mayores variabilidades metodológicas y menor alineación con los estándares actuales.

La falta de cegamiento del investigador principal respecto a las URT seleccionadas constituye una limitación en el estudio comparativo entre metodologías. Dado que las URT fueron seleccionadas y analizadas previamente, por el investigador principal y único del estudio, para extraer los datos necesarios para el análisis comparativo, esto podría haber condicionado de manera inadvertida los resultados obtenidos en los proyectos realizados con TU®. En particular, el conocimiento previo de los hallazgos de las URT podría haber influido en las decisiones tomadas durante el proceso de análisis con TU®, sesgando los resultados hacia aquellos previamente conocidos. Sin embargo, aunque esta falta de cegamiento representa una limitación, es importante señalar que el volumen y la complejidad de los datos manejados en este estudio hacen que sea difícil que esta influencia haya ocurrido de manera sistemática o con un nivel de detalle suficiente como para invalidar los resultados obtenidos. Las URT incluidas contenían múltiples Rdl con distintas métricas, TE, dirección del efecto, significación estadística y valoraciones de calidad. Este alto grado de heterogeneidad dificulta que un investigador, incluso con conocimiento previo, pueda influir deliberadamente en todos los aspectos del análisis realizado con TU®. Además, la metodología de TU®, basada en un enfoque automatizado y guiado por algoritmos, introduce una capa adicional de neutralidad al proceso. Aunque el investigador tuvo acceso al diseño

y resultados de los proyectos URT, la selección y representación final de los resultados en TU® dependió en gran medida de algoritmos predefinidos y procesos automatizados, lo que limita la posibilidad de que el sesgo cognitivo haya afectado significativamente los resultados.

Otra limitación significativa se encuentra en el proceso de transformación de los TE, que fue necesario para permitir una comparación directa entre las dos metodologías. TU® utiliza la métrica R_{TU} , mientras que las URT emplean una variedad de métricas, como odds ratios, hazard ratios, diferencias de medias..., las cuales tuvieron que ser mayoritariamente transformadas a una métrica común, la d de Cohen. Aunque esta transformación permitió la comparación, también introdujo desafíos metodológicos significativos, como se evidenció en las variaciones observadas entre los resultados obtenidos en los diferentes niveles de análisis (Niveles O, A y B). Esta transformación podría limitar la capacidad de capturar la heterogeneidad subyacente de las URT, afectando la interpretación de los resultados y, potencialmente, la evaluación de la concordancia de los TE entre las dos metodologías. Por ello, el proceso de normalización añade una complejidad que requiere considerarse al interpretar los hallazgos comparativos.

Uno de los aspectos clave de la comparación entre las metodologías TU® y las URT fue la evaluación de la calidad de la evidencia en los resultados. Este análisis estuvo condicionado por la decisión metodológica de incluir únicamente revisiones que emplearan el sistema GRADE. La elección se fundamentó en el reconocimiento de GRADE como un estándar internacionalmente aceptado y ampliamente utilizado para evaluar la certeza en síntesis secundaria, terciaria y guías de práctica clínica. Esta selección limitó el número de URT incluidas en el análisis, reduciéndolo a 25 Rdl. Aunque esta muestra fue adecuada para un primer análisis, un número tan bajo de valores podría no ser suficiente para garantizar la validez estadística de los resultados ni su generalización a otros contextos de síntesis terciaria.

Otras limitaciones que presenta este estudio, pero que sobrepasan los objetivos específicos planteados, pero es relevante destacarlas como áreas clave que podrían abordarse en investigaciones futuras para optimizar y ampliar la utilidad de herramientas como TU® en el ámbito de la síntesis de evidencia. En primer lugar, no se evaluó la capacidad de TU® para mantener actualizada la información generada, a pesar de que los desarrolladores destacan esta característica de la herramienta. Validar esta capacidad es esencial, ya que la actualización constante de la evidencia es crítica en el ámbito de la salud, donde la información evoluciona rápidamente. Por otro lado, aunque se recogieron datos sobre las referencias utilizadas para cada Rdl, no se exploraron en profundidad las diferencias entre metodologías. Analizar estos aspectos podría haber aportado información valiosa sobre qué enfoques optimizan mejor la síntesis terciaria, integrando múltiples RS/MA en lugar de reproducir información de una sola fuente, y ofreciendo resultados más ajustados a la realidad. Y, por último, no se abordó el análisis económico asociado al uso de TU®. Dado que muchas herramientas avanzadas de IA generativa tienen altos costos, sería importante analizar el precio y la relación coste-beneficio de TU® en comparación con la metodología tradicional y otras herramientas informáticas en el ámbito de la síntesis de evidencia.

6.1.2 Limitaciones de la Encuesta sobre la Percepción de TU®

La encuesta dirigida a profesionales sanitarios de servicios de geriatría, diseñada para evaluar la percepción de TU® como herramienta de síntesis de evidencia, presentó diversas restricciones metodológicas que deben ser consideradas. Una de las más destacadas fue en el uso de una IA generativa (ChatGPT 4.0, OpenAI) para la redacción del cuestionario empleado en la encuesta. Si bien esta herramienta permitió estructurar y generar las preguntas con rapidez y en línea con los principios del modelo SERVQUAL y otras referencias metodológicas utilizadas, su empleo plantea interrogantes sobre la validación y la robustez del instrumento resultante. Aunque no se trata del primer caso de uso de IA generativa en el diseño de encuestas, y estudios previos han reportado que cuestionarios creados con herramientas como GPT-3 pueden ser válidos y fiables (239).

Otra limitación principal fue la representatividad de la muestra. La encuesta se realizó únicamente en cinco hospitales, lo que puede no reflejar adecuadamente la diversidad de servicios geriátricos en España ni en otros contextos internacionales. Además, los participantes de la encuesta estuvieron limitados a un grupo de las partes interesadas en salud, que son los proveedores de cuidados, dejando fuera a otros grupos de interés clave, como gestores sanitarios, pacientes, cuidadores y especialistas en metodologías de síntesis de evidencia. La exclusión de estos actores limita la comprensión integral de la utilidad y aceptación de TU® en un contexto más amplio.

El diseño de la encuesta también presentó desafíos metodológicos que pudieron introducir sesgos. Por ejemplo, la mayoría de las encuestas se completaron durante las presentaciones del proyecto, esto pudo haber inducido un sesgo de respuesta positivo, donde los participantes, influidos por el contexto favorable de la presentación, podrían haber emitido evaluaciones más positivas de lo que harían en un entorno neutral. Este fenómeno, conocido como efecto halo (240), se produce cuando una impresión general positiva impacta las respuestas relacionadas con atributos específicos, en este caso, la percepción de utilidad y eficacia de TU®. Los participantes, al estar en presencia del investigador y bajo el contexto de una presentación favorable hacia TU®, podrían haber sentido presión para emitir opiniones más positivas.

La falta de un análisis cualitativo detallado de las respuestas limita el alcance de la encuesta. Aunque se recogieron datos cuantitativos útiles, la ausencia de un análisis cualitativo dejó sin explorar matices importantes sobre la percepción de TU®, especialmente en términos de sus ventajas, limitaciones percibidas y posibles mejoras no exploradas en los ítems preestablecidos de la encuesta.

El conocimiento previo de los participantes sobre herramientas de síntesis de evidencia, incluidas las URT, no fue explorado en la encuesta. Este factor podría haber influido de manera significativa en las respuestas, especialmente si algunos participantes tenían experiencia práctica con estas metodologías tradicionales y otros no. La falta de preguntas específicas sobre este tema limita la capacidad de interpretar hasta qué punto la experiencia previa en síntesis de evidencia pudo haber sesgado las respuestas, especialmente en

cuanto a la comparación entre TU® y las URT. En futuros estudios, sería valioso incluir preguntas que exploren este tipo de antecedentes, lo que permitiría un análisis más ajustado y comprensivo de los resultados.

En conjunto, las limitaciones identificadas reflejan tanto la naturaleza exploratoria del presente trabajo como los desafíos inherentes a la validación de metodologías innovadoras como TU®. En el estudio comparativo entre TU® y las URT, cuestiones como el reducido número de URT incluidas, la falta de cegamiento del investigador y los desafíos metodológicos asociados a la normalización de métricas subrayan áreas clave para mejorar en futuros estudios. Por su parte, la encuesta sobre la percepción de TU® presentó limitaciones relacionadas con la representatividad de la muestra, el diseño del cuestionario generado por IA generativa y la ausencia de análisis cualitativos, que podrían haber permitido captar matices importantes sobre la utilidad percibida de la herramienta.

6.2 FORTALEZAS DEL ESTUDIO

La síntesis de evidencia apoyada en inteligencia artificial (IA) y desarrollos informáticos se enfrenta con importantes desafíos, como la falta de transparencia en los procesos, la dificultad para mantener los resultados actualizados, la escasa validación de las herramientas frente a métodos tradicionales y las limitaciones actuales de las herramientas automatizadas para abordar de manera rigurosa y consistente los procesos complejos que exige la síntesis de la evidencia (241). Estos problemas, particularmente estudiados en el contexto de revisiones secundarias, han generado preocupaciones sobre la precisión y la reproducibilidad de las herramientas basadas en IA. Nuestro estudio aborda algunas de estas cuestiones al validar un modelo semiautomatizado de síntesis de evidencia terciaria, comparándolo con métodos tradicionales.

La fortaleza más destacada de este estudio radica en que es posiblemente uno de los primeros trabajos para validar una herramienta automatizada, TU®, específicamente en el ámbito de la síntesis terciaria frente a la metodología tradicional representada por las URT. Aunque existen algunos estudios que han evaluado herramientas de automatización e IA en síntesis secundaria, este trabajo podría ser pionero en comparar una herramienta semiautomatizada con metodologías tradicionales en el contexto de la síntesis terciaria. El estudio aporta evidencia concreta sobre la capacidad de TU® para generar resultados coherentes y en plazos reducidos.

El estudio demuestra que TU® puede operar de manera efectiva utilizando únicamente los *abstracts* de RS/MA, lo que representa una fortaleza importante frente a otras herramientas de síntesis de evidencia que dependen del acceso al texto completo. Aunque los *abstracts* han sido tradicionalmente considerados fuentes con información limitada, insuficientes por sí solos para obtener conclusiones robustas, este estudio sugiere que, combinados con un enfoque adecuado como el de TU®, pueden ser aprovechados de manera efectiva para generar síntesis útiles y coherentes. Las herramientas de síntesis actualmente en desarrollo para las distintas fases de síntesis de evidencia suelen estar limitadas por

barreras técnicas y económicas, mientras que TU® reduce estas restricciones y amplía su aplicabilidad, incluso cuando los *abstracts* ofrecen información limitada debido a restricciones editoriales. A través del enfoque de retrieval-augmented generation, aplicada en programas de IA y también en TU®, se busca que la información buscada y usada sea relevante para la pregunta de investigación. La estrategia de generación aumentada por recuperación combina la recuperación de información relevante de una base de datos con la generación de texto basada en esa información. El sistema selecciona estratégicamente las fuentes más adecuadas a la pregunta de investigación mediante algoritmos avanzados de búsqueda, lo que ayuda a reducir errores como las "alucinaciones" típicas de los sistemas de IA generativa. TU® para usar los *abstracts* como fuente principal, selecciona únicamente las RS/MA mediante algoritmos incorporados en su programa. Este proceso se complementa con la supervisión humana, que refuerza la calidad y consistencia de los resultados. Estos hallazgos destacan cómo la combinación de algoritmos selectivos y supervisión humana, pueden superar las limitaciones de trabajar con información aparentemente limitada, generando síntesis coherentes y útiles en el ámbito de la evidencia científica.

Los resultados del estudio demuestran que una única base de datos, MEDLINE vía PubMed, puede generar resultados de síntesis comparables con revisiones terciarias tradicionales que utilizan múltiples bases de datos. Esto cuestiona el requerimiento establecido de incluir diversas fuentes y sugiere un área de investigación futura sobre la eficiencia de limitarse a bases de datos específicas.

Este estudio demuestra que TU® es capaz de generar resultados en plazos significativamente más cortos que los métodos tradicionales, una característica clave en contextos médicos donde las decisiones clínicas requieren evidencia sólida y accesible con rapidez. A diferencia de los métodos tradicionales, en los que la medición del tiempo empleado no es habitual, este estudio incluyó una evaluación específica de los tiempos de proceso, lo que supone una fortaleza significativa. Esto resulta especialmente relevante en un momento en que la necesidad de información rápida y las restricciones presupuestarias exigen herramientas eficientes.

6.3 LÍNEAS FUTURAS DE INVESTIGACIÓN PARA LA VALIDACIÓN DE TU®

Para consolidar la validez y aplicabilidad de TU® como metodología y herramienta de síntesis terciaria de evidencia, se plantean diversas líneas futuras de investigación orientadas a superar las limitaciones identificadas y explorar nuevas oportunidades de mejora.

6.3.1 Ampliación del tamaño de las muestras

Una de las prioridades en futuros estudios de validación de la herramienta será incrementar el número de URT incluidas en futuros estudios comparativos, con el objetivo de aumentar la representatividad y robustez de los hallazgos. Esto es particularmente relevante para URT que emplean GRADE como sistema de evaluación de certeza, dado que este marco metodológico ha demostrado ser un estándar de alta calidad ampliamente aceptado.

La inclusión de un mayor número de URT permitirá explorar con mayor profundidad las concordancias y discrepancias entre TU® y las URT, proporcionando una base más sólida para evaluar la capacidad de TU® de replicar los hallazgos obtenidos por las metodologías tradicionales.

6.3.2 Comparaciones con diseño ciego

Una prioridad fundamental será realizar estudios comparativos entre TU® y las URT implementando un diseño ciego. Esto implica que los investigadores que ejecuten los proyectos con TU® no tengan conocimiento previo de los resultados reportados por las URT, limitándose únicamente a conocer la pregunta de investigación planteada. Este enfoque reducirá significativamente el riesgo de sesgo por conocimiento previo y proporcionará una evaluación más objetiva de la capacidad de TU® para replicar y ampliar los hallazgos de las URT.

6.3.3 Evaluación con diversos perfiles de revisores

Otra línea importante será evaluar la capacidad de diferentes perfiles de revisores humanos para utilizar TU® de manera efectiva. Esto incluiría geriatras, médicos de otras especialidades, expertos en metodología y sanitarios no médicos. Los estudios buscarán determinar la concordancia de los resultados obtenidos por estos revisores con los de un revisor humano experto en la herramienta y analizar las posibles variaciones en función de la experiencia y el conocimiento de cada grupo. Este análisis permitirá identificar si TU® mantiene su robustez y facilidad de uso en manos de usuarios con diferentes niveles de formación.

6.3.4 Ampliación del ámbito de aplicación

Para explorar la versatilidad de TU®, se deberían realizar proyectos en áreas distintas a la geriatría, incluyendo otras especialidades médicas, áreas no médicas e incluso en campos ajenos a la medicina. Estas áreas deberán contar con un volumen sustancial de RS/MA para garantizar la aplicabilidad de la metodología y tener URT para permitir realizar las comparaciones. Este enfoque permitirá evaluar si TU® es igualmente eficaz en contextos con diferentes patrones de publicación y características metodológicas, ampliando su alcance y relevancia.

6.3.5 Encuestas de percepción a grupos diversos de las partes interesadas

Además de los profesionales sanitarios, se deberían diseñar encuestas de percepción dirigidas a otros grupos de partes interesadas, como gestores sanitarios, expertos en metodología de síntesis de evidencia, pacientes y cuidadores. Estas encuestas buscarán explorar la percepción de utilidad, facilidad de uso y aplicabilidad práctica de TU® en diferentes contextos, ampliando la comprensión sobre su aceptación y potencial impacto en la práctica clínica y la toma de decisiones en salud.

6.3.6 Validación de capacidades no exploradas

Es fundamental explorar tres áreas clave que no fueron objeto de evaluación directa pero que son esenciales para consolidar la aplicabilidad de TU®.

La capacidad de TU® para mantener los resultados actualizados en tiempo real, una característica importante destacada por sus desarrolladores debe validarse en estudios específicos. Este análisis es especialmente relevante en un contexto donde la evidencia en salud evoluciona rápidamente y la actualización constante de la información es fundamental para la toma de decisiones informadas.

Es necesario profundizar en cómo las referencias seleccionadas para cada Rdl afectan la calidad y validez de las síntesis generadas. Esto incluye analizar las diferencias entre metodologías que optimicen la integración de información procedente de múltiples RS/MA, en lugar de limitarse a reproducir los hallazgos de una única fuente. Tal análisis podría ofrecer una mejor comprensión del potencial de TU® para alcanzar resultados más completos y ajustados a la realidad.

Finalmente, se debe evaluar el coste económico asociado al uso de TU® en comparación con la metodología tradicional y otras herramientas avanzadas de IA generativa. Este análisis permitirá situar a TU® en el contexto de las revisiones terciarias tradicionales y las tecnologías actuales de IA en síntesis de evidencia, explorando su viabilidad económica y su relación coste-beneficio para facilitar su implementación práctica en diferentes contextos.

Estas líneas futuras de investigación no solo buscan fortalecer la validación de TU® como metodología y herramienta de síntesis terciaria, sino también posicionarla como un recurso adaptable y de amplio alcance en el ecosistema de la síntesis de evidencia. Con un enfoque en la mejora continua y la exploración de nuevas aplicaciones, TU® tiene el potencial de convertirse en un referente en la integración de evidencia científica en la toma de decisiones, tanto en medicina como en otros campos del conocimiento.

7

CONCLUSIONES

1. The Umbrella Collaboration® logra replicar la mayoría de los resultados de interés identificados por las umbrellla tradicionales con un alto nivel de concordancia, además de ampliar el alcance al identificar resultados de interés adicionales.
2. La métrica R_{TU} de The Umbrella Collaboration® muestra una fuerte correlación con medidas estadísticas consolidadas, como la d de Cohen, reforzando su capacidad para evaluar la magnitud de intervenciones y riesgos.
3. El análisis de sentimiento automatizado de The Umbrella Collaboration® presenta una concordancia aceptable con el sistema GRADE, lo que respalda su capacidad para valorar la certeza de la evidencia.
4. Se observó un alto grado de concordancia entre The Umbrella Collaboration® y las umbrellla tradicionales en la dirección y significación estadística de los resultados de interés, consolidando la robustez de los resultados.
5. The Umbrella Collaboration® demuestra, ser una metodología con una herramienta válida y robusta para la síntesis terciaria de evidencia, con resultados comparables a las revisiones umbrellla tradicionales.
6. The Umbrella Collaboration® destaca por su rapidez, completando proyectos en horas frente a los meses requeridos por las umbrellla tradicionales.
7. Los profesionales sanitarios de los diferentes servicios de Geriátría de la Comunidad de Madrid de forma uniforme sin diferencias estadísticamente significativas valoran favorablemente a The Umbrella Collaboration® en términos de utilidad y claridad, destacando su potencial para integrar evidencia en la práctica clínica.
8. Los profesionales mayores de los servicios de geriatría valoraron significativamente de manera más positiva The Umbrella Collaboration® en comparación con sus colegas más jóvenes, lo que sugiere interesantes diferencias intergeneracionales en la percepción de su utilidad.
9. Este estudio es pionero en la validación de una herramienta informática semiautomatizada apoyada en inteligencia artificial, para la síntesis terciaria de evidencia en comparación con la metodología tradicional. Aunque presenta ciertas limitaciones, establece una base sólida para futuras investigaciones. Es necesario ampliar las validaciones en contextos y muestras más diversos para consolidar la utilidad y aplicabilidad de The Umbrella Collaboration®.

8

ANEXOS

ANEXO 1

Metodología de Umbrella Reviews.

Ver tabla Tabla 1

1. Título e Información de Autoría:

a. El primer paso en la elaboración de un trabajo de síntesis terciaria de la evidencia comienza con la formulación de un título claro y la provisión de información detallada sobre los autores. El título debe ser informativo y reflejar de manera precisa el tema de la revisión, permitiendo a los lectores comprender de inmediato que el objetivo de la síntesis es responder a una pregunta de investigación específica. Algunos autores recomiendan la inclusión explícita del término de “umbrella review” u otros sinónimos que permitan una rápida y fácil identificación del tipo de estudio. JBI recomienda que el título incluya alguno de los elementos que definan la población, intervención, comparador y resultados según la fórmula PICO. La inclusión de detalles sobre los autores, como las afiliaciones institucionales y la información de contacto, es esencial para garantizar la transparencia del proceso de investigación. Es fundamental que el equipo de autores sea multidisciplinario y que abarquen diversas competencias necesarias para la investigación.

2. Introducción:

a. La sección de introducción de una URT debe ser lo suficiente completo e integral para contextualizar el tema de estudio y subrayar su relevancia dentro del campo académico y clínico. Es esencial que esta sección proporcione una justificación sólida para la realización de la revisión. La introducción debe considerar una serie de preguntas que invitan a realizar la URT, como son la relevancia clínica del tema, la volatilidad y variaciones de resultados del campo de estudio o la presencia de numerosas RS/MA sobre el mismo tema. Esta justificación debe identificar claramente lo conocido sobre el tema de investigación y de los posibles vacíos de conocimiento dentro de la literatura existente, especialmente en áreas sujetas a controversia, propensas a sesgos, o que han sido objeto de múltiples RS/MA previas. La introducción debe avalar la realización del URT y esbozar cómo la URT abordará los resultados y los vacíos de conocimientos, contribuyendo significativamente al avance del conocimiento en el área temática.

3. Fuentes informativas (bases de datos):

a. Es esencial en la elaboración de URT que la búsqueda sea lo más amplia posible, por lo que seleccionar múltiples fuentes para buscar información es fundamental para evitar introducir sesgos en los resultados finales. No existe consenso sobre el número óptimo de bases de datos que se debe consultar (242), la selección y por ello el número de bases de datos incluidos en la búsqueda va a depender del tema de investigación. Para intervenciones en salud, MEDLINE y EMBASE son las bases de datos más frecuentemente utilizadas. Mientras que se acepta la inclusión de bases de datos especializadas como la Cochrane Database of Systematic Reviews, la JBI Database of Systematic Reviews and Implementation Reports, la Database of Abstracts of Reviews of Effects y el registro PROSPERO. En función del tema de estudio se podrán incluir bases de datos específicos como PsycINFO, AMED, PEDro, MANTIS, CINHAI... (25).

4. Definición Clara de la Pregunta de Investigación:

a. La definición precisa de la pregunta de investigación, y sus objetivos y alcance, son fundamentales en la metodología de las URT. Aunque las URT están abiertas a poder abarcar temas amplios de investigación, la pregunta principal de investigación de las URT debe ser clara y específica de manera que permita delimitar con exactitud las poblaciones, intervenciones, comparadores y resultados de interés (PICOS).

5. Registro y Protocolo Predefinido:

a. La inscripción del protocolo de una URT en un registro como PROSPERO se identifica como un paso importante para garantizar la transparencia y credibilidad de la investigación. Este acto de registro, al detallar el enfoque estructurado del estudio se presentan las bases metodológicas y temáticas de manera clara. Dicho registro facilita evitar la redundancia en la investigación y permite que la comunidad científica verifique la premeditación y rigurosidad del enfoque de síntesis, fortaleciendo así la confianza en los hallazgos de la URT. Se acepta que se puedan introducir modificaciones sobre el protocolo inicial, en función de los resultados que se presenten a lo largo del proceso de elaboración, estas modificaciones post-hoc deben quedar claramente documentadas en el documento final.

6. Desarrollo de Criterios de Inclusión:

a. La selección de RS/MA para su inclusión en una URT se fundamenta en criterios de inclusión preestablecidos, que deben ser claros y precisos para asegurar la relevancia y coherencia de la evidencia analizada. Estos criterios abarcan aspectos como el tipo de estudios, las poblaciones, las intervenciones, comparadores, medidas de resultado, así como el diseño de estudio que será incluido, que es la RS/MA, especificando si los estudios primarios incluidos son ensayos clínicos aleatorios o estudios observacionales.

7. Estrategia de Búsqueda Exhaustiva:

a. La elaboración de una estrategia de búsqueda exhaustiva y sistemática constituye un pilar fundamental en la metodología de síntesis terciaria de la evidencia. Es fundamental en esta fase contar en el equipo de investigación con un documentalista o especialista en información para dirigir la búsqueda de manera que sea lo más precisa posible sin incurrir en utilización no productiva de recursos ni llegar a la futilidad bibliográfica (243). La precisión en la definición de términos de búsqueda y la consideración de posibles restricciones por fecha, idioma y estado de publicación son esenciales para acceder a un espectro completo de estudios pertinentes. Es imperativo utilizar métodos explícitos y reproducibles para la identificación de RS/MA que satisfagan los criterios de inclusión predefinidos. La estrategia de búsqueda debe ser diseñada meticulosamente para abarcar todas las RS/MA relevantes, detallando las bases de datos consultadas, términos de búsqueda específicos y las estrategias empleadas para identificar estudios adicionales como literatura gris o informes que no estuviesen publicados de forma comercial. Este enfoque riguroso no solo promueve la integridad y la confiabilidad de los hallazgos de la URT, sino que también facilita la reproducibilidad de la búsqueda, permitiendo que otros investigadores puedan seguir y verificar el proceso.

8. Selección Rigurosa de Estudios:

a. Para garantizar la exactitud y transparencia de los resultados de la síntesis terciaria de la evidencia, es esencial realizar un proceso claro y reproducible para valorar todos los estudios encontrados en la búsqueda para determinar si cumplen con los criterios de

inclusión preestablecidos. Es fundamental detallar los procedimientos empleados para evaluar los estudios identificados para su inclusión, especificando los criterios de selección y delineando los mecanismos para resolver cualquier discrepancia entre los revisores.

9. Características de los estudios incluidos:

a. La presentación de los datos descriptivos de las RS/MA incluidos en una URT permite analizar y comprender los resultados de la evidencia sintetizada. Este proceso meticuloso implica la extracción detallada de información relevante de cada RS/MA incluida, abarcando desde los datos básicos de la propia RS/MA y de los estudios primarios incluidos, tales como: autor, año, fecha de la última actualización, número de estudios primarios y participantes, los objetivos, métodos, criterios de inclusión y estrategia de búsqueda, así como detalles de la población, intervención, controles y resultados. Es importante reseñar que los autores deben a la vez considerar cuáles de estas características son potencialmente relevantes, priorizando la información que permita una evaluación exhaustiva del cuerpo de evidencia. Toda esta información se puede presentar en una tabla para facilitar la comprensión del lector.

10. Recolección de Datos de RS Incluidas:

a. La metodología de las URT implica la extracción rigurosa de datos críticos de las RS/MA incluidas, que no solo incluye la extracción de datos descriptivos de las RS/MA y los estudios primarios incluidos como se mencionó en el apartado anterior, sino que debe abarcar la recogida de datos adecuados para la valoración de riesgo de sesgos, la valoración de la calidad metodológica y de la certeza de la evidencia. La mayoría de las guías y organizaciones recomiendan que esta fase sea realizada de manera independiente por dos revisores.

b. Se recomienda emplear formularios estandarizados para facilitar la lectura y comprensión de los resultados, y que otros autores puedan reproducir la metodología para sintetizar estos resultados. En función de si las revisiones sistemáticas tengan o no metaanálisis se recogerán distintos tipos de datos si existen, en caso de no estar presentes en las RS/MA y limite la capacidad de síntesis del URT se limitará a justificarlo en los resultados. Otra opción es recurrir a los estudios primarios, que puede hacer dudar de la justificación de realizar la síntesis terciaria, y no plantarse hacer en su lugar una nueva RS/MA.

11. Evaluación del Solapamiento (*Overlap*):

a. La presencia, casi inevitable, de solapamiento de estudios primarios en las RS/MA incluidas en las URT representa un desafío metodológico que requiere un análisis detallado por parte de los autores. Este fenómeno se presenta cuando dos o más revisiones sistemáticas abordan la misma pregunta de investigación e incluyen algunos de los mismos estudios primarios. El solapamiento introduce dos retos importantes en las URT. Por un lado, en el caso de agregar datos de resultados de interés para obtener un resultado de interés global, aquellos estudios primarios que aparecen en más de una RS/MA tendrán un exceso de peso en esos resultados finales. Y, por otro lado, las URT con un grado elevado de solapamiento entre las RS/MA son más complejas de realizar, y precisan de más tiempo y recursos, no siempre disponibles para este tipo de trabajos de síntesis.

b. Existen dos maneras habituales de abordar este solapamiento en función de los objetivos de las URT. Si el objetivo es explorar nuevas preguntas de investigación sobre sub-poblaciones de participantes incluidos en las RS/MA existentes, resulta esencial re-extraer y re-analizar datos de un conjunto seleccionado de RS/MA sin solapamiento

para garantizar la precisión y la relevancia de los hallazgos. Lo que obliga a identificar el solapamiento al detalle, aumentando la complejidad de la URT. En cambio, si la URT busca proporcionar un panorama descriptivo completo del estado actual de la evidencia sobre un tema, está justificado incluir los resultados de todas las RS/MA con independencia del grado de solapamiento. De manera que este abordaje simplifica el proceso de elaboración de la URT.

c. En el caso de decidir valorar el grado de solapamiento de estudios primarios incluidos en las RS/MA incluidas en la URT, se recomienda crear una matriz de citaciones para demostrar visualmente el grado de solapamiento, facilitar la descripción del número y tamaño de solapamiento de estudios primarios y calcular el "área cubierta corregida" que aporta una medida de la extensión de solapamiento entre estudios primarios de las RS/MA (139).

12. Evaluación de la Calidad Metodológica y Riesgo de Sesgos:

a. La evaluación de la calidad metodológica y el riesgo de sesgos en las RS/MA incluidas y sus estudios primarios es un componente crítico en la metodología de síntesis terciaria de la evidencia. La aplicación de herramientas y criterios estandarizados, tales como AMSTAR 2 y ROBIS, facilita una evaluación rigurosa y objetiva de estos aspectos, asegurando así la fiabilidad y la validez de las evidencias recopiladas en una URT. Este proceso meticuloso no solo evalúa la rigurosidad con la que se realizaron y reportaron las RS/MA, sino que también identifica posibles desviaciones sistemáticas que podrían afectar la interpretación de los hallazgos de la URT.

13. Análisis y presentación de los Resultados:

a. Existen dos enfoques principales para el análisis y presentación de datos de resultados: la presentación resumida de datos en su forma original según figuran en las RS/MA (tanto de los datos en formato narrativo como de los datos metaanalizados) y el reanálisis de los datos de los resultados de interés de las RS/MA incluidas en la URT, para ofrecer nuevas interpretaciones o perspectivas. La selección entre estos enfoques depende intrínsecamente de los objetivos de la URT, la naturaleza de la pregunta de investigación y la heterogeneidad presente en las RS/MA incluidas. En el caso de que el objetivo de la URT sea responder a una nueva pregunta no considerada en las RS/MA sobre una subpoblación de participantes incluidos en las RS/MA de la URT, tiene sentido el reanálisis de los datos referidos exclusivamente a esa subpoblación. Mientras que, si el objetivo de la URT es presentar y describir el cuerpo total de la evidencia sobre un tema, puede ser más apropiada la primera opción de presentar los resultados de forma más o menos resumida de todas las RS/MA según están referidas en ellas (136,244).

b. El reanálisis también tiene sentido cuando se desea la estandarización de medidas de tamaños de efecto utilizados en las RS/MA originales, permitiendo una presentación uniforme de resultados que facilita su comparación y la síntesis global de evidencias. Un tercer motivo para realizar el reanálisis de los datos originales de las RS/MA es poder examinar datos que no fueron metaanalizados en las RS/MA originales, ampliando el alcance de la revisión en paraguas y posibilitando el descubrimiento de nuevas perspectivas o la confirmación de hallazgos existentes.

c. En la síntesis terciaria de la evidencia, se pueden adoptar diversas estrategias para presentar los datos de resultados, buscando siempre maximizar la claridad y la comprensión de los resultados. Una estrategia consiste en la elaboración de resúmenes

narrativos agrupados o no según medidas similares de resultados de interés, posiblemente complementados con tablas, que proporcionan una visión general accesible y detallada de los resultados. Alternativamente, se puede optar por categorizar los resultados según su relevancia clínica o efectividad, distinguiendo entre hallazgos "clínicamente importantes" o "no clínicamente importantes", y entre intervenciones "eficaces" o "ineficaces". Esta clasificación se enfoca más en la aplicación práctica que en la significancia estadística. Otra opción es la presentación de los datos dentro de un marco conceptual nuevo que crucen intervenciones con resultados, mostrando la cantidad de estudios y participantes que aportan datos para cada resultado y la dirección y significación del efecto observado.

14. Evaluación Certeza Resultados:

a. La evaluación de la certeza de la evidencia en los resultados de interés es un componente esencial en el proceso de síntesis terciaria, fundamental para establecer la confianza que se pueda tener en las estimaciones del efecto de las intervenciones analizadas y para fundamentar recomendaciones basadas en la evidencia. La herramienta más recomendada, y que tiene cada día más predicamento en la síntesis terciaria, es la herramienta GRADE (148), que permite una valoración rigurosa y reconocida de la certeza de la evidencia de cada resultado de interés. Es clave valorar de manera sistemática y detallada la certeza de la evidencia obtenida, tanto de los datos de resultados como de las RS/MA incluidas. Este proceso implica no solo extraer y reportar evaluaciones GRADE ya realizadas en las RS/MA, sino también llevar a cabo evaluaciones propias a nivel de la URT cuando sea necesario.

15. Limitaciones de la síntesis terciaria

a. Al igual que cualquier publicación científica, al realizar un estudio de síntesis terciaria de evidencia, es fundamental identificar y discutir las limitaciones inherentes a la metodología de las URT, pero también a las limitaciones de las RS/MA y de los estudios primarios que forman parte de estos. Es crucial reportar estas limitaciones para ofrecer una perspectiva equilibrada, reconociendo los posibles sesgos o factores que podrían afectar la interpretación y aplicabilidad de los hallazgos (37,245).

16. Discusión implicación de resultados

a. La discusión sobre las implicaciones de los hallazgos de una URT debe abordar su relevancia para la investigación futura, las políticas sanitarias y la práctica clínica. Es esencial considerar cómo estos resultados pueden influir en los proveedores de salud, gestores sanitarios y pacientes, destacando la importancia de aplicar la evidencia sintetizada para mejorar la toma de decisiones basada en evidencia a nivel de salud pública y de la población general de forma individual. Además, se debe reflexionar sobre las necesidades de investigación futura que surgen de la revisión, identificando lagunas en el conocimiento existente y sugiriendo áreas prioritarias para estudios adicionales.

17. Conclusions:

a. Resumir los hallazgos de una síntesis terciaria implica no solo la presentación concisa de los datos y resultados obtenidos sino también una interpretación cuidadosa de estos. Las conclusiones derivadas deben estar fundamentadas en la calidad de los estudios primarios y las RS/MA incluidas, así como en la rigurosidad de los métodos empleados para el análisis terciario de los datos. Es crucial ejercer precaución al realizar comparaciones indirectas entre diferentes intervenciones, para evitar conclusiones erróneas. Este enfoque asegura que las interpretaciones y conclusiones sean válidas, fiables y reflejen fielmente la evidencia analizada.

ANEXO 2

Metodología de la transformación de las métricas a la d de Cohen.

En la comparación de los tamaños de efecto (TE) de los resultados de interés (RdI) obtenidos mediante la metodología TU® y las Umbrella Reviews Tradicionales (URT), nos enfrentamos con el desafío de que los TE se expresan en diferentes métricas estadísticas. Mientras que las URT emplean una variedad de métricas como la Diferencia de Medias Estandarizada (SMD), la Diferencia de Medias (MD), el Riesgo Relativo (RR), el Odds Ratio (OR) y el Hazard Ratio (HR), la metodología TU® utiliza una métrica propia denominada R_{TU} , desarrollada ad hoc para este proyecto. Para facilitar una comparación directa entre estas metodologías, decidimos transformar las métricas empleadas en las URT a la d de Cohen, mientras que mantenemos la métrica R_{TU} para los análisis generados por TU®. Así, la comparación se centra en el contraste entre la d de Cohen, representando a las URT, y R_{TU} como métrica específica de TU®, permitiendo una interpretación consistente dentro de las limitaciones inherentes a esta nueva métrica.

La elección de la d de Cohen como métrica común se basa en varias consideraciones teóricas y prácticas, sustentadas por una extensa literatura académica.

En primer lugar, la d de Cohen es ampliamente reconocida y utilizada en el ámbito de las ciencias del comportamiento y la investigación clínica debido a su capacidad para proporcionar una medida estandarizada del TE. Como describe Cohen en su obra "Statistical Power Analysis for the Behavioral Sciences" (145), la d de Cohen permite traducir los resultados de estudios en términos de magnitud del efecto, lo que facilita su interpretación y comparabilidad. Esta métrica convierte las diferencias entre grupos o las relaciones de asociación en una escala común, lo que es esencial cuando se integran y comparan resultados de estudios que utilizan diferentes enfoques de metodología estadística y métricas.

La utilización de la d de Cohen también está respaldada por su simplicidad y eficacia en la comunicación de resultados a audiencias no especializadas. Jacobs Cohen ((145) enfatiza que una de las fortalezas de la d de Cohen es su capacidad para expresar el TE en términos que son fácilmente comprensibles, incluso para aquellos sin una formación estadística avanzada. Este aspecto es crucial para nuestra investigación, ya que uno de los objetivos de TU® es hacer que los resultados sean accesibles y comprensibles para todas las partes interesadas, incluidas las que no tienen una formación técnica.

Además, la d de Cohen ha sido la métrica elegida en numerosos estudios debido a su capacidad para estandarizar medidas de efecto diversas, como el OR, RR y HR. En el capítulo 7 de "Introduction to Meta-Analysis" (246) explican detalladamente cómo la d de Cohen puede ser derivada de diferentes medidas de efecto, proporcionando fórmulas y procedimientos específicos para estas transformaciones. Esta flexibilidad y aplicabilidad generalizada hacen que la d de Cohen sea una elección ideal para nuestra investigación, donde las medidas de efecto originales varían considerablemente.

El artículo de Zhang (247) sobre la relación entre el RR y el OR también subraya la utilidad de convertir estas métricas a la d de Cohen. Ellos explican que, especialmente cuando la incidencia del evento es alta, las medidas como el OR pueden sobreestimar el RR, pero al

convertirlas a la d de Cohen, se logra una estimación más consistente y comparable del TE. Este tipo de ajuste es fundamental para asegurar que nuestras comparaciones entre TU® y URT sean precisas y fiables.

Finalmente, el trabajo de Ellis (248) en "The Essential Guide to Effect Sizes" refuerza la utilidad de la d de Cohen en la interpretación de los TE. Ellis proporciona una guía clara sobre cómo interpretar los valores de d , categorizándolos en efectos triviales, pequeños, medianos y grandes. Esta categorización no solo facilita la comprensión de los resultados, sino que también estandariza la interpretación de la magnitud del efecto, lo que es esencial para comunicar los hallazgos de manera efectiva.

En resumen, la elección de la d de Cohen como métrica común para la unificación de los TE en nuestra investigación está justificada por su reconocimiento y aceptación en la literatura académica, su capacidad para estandarizar diversas métricas de efecto, y su facilidad de interpretación y comunicación.

A continuación, se detalla el proceso de transformación de cada métrica a la d de Cohen.

El primer paso en nuestra metodología fue revisar el capítulo 7 del libro "Introduction to Meta-Analysis" de Borenstein et al. (249). En este capítulo, titulado "Converting Among Effect Sizes", se proporciona un marco teórico y práctico para la conversión de diferentes métricas de efecto a un tamaño de efecto común, como la d de Cohen. Este capítulo es fundamental para entender cómo podemos comparar resultados que originalmente se expresan en diferentes métricas, como OR, RR, HR y SMD. La transformación de estas métricas permite una interpretación uniforme y facilita la comparación entre estudios que originalmente reportan sus resultados en formatos distintos.

Para transformar los OR y RR a la d de Cohen, seguimos la fórmula básica sugerida por Borenstein et al. (249):

$$d = \frac{\ln(OR)}{1.81}$$

Esta fórmula se basa en la relación entre el logaritmo natural del OR y una constante, que proporciona una estimación del TE comparable a la d de Cohen. Para simplificar los cálculos y evitar errores, utilizamos una constante alternativa derivada de la raíz cuadrada de 3 dividida por pi ($(\sqrt{3}/\pi \approx 0.551)$). Los pasos son sencillos: primero, se calcula el logaritmo natural (LN) del OR, y luego se multiplica el resultado por la constante 0.551. Por ejemplo, para un OR = 2, el LN es 0.639 y al multiplicarlo por 0.551 obtenemos 0.352. Este proceso se repite para diferentes valores de OR, lo que nos permite transformar estos tamaños de efecto en la d de Cohen para la comparación entre resultados de ambas metodologías.

Para la conversión de los HR a d de Cohen, seguimos un proceso similar. Según el artículo de Zhang et al. (247), la fórmula para convertir HR a d de Cohen es (250):

$$d = \frac{\ln(HR) \times 2}{\sqrt{3}}$$

Por ejemplo, para un HR = 2, el LN es 0.693 y al multiplicarlo por 1.155 obtenemos 0.800.

La transformación de la SMD a d de Cohen se realiza utilizando el intervalo de confianza (IC) de los outcomes. El IC es igual a la diferencia de medias estandarizadas ± 1.96 veces el error estándar (SE). Los pasos son: calcular la amplitud del IC (diferencia entre el valor mayor y el menor), dividir la amplitud por 1.96 para obtener el SE, y asumir que SE = SD (desviación

estándar), dado que no tenemos el tamaño muestral. Por ejemplo, para un estudio con SMD 1.782 (0.722 a 2.842), la amplitud es 2.12, el SE es 1.08 y la d de Cohen es 1.65.

Finalmente, la interpretación de los valores de d de Cohen se realiza según la Tabla 2.1 de Ellis (248)(Ellis, 2010. Página 41). Los valores se interpretan como: de 0 a 0.2 (trivial), de 0.2 a 0.5 (pequeño), de 0.5 a 0.8 (medio) y > 0.8 (grande). Este marco nos permite traducir los TE en interpretaciones coloquiales, claras y comprensibles para todas las partes interesadas. Las clases de TE de Cohen tienen dos puntos fuertes. Primero, son fáciles de entender. Solo se necesita comparar los valores con sus umbrales para obtener una interpretación coloquial inmediata del resultado. Segundo, aunque son arbitrarios, están lo suficientemente fundamentados en la lógica como para que Cohen esperara que “sean considerados razonables por personas razonables” (145). Al decidir los límites para las tres clases de tamaño originales (pequeño, mediano y grande), Cohen comenzó definiendo un efecto medio como uno que es “visible a simple vista para un observador atento” (145). Para usar su ejemplo, un efecto medio es equivalente a la diferencia de altura entre niñas de catorce y dieciocho años, que es aproximadamente una pulgada. Luego definió un efecto pequeño como uno que es menor que un efecto medio, pero mayor que un efecto trivial. Los efectos pequeños son equivalentes a la diferencia de altura entre niñas de quince y dieciséis años, que es aproximadamente media pulgada. Finalmente, un efecto grande se definió como uno que estaba tan por encima de un efecto medio como un efecto pequeño estaba por debajo de él. En este caso, un efecto grande es equivalente a la diferencia de altura entre niñas de trece y dieciocho años, que es un poco más de una pulgada y media.

En conclusión, la transformación de diferentes métricas de efecto a la d de Cohen ha permitido una comparación directa y coherente entre los resultados obtenidos por la metodología TU® y las URT.

ANEXO 3

Referencias Excluidas y Motivos de Exclusión en la Búsqueda Inicial

Nº	Autor año	Título	Motivo exclusión
1	Stites, 2023 (251)	A systematic review of measures of gender and biological sex: Exploring candidates for Alzheimer's disease and related dementias (AD/ADRD) research.	1
2	Sousa-Pinto, 2022 (252)	Real-world data using mHealth apps in rhinitis, rhinosinusitis and their multimorbidities.	2
3	Wang, 2022 (253)	The effectiveness of case management for cancer patients: an umbrella review.	2
4	Hoffmann, 2022 (254)	Exercise Reduces the Number of Overall and Major Osteoporotic Fractures in Adults. Does Supervision Make a Difference? Systematic Review and Meta-Analysis	1, 2
5	Lei, 2022 (255)	Effectiveness of mHealth/eHealth interventions on obesity treatment: a protocol for umbrella review of meta-analyses.	2, 3
6	Kast, 2022 (256)	Effects of different exercise intensity on bone mineral density in adults: a comparative systematic review and meta-analysis.	1, 2
7	Chen, 2022 (257)	Vitamin C Intake and Cancers: An Umbrella Review.	2
8	Tannou, 2022 (258)	Effectiveness of smart living environments to support older adults to age in place in their community: an umbrella review protocol.	3
9	Martínez-García, 2022 (259)	. Incidence of Arterial Hypertension in People With Periodontitis and Characterization of the Oral and Subgingival Microbiome: A Study Protocol.	1, 2, 3
10	Chauhan, 2022 (260)	Prevalence and determinants of activity of daily living and instrumental activity of daily living among elderly in India.	1
11	Kroke, 2022 (261)	Dietary protein intake and health-related outcomes: a methodological protocol for the evidence evaluation and the outline of an evidence to decision framework underlying the evidence-based guideline of the German Nutrition Society.	1, 2, 3
12	Mages, 2021 (262)	Exercise Effects on Bone Mineral Density in Men.	1, 2
13	Moretto, 2021 (263)	Reversible facial nerve paralysis in a cat suspected to be associated with systemic hypertension.	1, 2
14	Zitzmann, 2022 (264)	The effect of different training frequency on bone mineral density in older adults. A comparative systematic review and meta-analysis.	1
15	Sawyer, 2021 (265)	Dynamics of the complex food environment underlying dietary intake in low-income groups: a systems map of associations extracted from a systematic umbrella literature review.	1,2
16	Michalski, 2020 (266)	Magnetic resonance imaging of the knee joint in juvenile idiopathic arthritis.	1, 2
17	Lauretani, 2020 (267)	Comprehensive Model for Physical and Cognitive Frailty: Current Organization and Unmet Needs.	1
18	Sikkes, 2021 (268)	ISTAART Non-pharmacological Interventions Professional Interest Area; Bahar-Fuchs A. Toward a theory-based specification of non-pharmacological treatments in aging and dementia: Focused reviews and methodological recommendations.	1

Nº	Autor año	Título	Motivo exclusión
19	Florisson, 2021 (269)	Are older adults insufficiently included in clinical trials?-An umbrella review.	4
20	Falandry, 2020 (270)	Interventions to improve physical performances of older people with cancer before complex medico-surgical procedures: Protocol for an umbrella review of systematic reviews and meta-analyses.	3
21	Gremese, 2020 (271)	Basic immunology may lead to translational therapeutic rationale: SARS-CoV-2 and rheumatic diseases.	1, 2
22	Soysal, 2021 (272)	Handgrip strength and health outcomes: Umbrella review of systematic reviews with meta-analyses of observational studies.	2
23	Solmi, 2020 (273)	Factors Associated With Loneliness: An Umbrella Review Of Observational Studies.	2
24	aretzke, 2020 (274)	Role of Vitamin D in Preventing and Treating Selected Extraskkeletal Diseases-An Umbrella Review.	2
25	McGrath, 2020 (275)	What are the association patterns between handgrip strength and adverse health conditions? A topical review.	1, 2
26	Biondi-Zoccai, 2020 (276)	Atherothrombosis Prevention and Treatment with Anti-interleukin-1 Agents.	1
27	Wangmo, 2019 (277)	Ethical concerns with the use of intelligent assistive technology: findings from a qualitative study with professional stakeholders.	1, 2
28	Roller-Wirnsberge, 2019 (278)	Massive open online courses (MOOCs) for long-distance education in geriatric medicine across Europe : A pilot project launched by the consortium of the project "Screening for Chronic Kidney Disease among Older People": SCOPE project.	1
29	Soysal, 2019 (279)	Orthostatic hypotension and health outcomes: an umbrella review of observational studies.	2
30	MacEntee, 2019 (280)	A typology of systematic reviews for synthesising evidence on health care.	1
31	Kivnick, 2020 (281)	"Who's Been Putting Socks in My Drawer?" Narrative Case Study of an Elder Role Model.	1
32	Grabovac, 2020 (282)	Human Immunodeficiency Virus Infection and Diverse Physical Health Outcomes: An Umbrella Review of Meta-analyses of Observational Studies.	2
33	Pescatello, 2019 (283)	Physical Activity to Prevent and Treat Hypertension: A Systematic Review.	1, 2
34	Chen, 2020 (284)	A Self-Reliant Umbrella: Defining Successful Aging Among the Old-Old (80+) in Shanghai.	4
35	Sharka, 2019 (285)	Oral health-related quality of life and satisfaction of edentulous patients using conventional complete dentures and implant-retained overdentures: An umbrella systematic review.	2
36	Bangsbo, 2019 (286)	Copenhagen Consensus statement 2019: physical activity and ageing.	1
37	Smith, 2019 (287)	Telomere length and health outcomes: An umbrella review of systematic reviews and meta-analyses of observational studies.	2
38	Smith, 2019 (288)	Altered Mental Status in the Emergency Department.	1

Nº	Autor año	Título	Motivo exclusión
39	Zuo, 2019 (289)	Free Energy Profile for Penetration of Pittsburgh Compound-B into the Amyloid β Fibril.	1
40	Veronese, 2020 (290)	Magnesium and health outcomes: an umbrella review of systematic reviews and meta-analyses of observational and intervention studies.	2
41	Vink, 2018 (291)	Music-Based Therapeutic Interventions for People with Dementia: A Mini-Review.	1
42	Veronese, 2019 (292)	Is chocolate consumption associated with health outcomes? An umbrella review of systematic reviews and meta-analyses.	2
43	Veronese, 2018 (293)	Dietary fiber and health outcomes: an umbrella review of systematic reviews and meta-analyses.	2
44	Ying, 2018 (294)	Iterating a framework for the prevention of caregiver depression in dementia: a multi-method approach.	1
45	Parker, 2018 (295)	What is Comprehensive Geriatric Assessment (CGA)? An umbrella review.	4
46	Finnigan-Fox, 2017 (296)	Hospice, She Yelped: Examining the Quantity and Quality of Decision Support Available to Patient and Families Considering Hospice.	1
47	Catalá-López, 2017 (297)	Anorexia nervosa and cancer: a protocol for a systematic review and meta-analysis of observational studies.	2, 3
48	Shenkin, 2017 (298)	Systematic reviews: guidance relevant for studies of older people.	1, 4
49	Bortolato, 2017 (299)	Systematic assessment of environmental risk factors for bipolar disorder: an umbrella review of systematic reviews and meta-analyses.	2
50	Apóstolo, 2017 (300)	Predicting risk and outcomes for frail older adults: an umbrella review of frailty screening tools.	4
51	Catalá-López, 2017 (301)	Cancer and central nervous system disorders: protocol for an umbrella review of systematic reviews and updated meta-analyses of observational studies.	3
52	Evenepoel, 2017 (302)	Bone biopsy practice patterns across Europe: the European renal osteodystrophy initiative-a position paper.	1
53	Bakker, 2016 (303)	Er is disciplinebreed sprake van relatieve onbevoegdheid en onbekwaamheid op het gebied van kwetsbare ouderen.	1, 4
54	Jadczak, 2016 (194)	Effectiveness of exercise interventions on physical function in community-dwelling frail older people: an umbrella review protocol.	3
55	Hategan, 2016 (304)	Aging with autism spectrum disorder: an emerging public health problem.	1
56	Hermans, 2016 (305)	Usefulness, feasibility and face validity of the interRAI Palliative Care instrument according to care professionals in nursing homes: A qualitative study.	1
57	Tamaki, 2016 (306)	Japan Prosthodontic Society position paper on "occlusal discomfort syndrome".	1, 2
58	Apóstolo, 2016 (307)	Predicting risk and outcomes for frail older adults: a protocol for an umbrella review of available frailty screening tools.	3

Nº	Autor año	Título	Motivo exclusión
59	Wan, 2015 (308)	Applications of Induced Pluripotent Stem Cells in Studying the Neurodegenerative Diseases.	1
60	Morris, 2014 (309)	Predicting risk of hospital and emergency department use for home care elderly persons through a secondary analysis of cross-national data.	1
61	Dybedal, 2013 (310)	Neuropsychological functioning in late-life depression.	1
62	Cherry, 2013 (311)	Social engagement and health in younger, older, and oldest-old adults in the Louisiana Healthy Aging Study.	1
63	Lee, 2013 (312)	Defining and managing overactive bladder: disagreement among the experts.	1
64	Sandoval, 2012 (313)	Equine wellness care in ambulatory practice.	1, 2
65	Porter, 2008 (314)	Reducing my risk of intrusion: an intention of old homebound women who live alone.	1
66	Bronstein, 2006 (315)	Developing an aging prepared community: collaboration among counties, consumers, professionals and organizations.	1, 2
67	Auger, 2006 (316)	Making sense of pragmatic criteria for the selection of geriatric rehabilitation measurement tools.	1
68	Sonnenblick, 2004 (317)	The geriatric department in a general hospital.	1
69	McHorney, 2003 (318)	Ten recommendations for advancing patient-centered outcomes measurement for older persons.	1
70	Yeaworth, 2000 (319)	Frontotemporal dementia: a different kind of dementia.	1
71	Robben, 1999 (320)	Beroepsvorming in tweevoud: de klinisch geriater en de verpleeghuisarts, 1945-1990.	1
72	Stessman, 1997 (321)	Home hospitalization in the spectrum of community geriatric care.	1
73	Morse, 1995 (322)	Achieving high quality long-term care for elderly people: consumers' wishes and providers' responsibilities.	1
74	Hall, 1989 (323)	Multidimensionality of health status in an elderly population. Construct validity of a measurement battery. -	1
75	Brousseau, 1975 (324)	"Umbrella" projects are questioning what is, what will be, what should be the status of the aged.	1

ANEXO 4

Resultados por proyectos de resultados de interés identificados por las dos metodologías

PROYECTO 1. ARTOSIS DE RODILLA Y EVENTOS DE SALUD ADVERSOS

El URT utilizado como modelo de referencia (165)(Veronese, 2023). En la metodología TU[®] se utilizó como código de identificación del proyecto y de los Rdl: KOA.

El URT identificó 5 Rdl; mientras que TU[®] identificó 12 Rdl (**Tabla 47**). De los 5 Rdl identificados por el URT, TU[®] identificó los 5 (100%), por lo que TU[®] dejó sin identificar de URT 0 Rdl (0%) (**Figura 29**). De los 12 Rdl identificados por TU[®], el URT identificó 5 (41.66%), mientras dejó de identificar 7 (58.33%) (**Figura 30**).

Tabla 47

Proyecto 1. Rdl identificados por URT y TU[®]: total, coincidentes y no coincidentes.

	URT	TU [®]
Rdl identificados total	5	12
Rdl de URT identificados por TU [®] (N) (% de URT)	N = 5	(100.00 %)
Rdl de URT NO identificados por TU [®] (N) (% de URT)	N = 0	(0.00 %)
Rdl de TU [®] identificados por URT (N) (% de TU [®])	N = 5	(41.66 %)
Rdl de TU [®] NO identificados por URT (N) (% de TU [®])	N = 7	(58.33 %)

Figura 29

Proyecto 1. Rdl identificados por URT (N=5) concordancia TU[®].

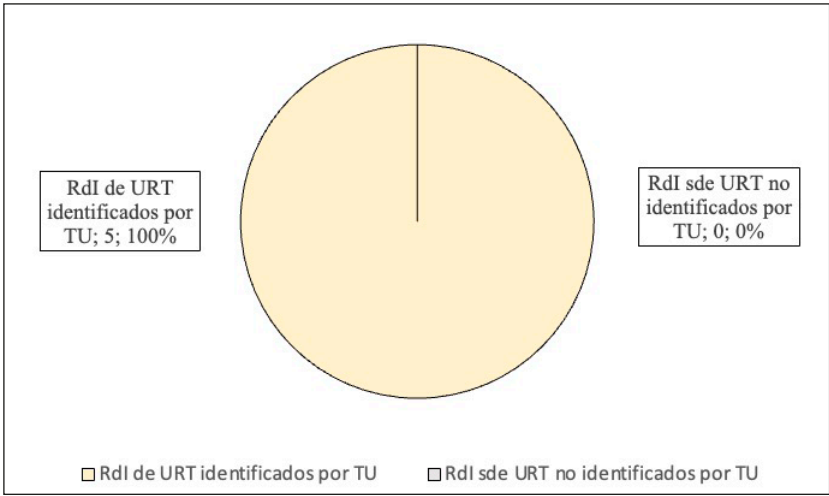
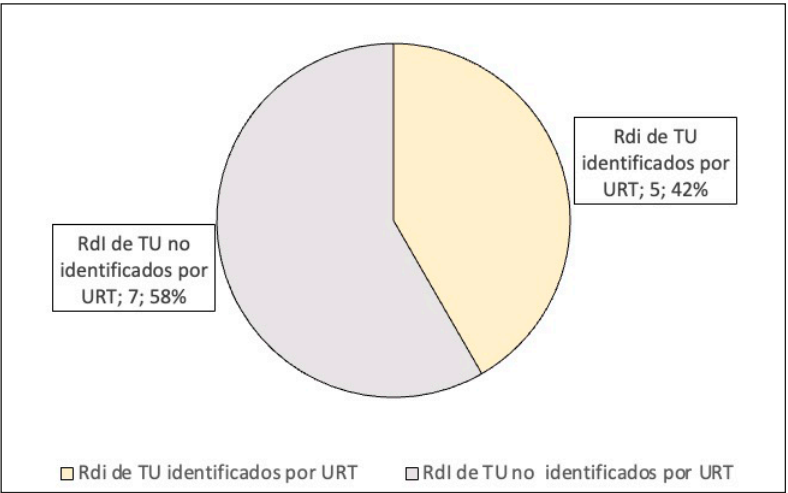


Figura 30

Proyecto 1. Rdl identificados por TU[®] (N=12) concordancia URT.



En la **Tabla 48** se desglosan de forma individual por término los Rdl identificados por URT y los Rdl coincidentes emparejados de TU[®]. En la **Tabla 49** se desglosan de forma individual y por término los Rdl identificados por TU[®] no identificados por URT.

Tabla 48

Proyecto 1. Rdl identificados por URT y emparejamiento por TU[®].

URT (Rdl identificado)	TU [®] (Rdl identificado)
1 All-cause Mortality	KOA. All-cause Mortality Risk
2 Cardiovascular Mortality	KOA. Cardiovascular Mortality Risk
3 Cardiovascular Disease Risk	KOA. Cardiovascular Disease Risk
4 Falls Risk	KOA. Fall Risk
5 Subclinical Atherosclerosis Risk	KOA. Subclinical Atherosclerosis Risk

Tabla 49

Proyecto 1. Rdl identificados por TU[®] no identificados por URT

TU [®] (Rdl identificado)
1 KOA. Metabolic Syndrome
2 KOA. Weight Gain TKR women
3 KOA. Hyperglycemia
4 KOA. Worse Balance Tests
5 KOA. Hypertension
6 KOA. Weight Gain TKR men
7 KOA. Dyslipidemia

PROYECTO 2. ÍNDICE DE DIETA INFLAMATORIA Y SALUD HUMANA

El URT utilizado como modelo de referencia es (176)(Marx, 2021). En la metodología TU[®] se utilizó como código de identificación del proyecto y de los resultados de interés: DII.

El URT encontró 38 Rdl; mientras que TU[®] identificó 71 Rdl (Tabla 50). De los 38 Rdl identificados por el URT, TU[®] identificó 30 (78.94%), mientras que TU[®] dejó de identificar de URT 8 Rdl (21.05%) (Figura 31). De los 71 Rdl identificados por TU[®], el URT identificó 30 (42.25%), mientras dejó de identificar 41 (57.74%) (Figura 32).

Tabla 50
Proyecto 2. Rdl identificados por URT y TU[®]: total, coincidentes y no coincidentes.

	URT	TU [®]
Rdl identificados totales	38	71
Rdl de URT identificados por TU [®] (N) (% de URT)	N = 30	(78.94%)
Rdl de URT NO identificados por TU [®] (N) (%)	N = 8	(21.05%)
Rdl de TU [®] identificados por URT (N) (% de TU [®])	N = 30	(42.25%)
Rdl de TU [®] NO identificados por URT (N) (%)	N = 41	(57.74%)

Figura 31
Proyecto 2. Rdl identificados por URT (N=38) concordancia TU[®].

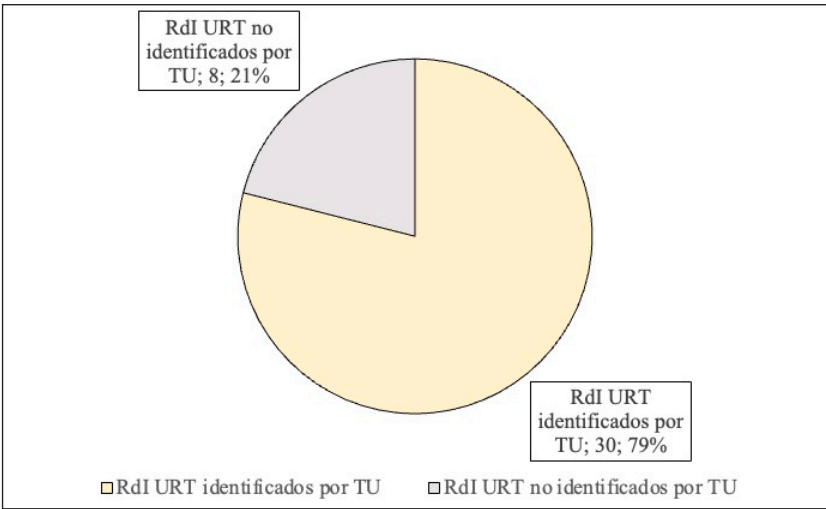
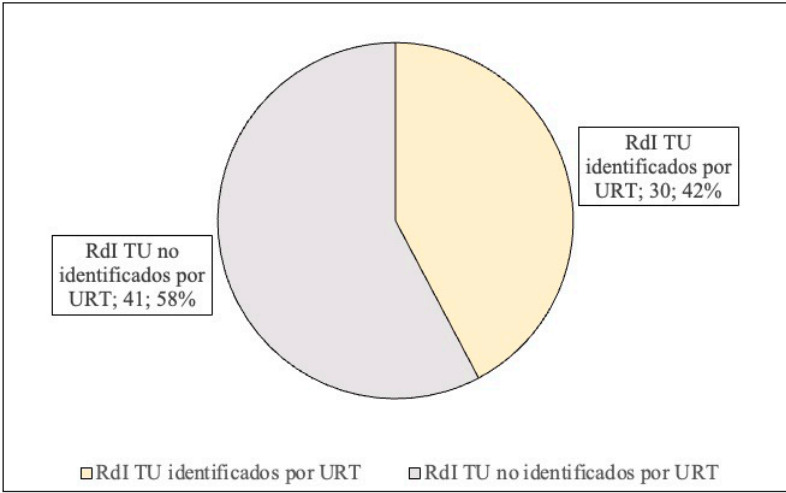


Figura 32
Proyecto 2. Rdl identificados por TU[®] (N=71) concordancia URT.



En la **Tabla 51** se desglosan de forma individual por término los Rdl identificados por URT y los Rdl coincidentes emparejados de TU[®]. En la **Tabla 52** se desglosan de forma individual y por término los Rdl identificados por TU[®] no identificados por URT.

Tabla 51
Proyecto 2. Rdl identificados por URT y emparejamiento por TU[®].

URT (Rdl identificado)	TU [®] (Rdl identificado)
1 Myocardial Infarction	-
2 Oral Cancer	DII. Oral Cancer. Hsc. Risk
3 Respiratory Cancer	DII. Respiratory Cancer. Hsc. Risk
4 Pancreatic Cancer	DII. Pancreatic Cancer. Hsc. Risk
5 Colorectal Cancer	DII. Colorectal Cancer. Hsc. Risk
6 Overall Cancer	DII. Overall Cancer. Hsc. Risk
7 All-cause Mortality	DII. All-cause Mortality. Hsc. Risk
8 Waist Circumference	DII. Waist Circumference. Hsc. Higher Value
9 HbA1C	DII. HbA1c. Hsc. Higher Value
10 Depression	DII. Depression. Hsc. Risk
11 Prostate Cancer	DII. Prostate Cancer. Hsc. Risk
12 Pharyngeal Cancer	DII. Pharyngeal Cancer. Hsc. Risk
13 Ovarian Cancer	DII. Ovarian Cancer. Hsc. Risk
14 Breast Cancer	DII. Breast Cancer. Hsc. Risk
15 Lung Cancer	DII. Lung Cancer
16 Esophageal Cancer	DII. Esophageal Cancer. Hsc. Risk

17 Systolic Blood Pressure	DII. Systolic BP. Hsc. Risk
18 HOMA-IR	DII. HOMA-IR. Hsc. Higher Value
19 Insulin	DII. Insulin. Hsc. Higher Value
20 Fasting Blood Glucose	DII. Fasting Blood Glucose. Hsc. Higher Value
21 Cardiovascular	DII. CVD. Hsc. Risk
22 Hypertension	DII. Hypertension. Hsc. Risk
23 Kidney Cancer	DII. Kidney Cancer. Hsc. Risk
24 Gastric Cancer	DII. Gastric Cancer. Hsc. Risk
25 CHD Mortality	-
26 CVD Mortality	DII. CVD Mortality. Hsc. Risk
27 Cancer Mortality	DII. Cancer Mortality. Hsc. Risk
28 Diastolic Blood Pressure	-
29 Waist to hip Ratio	-
30 Central Obesity	DII. Central Obesity
31 Hyperglycemia	DII. Hyperglycaemia
32 Metabolic Syndrome	DII. Metabolic Syndrome. Hsc. Risk
33 Angina	-
34 Stroke	-
35 IHD-CHD Risk	-
36 Urothelial Cancer	DII. Urothelial Cancer. Hsc. Risk
37 Endometrial Cancer	-
38 Laryngeal Cancer	DII. Laryngeal Cancer. Hsc. Risk

Tabla 52
Proyecto 2. Rdl identificados por TU® no identificados por URT

TU® (Rdl identificado)
1 DII. Anxiety Men
2 DII. Bladder Cancer
3 DII. Bone Mineral Density Femoral Neck
4 DII. Colorectal Adenoma Recurrence
5 DII. Upper Aerodigestive Cancer. Hsc. Risk
6 DII. Small-for-Gestational-Age Babies
7 DII. Schizophrenia

8 DII. Sarcopenia
9 DII. Rectal Cancer
10 DII. Pregnancy Childhood Depression and Anxiety
11 DII. Pregnancy Childhood Aggressive Behavior
12 DII. Pregnancy Childhood ADHD Symptoms
13 DII. Osteoporosis
14 DII. Low Birth Weight
15 DII. Hypertriacylglycerolaemia
16 DII. Hormone-dependent Cancer. Hsc. Risk
17 DII. Gynecological Cancer. Hsc. Risk
18 DII. Gestational Diabetes Mellitus
19 DII. Frailty Weakness
20 DII. Frailty
21 DII. Fracture
22 DII. Distress
23 DII. Digestive Cancer. Hsc. Risk
24 DII. Diabetes Type 2
25 DII. Diabetes Mellitus
26 DII. Depression Women
27 DII. Depression Men
28 DII. Colorectal Cancer Women
29 DII. Colorectal Cancer Men
30 DII. Colon Cancer
31 DII. Cardiometabolic Diseases Mortality
32 DII. Cardiometabolic Diseases
33 DII. Bone Mineral Density Total Hip
34 DII. Bone Mineral Density Lumbar Spine
35 DII. Bone Mineral Density
36 DII. Bladder Cancer
37 DII. Anxiety Women
38 DII. Anxiety
39 DII. ADHD Total Symptomatology
40 DII. ADHD Inattention Symptomatology
41 DII. ADHD Hyperactivity Symptomatology

PROYECTO 3. EFECTOS DE EJERCICIO EN ANCIANOS CON SARCOPENIA

El URT utilizado como modelo de referencia es (167)(Shen, 2022). En la metodología TU® se utilizó como código de identificación del proyecto y de los Rdl: EXS.

El URT identificó 7 Rdl; mientras que TU® identificó 34 Rdl (Tabla 53). De los 7 Rdl identificados por el URT, TU® identificó 6 (85.71%) (Figura 33), mientras que TU® dejó de identificar de URT 1 Rdl (14.28%). De los 34 Rdl identificados por TU®, el URT identificó 6 (17.64%), mientras dejó de identificar 28 (82.35%) (Figura 34).

Nota: el URT identificó dos Rdl relacionados con la velocidad de la marcha (Usual Walking Speed y Maximum Walking Speed), mientras que TU® identificó relacionado con el mismo concepto 1 Rdl (EXS. Gait Speed). Asumimos que “Usual Walking Speed” del URT es el mismo concepto que “EXE. Gait Speed” de TU®.

Tabla 53

Proyecto 3. Rdl identificados por URT y TU®: total, coincidentes y no coincidentes.

	URT	TU®
Rdl identificados totales	7	34
Rdl de URT identificados por TU® (N) (% de URT)	N = 6	(85.71%)
Rdl de URT NO identificados por TU® (N) (% de URT)	N = 1	(14.28%)
Rdl de TU® identificados por URT (N) (% de TU®)	N = 6	(17.64%)
Rdl de TU® NO identificados por URT (N) (% de TU®)	N = 28	(82.35%)

Figura 33

Proyecto 3. Rdl identificados por URT (N=7) concordancia TU®.

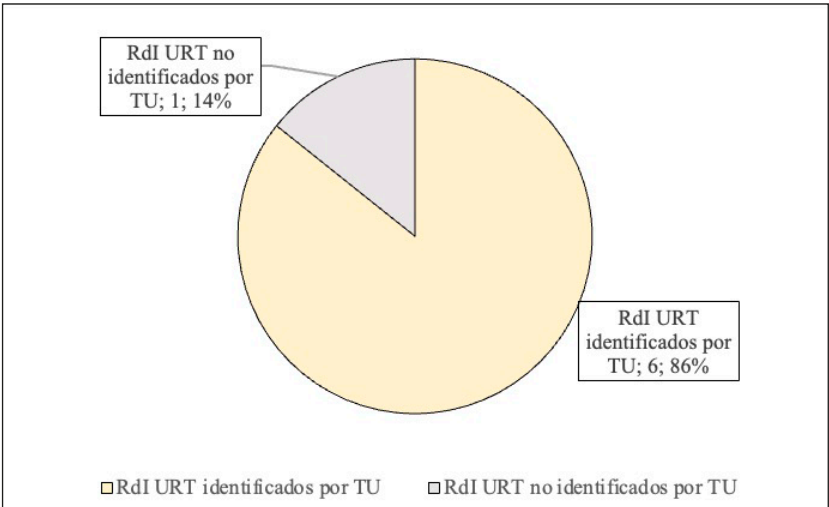
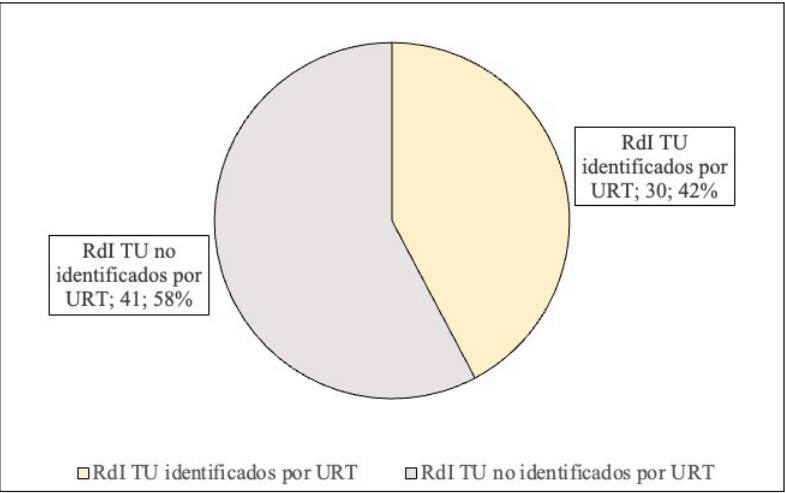


Figura 34

Proyecto 3. Rdl identificados por TU® (N=34) concordancia URT.



En la Tabla 54 se desglosan de forma individual por término los Rdl identificados por URT y los Rdl coincidentes emparejados de TU®. En la Tabla 55 se desglosan de forma individual y por término los Rdl identificados por TU® no identificados por URT.

Tabla 54

Proyecto 3. Rdl por URT y emparejamiento por TU®.

URT (Rdl identificado)	TU® (Rdl identificado)
1 Grip Strength	EXS. Hand Grip Strength
2 Knee Extension Strength	EXS. Knee-Extension Strength
3 Usual Walking Speed	EXS. Gait Speed
4 Maximum Walking Speed	
5 Timed Up and Go Test	EXS. Timed Up and Go Test
6 Timed Five Chair Stand	EXS. Timed Five Chair Stands
7 Dynamic Balance	EXS. Dynamic Balance

Tabla 55

Proyecto 3. Rdl identificados por TU® no identificados por URT.

TU® (Rdl identificado)
1 EXS. Body Fat Reduction
2 EXS. Waist Circumference
3 EXS. Trunk Fat Mass
4 EXS. Total Fat Mass
5 EXS. TNF-α

6 EXS. Skeletal Muscle Index
7 EXS. Sit to Stand Test
8 EXS. Short Physical Performance Battery
9 EXS. Rheumatoid Arthritis Muscle Mass
10 EXS. Reduce Falls
11 EXS. Physical Performance
12 EXS. Muscle Strength
13 EXS. Muscle Mass of Lower Extremities
14 EXS. Muscle Mass / Height2
15 EXS. Muscle Mass
16 EXS. Mini-Mental State Examination
17 EXS. Interleukin-6
18 EXS. Functional Mobility
19 EXS. Fear of Falling
20 EXS. ESKD Sit-to-Stand (STS) 60 score
21 EXS. ESKD Hand Grip Strength
22EXS. Diastolic Blood Pressure
23EXS. Depression
24 EXS. C Reactive Protein
25 EXS. Body Fat Reduction
26 EXS. Balance
27 EXS. Appendicular Skeletal Muscle Mass
28 EXS. 30-Second Chair Stand Test

PROYECTO 4. INTERVENCIONES PARA REDUCIR EFECTOS ADVERSOS EN ANCIANOS DESPUES DE ACUDIR A URGENCIAS

El URT utilizado como modelo de referencia es (161)(Conneely, 2022). En la metodología TU® se utilizó como código de identificación del proyecto y de los resultados de interés: EDI.

El URT identificó 10 Rdl; mientras que TU® identificó 22 Rdl (**Tabla 56**). De los 10 Rdl identificados por el URT, TU® identificó 7 (70.00%), mientras que TU® dejó de identificar de URT 3 Rdl (30.00%) (**Figura 35**). De los 22 Rdl identificados por TU®, el URT identificó 7 (31.81%), mientras dejó de identificar 15 (68.18%) (**Figura 36**).

Nota: El URT identificó 3 Rdl relacionados con la situación funcional (Functional Status/Decline 12 months, Functional Status/Decline 4 weeks y Functional Status/Decline 6 months) todas con una delimitación temporal. Mientras que TU® solo identifica un Rdl relacionado con la situación funcional (EDI. Functional Status) que no tiene una delimitación temporal, pero si detecta un Rdl relacionado con el mismo concepto.

Tabla 56
Proyecto 4. Rdl identificados por URT y TU®: total, coincidentes y no coincidentes.

	URT	TU®
Rdl identificados totales	10	22
Rdl de URT identificados por TU® (N) (% de URT)	N = 7	(70.00%)
Rdl de URT NO identificados por TU® (N) (% de URT)	N = 3	(30.00%)
Rdl de TU® identificados por URT (N) (% de TU®)	N = 7	(31.81%)
Rdl de TU® NO identificados por URT (N) (% de TU®)	N = 15	(68.18%)

Figura 35
Proyecto 4. Rdl identificados por URT (N=10) concordancia TU®.

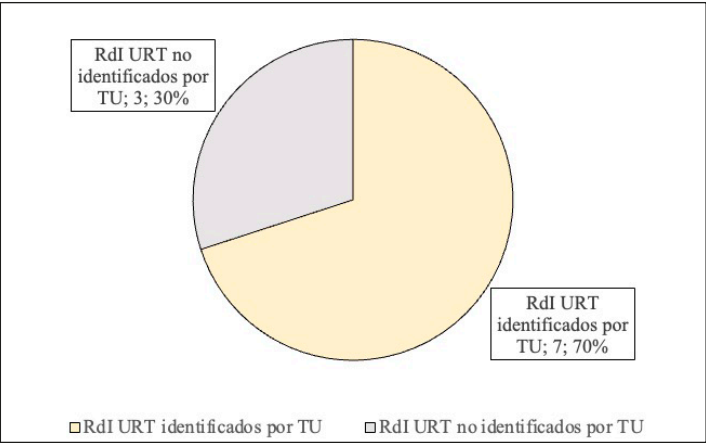
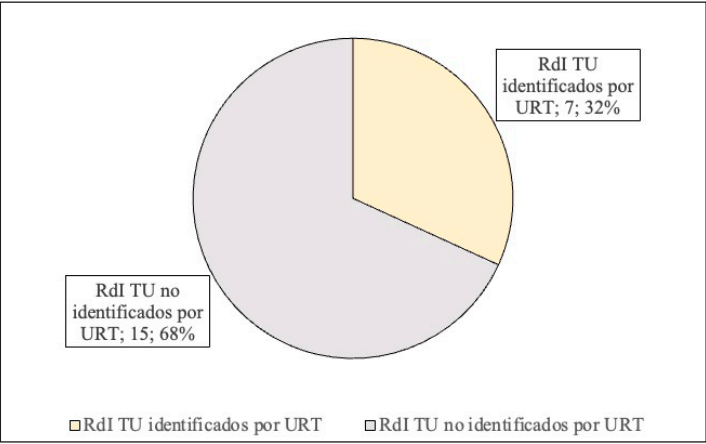


Figura 36
Proyecto 4. Rdl identificados por URT (N=10) concordancia TU®.



En la **Tabla 57** se desglosan de forma individual por término los Rdl identificados por URT y los Rdl coincidentes emparejados de TU®. En la **Tabla 58** se desglosan de forma individual y por término los Rdl identificados por TU® no identificados por URT.

Tabla 57
Proyecto 4. Rdl identificados por URT y emparejamiento por TU®.

URT (Rdl identificado)	TU® (Rdl identificado)
1 Functional Status/Decline 12 months	-
2 Functional Status/Decline 4 weeks	-
3 Functional Status/Decline 6 months	-
4 Quality of Life	EDI. QoL
5 Mortality	EDI. Mortality
6 Patient Experience or Satisfaction	EDI. Patient Experience
7 ED Revisits	EDI. ED Revisits
8 Hospital Admissions	EDI. Hospital Admission
9 Rate of Falls	EDI. Falls
10 Number of Fallers	EDI. Number of Fallers

Tabla 58
Proyecto 4. Rdl identificados por TU® no identificados por URT.

TU® (Rdl identificado)
1 EDI. Reduction in Falls
2 EDI. Participation in Daily Occupations 6 months
3 EDI. Participation in Daily Occupations 12 months
4 EDI. Nursing Home Admissions
5 EDI. Number of Fallers
6 EDI. Monthly Rate of Falling
7 EDI. Mental Status
8 EDI. Length of Hospital Stay
9 EDI. Intensive Interventions Improve Outcomes
10 EDI. Functional Status
11 EDI. Fall-Related ED Presentations
12 EDI. Fall Related Injuries
13 EDI. Delirium Prevalence
14 EDI. Cognitive Impairment Recognition
15 EDI. Rate of Fractured Neck of Femur

PROYECTO 5. EFECTO DE DIGOXINA SOBRE MORTALIDAD EN FIBRILACION AURICULAR

El URT utilizado como modelo de referencia es (169)(Gazzaniga, 2023). En la metodología TU® se utilizó como código de identificación del proyecto y de los resultados de interés: DGM.

El URT identificó 4 Rdl; mientras que TU® identificó 25 Rdl (**Tabla 59**). De los 4 Rdl identificados por el URT, TU® identificó los 4 (100.00%), por lo que TU® dejó sin identificar de URT 0 Rdl (0.00%) (**Figura 37**). De los 25 Rdl identificados por TU®, el URT identificó 4 (16.00%), mientras dejó de identificar 21 (84.00%) (**Figura 38**).

Tabla 59
Proyecto 5. Rdl identificados por URT y TU®: total, coincidentes y no coincidentes.

	URT	TU®
Rdl identificados totales	4	25
Rdl de URT identificados por TU® (N) (% de URT)	N = 4	(100.00%)
Rdl de URT NO identificados por TU® (N) (% de URT)	N = 0	(0.00%)
Rdl de TU® identificados por URT (N) (% de TU®)	N = 4	(16.00%)
Rdl de TU® NO identificados por URT (N) (% de TU®)	N = 21	(84.00%)

Figura 37
Proyecto 5. Rdl identificados por URT (N=4) concordancia TU®.

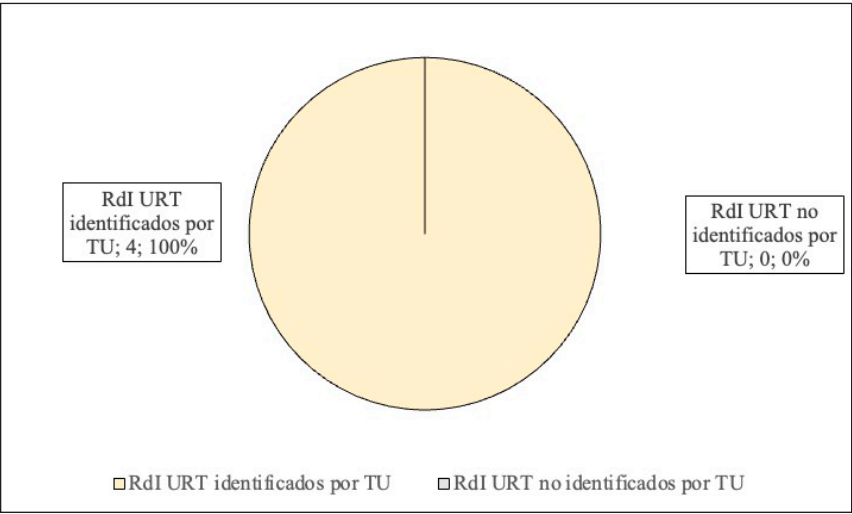
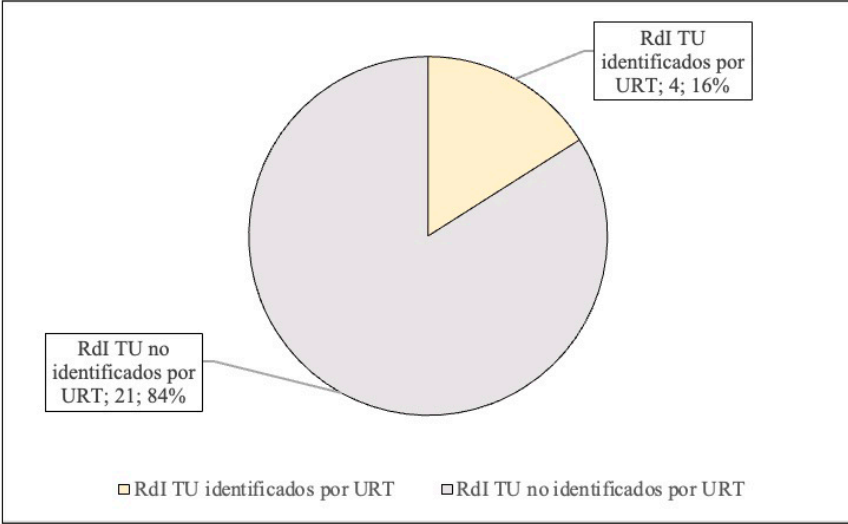


Figura 38
Proyecto 5. Rdl identificados por TU® (N=25) concordancia URT.



En la **Tabla 60** se desglosan de forma individual por término los Rdl identificados por URT y los Rdl coincidentes emparejados de TU®. En la **Tabla 61** se desglosan de forma individual y por término los Rdl identificados por TU® no identificados por URT.

Tabla 60
Proyecto 5. Rdl identificado por URT y emparejamiento por TU®.

URT (Rdl identificado)	TU® (Rdl identificado)
1 All-cause Mortality in whole population	DGM. All-cause Mortality risk
2 Cardiovascular Mortality	DGM. Cardiovascular Mortality
3 All-cause Mortality patients AF only	DGM. All-cause Mortality Atrial Fibrillation
4 All-cause Mortality patients AF + HF	DGM. All-cause Mortality Atrial Fibrillation with HF Risk

Tabla 61
Proyecto 5. Rdl identificado por TU® no identificado por URT.

TU® (Rdl identificado)
1 DGM. Acute Conversion SR
2 DGM. Cancer-Specific Mortality
3 DGM. Heart Failure Risk
4 DGM. Mortality HF Sinus Rhythm
5 DGM. Sudden Cardiac Death Atrial Fibrillation Risk
6 DGM. Stroke Risk
7 DGM. Serious Adverse Event

8 DGM. Prostate Cancer Mortality
9 DGM. Prostate Cancer
10 DGM. Hospitalization HF Sinus Rhythm
11 DGM. Deterioration in Clinical Status HF Sinus Rhythm
12 DGM. Cardiovascular Mortality Atrial Fibrillation Risk
13 DGM. Breast Cancer
14 DGM. All-cause Mortality without HF
15 DGM. All-cause Mortality without BBlockers
16 DGM. All-cause Mortality with BBlockers
17 DGM. All-cause Mortality New Users
18 DGM. All-cause Mortality HF Risk
19 DGM. All-cause Mortality Atrial Fibrillation without HF Risk
20 DGM. All-cause Mortality Atrial Fibrillation plus B-blockers
21 DGM. All-cause Hospitalization Atrial Fibrillation Risk

PROYECTO 6. EFECTOS DE SUPLEMENTOS DE VITAMINA D SOBRE BIOMARCADORES DE DIABETES MELLITUS TIPO 2

El URT utilizado como modelo de referencia es (170)(Musazadeh, 2023). En la metodología TU® se utilizó como código de identificación del proyecto y de los resultados de interés: VDD.

El URT identificó 8 Rdl; mientras que TU® identificó 149 Rdl (**Tabla 62**). De los 8 Rdl identificados por el URT, TU® identificó los 8 (100.00%), por lo que TU® dejó sin identificar de URT 0 Rdl (0.00%) (**Figura 39**). De los 149 Rdl identificados por TU®, el URT identificó 8 (5.36%), mientras dejó de identificar 141 (94.63%) (**Figura 40**).

Nota: en esta comparación la diferencia entre el número de Rdl encontrados por las dos metodologías viene determinado por los diferentes criterios de inclusión de Rdl utilizados por ambas metodologías. Mientras que el URT limitó los Rdl a marcadores bioquímicos de control (FBG, HbA1c, HOMA-IR), el TU® incluyó cualquier marcador bioquímico y a la vez detalló los mismos valores de los Rdl de los URT, pero con más detalles.

Tabla 62
Proyecto 6. Rdl identificados por URT y TU®: total, coincidentes y no coincidentes.

	URT	TU®
Rdl identificados totales	8	149
Rdl de URT identificados por TU® (N) (% de URT)	N = 8	(100.00%)
Rdl de URT NO identificados por TU® (N) (% de URT)	N = 0	(0.00%)
Rdl de TU® identificados por URT (N) (% de TU®)	N = 8	(5.36%)
Rdl de TU® NO identificados por URT (N) (% de TU®)	N = 41	(94.63%)

Figura 39
Proyecto 6. Rdl identificados por URT (N=8) concordancia TU®.

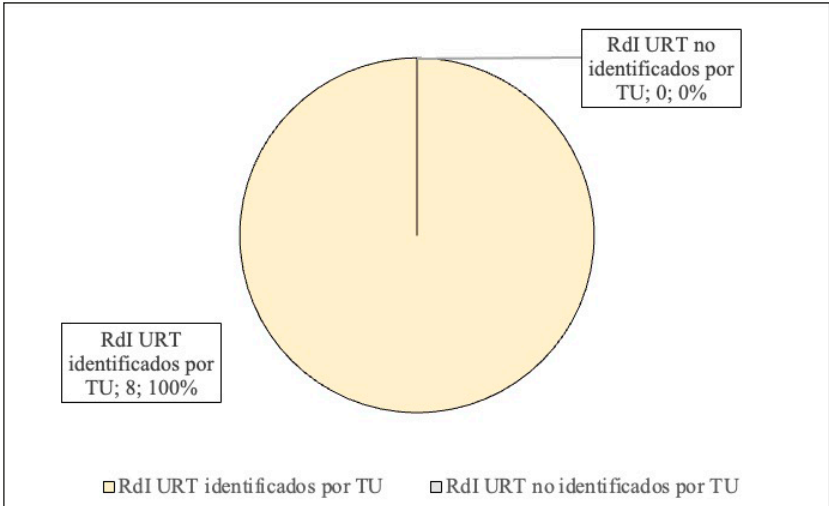
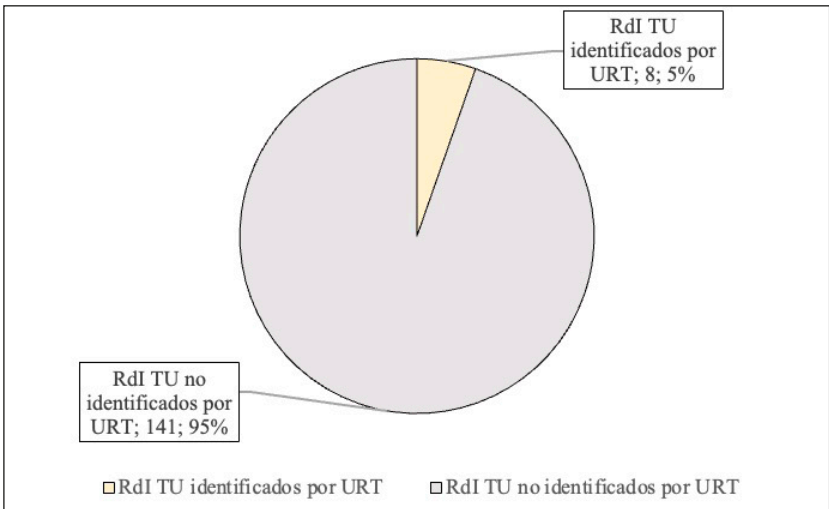


Figura 40
Proyecto 6. Rdl identificados por TU® (N=149) concordancia URT.



En la **Tabla 63** se desglosan de forma individual por término los Rdl identificados por URT y los Rdl coincidentes emparejados de TU®. En la **Tabla 64** se desglosan de forma individual y por término los Rdl identificados por TU® no identificados por URT.

Tabla 63
Proyecto 6. Rdl identificado por URT y emparejamiento por TU®.

URT (Rdl identificado)	TU® (Rdl identificado)
1 Vit D supplementation on FBG levels WMD	VDD. Vit D supplement FBG
2 SMD Vit D supplementation on FBG levels	VDD. Vit D supplement FBG
4 3 Vit D supplementation on HbA1c levels WMD	VDD. Vit D supplement HbA1c
5 SMD Vit D supplementation on HbA1c levels	VDD. Vit D supplement HbA1c
6 Vit D supplementation on HOMA-IR levels WMD	VDD. Vit D supplement Homeostatic Model Assessment for Insulin Resistance (HOMA-IR)
7 SMD Vit D supplementation on HOMA-IR levels	VDD. Vit D supplement Homeostatic Model Assessment for Insulin Resistance (HOMA-IR)
8 Vit D supplementation on Insulin levels WMD	VDD. Vit D supplement fasting insulin (FINS)
9 SMD Vit D supplementation on Insulin levels	VDD. Vit D supplement fasting insulin (FINS)

Tabla 64
Proyecto 6. Rdl identificado por TU® no identificado por URT.

TU® (Rdl identificado)
1 VDD. Vit D supplement Waist to hip Ratio
2 VDD. Vit D supplement Waist Circumstance (WC) (reduction)
3 VDD. Vit D supplement UAER
4 VDD. Vit D supplement Triglycerides
5 VDD. Vit D supplement TNF-alfa
6 VDD. Vit D supplement Testosterone Total
7 VDD. Vit D supplement SHBG (sex hormone-binding globulin)
8 VDD. Vit D supplement Serum Creatinine
9 VDD. Vit D supplement Serum Alanine Transaminase (ALT) (reduction)
10 VDD. Vit D supplement PTH
11 VDD. Vit D supplement Postprandial Blood Glucose (PBG)
12 VDD. Vit D supplement PCOS VLDL-C
13 VDD. Vit D supplement PCOS Triglyceride

14 VDD. Vit D supplement PCOS total cholesterol
15 VDD. Vit D supplement PCOS Testosterone total
16 VDD. Vit D supplement PCOS Testosterone free
17 VDD. Vit D supplement PCOS TAC (total antioxidant capacity)
18 VDD. Vit D supplement PCOS systolic blood pressure
19 VDD. Vit D supplement PCOS SHBG
20 VDD. Vit D supplement PCOS QUICKI
21 VDD. Vit D supplement PCOS PTH
22 VDD. Vit D supplement PCOS NO (nitric oxide)
23 VDD. Vit D supplement PCOS mF-G Score
24 VDD. Vit D supplement PCOS MDA (malondialdehyde)
25 VDD. Vit D supplement PCOS LH
26 VDD. Vit D supplement PCOS LDL-C
27 VDD. Vit D supplement PCOS HS-CRP
28 VDD. Vit D supplement PCOS HOMA-IR
29 VDD. Vit D supplement PCOS HOMA-B
30 VDD. Vit D supplement PCOS HDL-C
31 VDD. Vit D supplement PCOS GSH (total glutathione levels)
32 VDD. Vit D supplement PCOS Follicular Development
33 VDD. Vit D supplement PCOS FBG
34 VDD. Vit D supplement PCOS fasting insulin concentration
35 VDD. Vit D supplement PCOS Endometrial Thickness
36 VDD. Vit D supplement PCOS diastolic blood pressure
37 VDD. Vit D supplement PCOS DHEAS
38 VDD. Vit D supplement PCOS BMI
39 VDD. Vit D supplement PCOS (baseline Vit D deficiency) Trygliceride (red.)
40 VDD. Vit D supplement PCOS triglyceride
41 VDD. Vit D supplement Leptin
42 VDD. Vit D supplement LDL-Cholesterol
43 VDD. Vit D supplement Insulin Sensitivity Check Index
44 VDD. Vit D supplement insulin resistance by homeostasis model assessment (HOMA-IR) > 1 year study duration

45 VDD. Vit D supplement Insulin
46 VDD. Vit D supplement IL-6
47 VDD. Vit D supplement HS- CRP
48 VDD. Vit D supplement Homeostatic Model Assessment for Insulin Resistance (HOMA-IR)
49 VDD. Vit D supplement HOMA-IR baseline 25(OH)D $\geq$ 50 nmol/L
50 Vdd. Vit D supplement HOMA-B
51 VDD. Vit D supplement high-density lipoprotein-cholesterol (HDL-C) (increase)
52 VDD. Vit D supplement HbA1c (vit D deficiency)
53 VDD. Vit D supplement HbA1c (vit D > 50 mol/l))
54 VDD. Vit D supplement HbA1c (vit D < 50 mol/l))
55 VDD. Vit D supplement HbA1c (BMI<30kgm-2)
56 VDD. Vit D supplement HbA1c
57 VDD. Vit D supplement FBG (vit D deficiency)
58 VDD. Vit D supplement Fat Mass
59 VDD. Vit D supplement ESR (erythrocyte sedimentation rate)
60 VDD. Vit D supplement eGFR
61 VDD. Vit D supplement Cholesterol Total
62 VDD. Vit D supplement Calcium
63 VDD. Vit D supplement Body Weight (reduction)
64 VDD. Vit D supplement Body Weight
65 VDD. Vit D supplement Body Mass Index (reduction)
66 VDD. Vit D supplement Blood Pressure (systolic)
67 VDD. Vit D supplement Blood Pressure (diastolic)
68 VDD. Vit D supplement Blood Pressure
69 VDD. Vit D supplement Asia 2 h oral glucose tolerance test plasma glucose (2h-PG)
70 Vdd. Vit D supplement 2h-PG baseline 25(OH)D > 50 nmol/L
71 Vdd. Vit D supplement 2h-PG baseline 25(OH)D < 50 nmol/L
72 Vdd. Vit D supplement 2h-PG
73 VDD. Vit D supplement 24-hour Urine Protein
74 VDD. Vit D supp. Preeclampsia

75 VDD. Vit D supp. Preclampsia
76 VDD. Vit D supp. NAFLD Triglyceride
77 VDD. Vit D supp. NAFLD LDL-C
78 VDD. Vit D supp. NAFLD Insulin concentration
79 VDD. Vit D supp. NAFLD HOMA-IR
80 VDD. Vit D supp. NAFLD HDL-C
81 VDD. Vit D supp. NAFLD FBG
82 VDD. Vit D supp. NAFLD Cholesterol Total
83 VDD. Vit D supp. NAFLD BMI
84 VDD. Vit D supp. NAFLD AST
85 VDD. Vit D supp. NAFLD ALT
86 VDD. Vit D supp. Malondialdehyde (MDA)
87 VDD. Vit D supp. Gestational Diabetes Mellitus (GDM) Premature Delivery
88 VDD. Vit D supp. Gestational Diabetes Mellitus (GDM) Postpartum Hemorrhage (red.)
89 VDD. Vit D supp. Gestational Diabetes Mellitus (GDM) Polyhydramnios
90 VDD. Vit D supp. Gestational Diabetes Mellitus (GDM) Neonatal Hyerbilirubinemia (red.)
91 VDD. Vit D supp. Gestational Diabetes Mellitus (GDM) Maternal Hospitalization (red.)
92 VDD. Vit D supp. Gestational Diabetes Mellitus (GDM) Giant Children
93 VDD. Vit D supp. Gestational Diabetes Mellitus (GDM) Fetal Distress
94 VDD. Vit D supp. ESRD PTH
95 VDD. Vit D supp. ESRD FBG
96 VDD. Vit D supp. ESRD Fasting Insulin Levels (FINS)
97 VDD. Vit D supp. Diabetic Nephropathy Urine Albumina/Creatinine Ratio
98 VDD. Vit D supp. Diabetic Nephropathy Urine Albumin Excretion Rate (UAER)
99 VDD. Vit D supp. Diabetic Nephropathy HbA1c
100 VDD. Vit D supp. Diabetic Nephropathy Blood Pressure
101 VDD. Vit D supp. Diabetic Nephropathy 24-hour Proteinuria
102 VDD. Vit D supp. CVD Insulin-sensitivity Check Index

103 VDD. Vit D supp. CVD Insulin concentration
104 VDD. Vit D supp. CVD HS-CRP
105 VDD. Vit D supp. CVD HOMA-IR
106 VDD. Vit D supp. CVD HDL-cholesterol
107 VDD. Vit D supp. CVD FBG
108 VDD. Vit D supp. (CKD) Chronic Kidney Disease PTH
109 VDD. Vit D supp. (CKD) Chronic Kidney Disease Calcium (increase)
110 VDD. Vit D supp. (CKD) Chronic Kidney Disease Alkaline Phosphatase (red.)
111 VDD. Vit D supp Total Serum Antioxidant Capacity (TAC)
112 VDD. Vit D supp Total Glutathione (GSH) levels
113 VDD. Vit D supp Risk of Incident Diabetes
114 VDD. Vit D supp Risk of Incident Diabetes
115 VDD. Vit D supp Obesity Triglycerides (red.)
116 VDD. Vit D supp Obesity Systolic Blood Pressure (increase)
117 VDD. Vit D supp Obesity LDL-c (increase)
118 VDD. Vit D supp Obesity HOMA-IR
119 VDD. Vit D supp Obesity FBG
120 VDD. Vit D supp Nitric Oxide (NO)
121 VDD. Vit D supp CHF TNF-alfa
122 VDD. Vit D supp CHF PTH
123 VDD. Vit D supp CHF CRP
124 VDD. Vit D sup. Gestational Diabetes Mellitus (GDM) Triglyceride
125 VDD. Vit D sup. Gestational Diabetes Mellitus (GDM) Stillbirth
126 VDD. Vit D sup. Gestational Diabetes Mellitus (GDM) QUICKI
127 VDD. Vit D sup. Gestational Diabetes Mellitus (GDM) Neonatal Hospitalization
128 VDD. Vit D sup. Gestational Diabetes Mellitus (GDM) Neonatal Death
129 VDD. Vit D sup. Gestational Diabetes Mellitus (GDM) LDL-c
130 VDD. Vit D sup. Gestational Diabetes Mellitus (GDM) Insulin Concentration
131 VDD. Vit D sup. Gestational Diabetes Mellitus (GDM) Infant Length

132 VDD. Vit D sup. Gestational Diabetes Mellitus (GDM) Infant Birth Weight
133 VDD. Vit D sup. Gestational Diabetes Mellitus (GDM) HS-CRP
134 VDD. Vit D sup. Gestational Diabetes Mellitus (GDM) HOMA-IR
135 VDD. Vit D sup. Gestational Diabetes Mellitus (GDM) HOMA-B
136 VDD. Vit D sup. Gestational Diabetes Mellitus (GDM) HDL-c
137 VDD. Vit D sup. Gestational Diabetes Mellitus (GDM) Hb1Ac
138 VDD. Vit D sup. Gestational Diabetes Mellitus (GDM) FINS (fasting insulin)
139 VDD. Vit D sup. Gestational Diabetes Mellitus (GDM) FBG
140 VDD. Vit D sup. Gestational Diabetes Mellitus (GDM) Cholesterol total
141 VDD. Vit D sup. Gestational Diabetes Mellitus (GDM) Cesarean Section (red.)

PROYECTO 7. INTERVENCIONES PARA REDUCIR SOLEDAD

El URT utilizado como modelo de referencia es (171)(Veronese, 2021). En la metodología TU® se utilizó como código de identificación del proyecto y de los resultados de interés: LRI.

El URT identificó 6 Rdl; mientras que TU® identificó 49 Rdl (**Tabla 65**). De los 6 Rdl identificados por el URT, TU® identificó los 6 (100.00%), por lo que TU® dejó sin identificar de URT 0 Rdl (0.00%) (**Figura 41**). De los 49 Rdl identificados por TU®, el URT identificó 6 (12.24%), mientras dejó de identificar 43 (87.75%) (**Figura 42**).

Tabla 65

Proyecto 7. Rdl identificados por URT y TU®: total, coincidentes y no coincidentes.

	URT	TU®
Rdl identificados totales	6	49
Rdl de URT identificados por TU® (N) (% de URT)	N = 6	(100.00%)
Rdl de URT NO identificados por TU® (N) (% de URT)	N = 0	(0.00%)
Rdl de TU® identificados por URT (N) (% de TU®)	N = 6	(12.24%)
Rdl de TU® NO identificados por URT (N) (% de TU®)	N = 43	(87.75%)

Figura 41

Proyecto 7. Rdl identificados por URT (N=6) concordancia TU®.

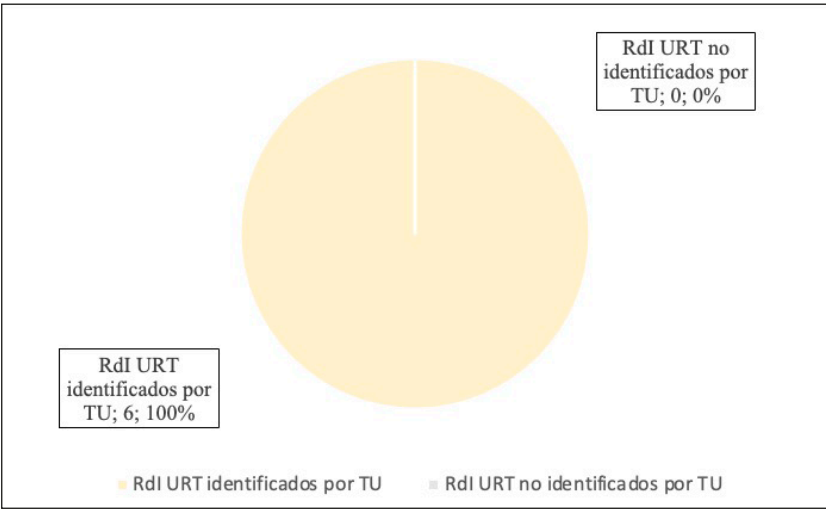
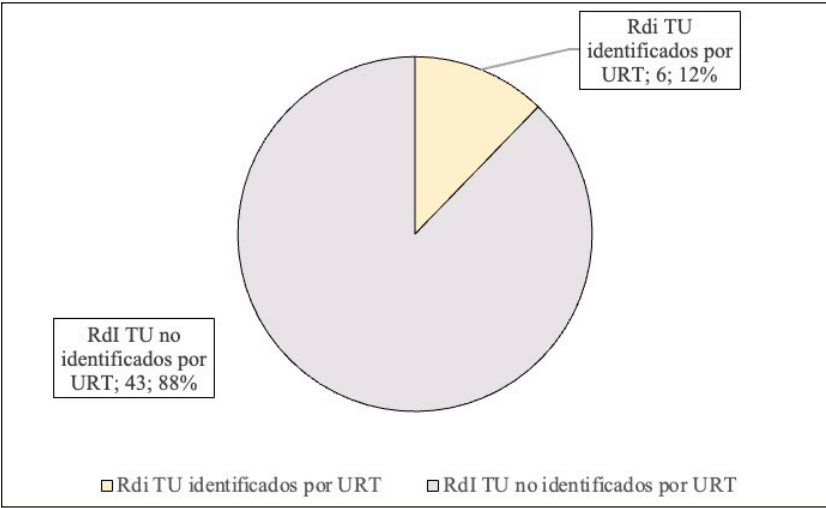


Figura 42

Proyecto 7. Rdl identificados por TU® (N=49) concordancia URT.



En la **Tabla 66** se desglosan de forma individual por término los Rdl identificados por URT y los Rdl coincidentes emparejados de TU®. En la **Tabla 67** se desglosan de forma individual y por término los Rdl identificados por TU® no identificados por URT.

Tabla 66
Proyecto 7. Rdl identificados por URT y emparejamiento por TU®.

URT (Rdl identificado)	TU® (Rdl identificado)
1 Befriending Interventions	LRI. Befriending Interventions
2 Technological interventions	LRI. Technological Interventions
3 Meditation / Mindfulness	LRI. Mindfulness-based intervention (MBI)
4 Animal therapy / robopets	LRI. Animal-Assisted Interventions
5 Social cognitive training	LRI. Social Cognition Treatment
6 Social support	LRI. Social Support

Tabla 67
Proyecto 7. Rdl identificados por TU® no identificados por URT.

TU® (Rdl identificado)
1 LRI. Internet training interventions Depression
2 LRI. Laughter therapy
3 LRI. Befriending Interventions Depression
4 LRI. Befriending Interventions QoL
5 LRI. Befriending Interventions Self-esteem
6 LRI. Befriending Interventions Well-Being
7 LRI. COVID-19 pandemic
8 LRI. Digital Technology Interventions (DTIs)
9 LRI. Robopets Depression
10 LRI. Robopets Loneliness
11 LRI. Robopets QoL
12 LRI. Smartphone-Based Video Calls
13 LRI. Remotely Delivered Intervention Video Calls 3 months
14 LRI. Remotely Delivered Intervention Video Calls 6 months

15 LRI. Video Game Participation
16 LRI. Trained Volunteers Services
17 LRI. Telehealth Interventions
18 LRI. Technological Interventions
19 LRI. Tai Chi
20 LRI. Spiritual Engagement (SE)
21 LRI. Social Networks Maintenance and Enhancement
22 LRI. Social Networking Sites (SNSs)
23 LRI. Social Connections
24 LRI. Robopets Agitation
25 LRI. Remotely Delivered Intervention Video Call
26 LRI. Remotely delivered intervention Individual format
27 LRI. Remotely delivered intervention 3-6 months
28 LRI. Remotely delivered intervention 3 months
29 LRI. Remotely Delivered Intervention
30 LRI. Reminiscence therapy
31 LRI. Psychological Interventions
32 LRI. Mindfulness-based stress reduction (MBSR) Depression
33 LRI. Mindfulness-based stress reduction (MBSR) Anxiety
34 LRI. Mindfulness-based stress reduction (MBSR)
35 LRI. Laughter Yoga
36 LRI. Internet training interventions
37 LRI. Individual-led Peer Support
38 LRI. Home-Based Interventions
39 LRI. Green Space Exposition
40 LRI. Dog-Assisted Interventions (DAIs)
41 LRI. Cognitively Based Compassion Training (CBCT)
42 LRI. Arts and Creativity Interventions
43 LRI. Horticultural therapy

PROYECTO 8. EFICACIA DE TRATAMIENTOS CONSERVADORES PARA ARTROSIS DE MANOS

El URT utilizado como modelo de referencia es (172)(Veronese, 2021). En la metodología TU® se utilizó como código de identificación del proyecto y de los resultados de interés: HOA.

El URT identificó 8 Rdl; mientras que TU® identificó 48 Rdl (Tabla 68). De los 8 Rdl identificados por el URT, TU® identificó los 6 (100.00%), por lo que TU® dejó sin identificar de URT 0 Rdl (0.00%) (Figura 43). De los 48 Rdl identificados por TU®, el URT identificó 8 (16.66%), mientras dejó de identificar 40 (83.33%) (Figura 44).

Tabla 68

Proyecto 8. Rdl identificados por URT y TU®: total, coincidentes y no coincidentes.

	URT	TU®
Rdl identificados totales	8	48
Rdl de URT identificados por TU® (N) (% de URT)	N = 8	(100.00%)
Rdl de URT NO identificados por TU® (N) (% de URT)	N = 0	(0.00%)
Rdl de TU® identificados por URT (N) (% de TU®)	N = 8	(16.66%)
Rdl de TU® NO identificados por URT (N) (% de TU®)	N = 40	(83.33%)

Figura 43

Proyecto 8. Rdl identificados por URT (N=8) concordancia TU®.

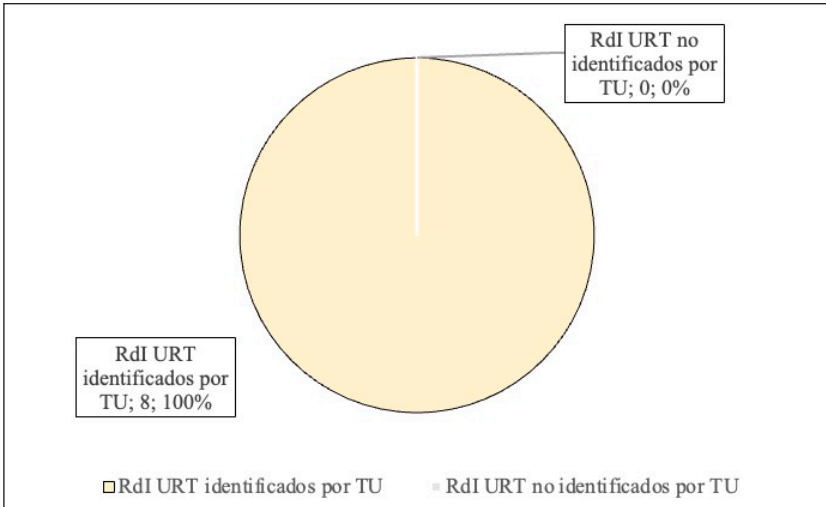
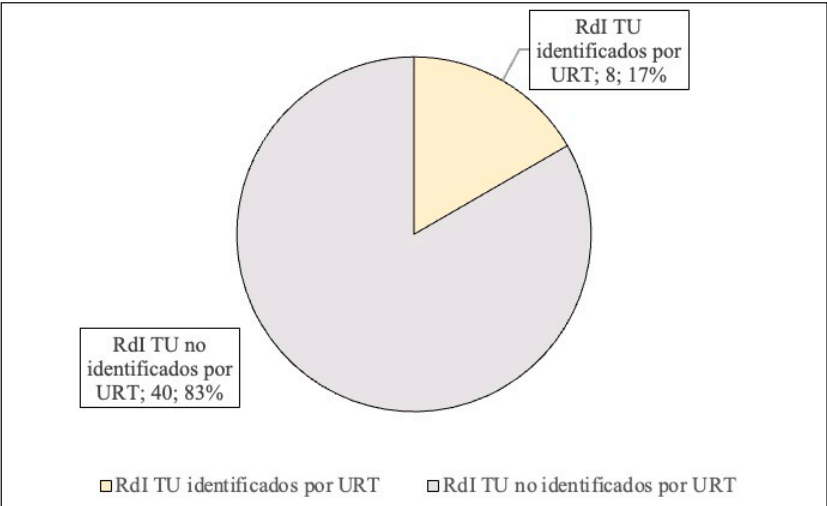


Figura 44

Proyecto 8. Rdl identificados pot TU® (N=48) concordancia URT.



En la **Tabla 69** se desglosan de forma individual por término los Rdl identificados por URT y los Rdl coincidentes emparejados de TU®. En la **Tabla 70** se desglosan de forma individual y por término los Rdl identificados por TU® no identificados por URT.

Tabla 69

Proyecto 8. Rdl identificados por URT y emparejamiento por TU®.

URT (Rdl identificado)	TU® (Rdl identificado)
1 Pain (follow-up 3-12 m) in thumb carpometacarpal OA / splint vs no splint	HOA. Thumb Pain (splint vs no splint)
2 Function (follow-up 3-12 m) in thumb carpometacarpal OA / splints vs no splint	HOA. Thumb Function (splint vs no splint)
3 Pain (long follow-up) in symptomatic hand OA / splints vs no splint	HOA. Hand Pain Long term (> 3months) (splint vs no splint)
4 Hand pain / Resistant Training vs no Resistant Training	HOA. Hand Pain (resistance training vs no resistance training)
5 Finger joint stiffness / Physical exercise vs no exercise	HOA. Finger Joint Stiffness (exercise vs no exercise)
6 Pain in hand OA / Physical exercise vs no exercise	HOA. Pain (exercise vs no exercise)
7 Pain in trapeziometacarpal OA / Physical and occupational therapy vs no intervention	HOA. Hand Pain (Physical and occupational therapy vs no intervention)
8 Function in thumb OA / IAHA (intra-articular hyaluronic acid) vs placebo	HOA. Thumb Function Intra-articular Hyaluronic Acid

Tabla 70
Proyecto 8. Rdl identificados por TU® no identificados por URT.

TU® (Rdl identificado)
1 HOA. Adverse Events (exercise vs no exercise)
2 HOA. Adverse Events Withdrawal (exercise vs no exercise)
3 HOA. Function (exercise vs no exercise)
4 HOA. Grip Strength (resistance training vs no resistance training)
5 HOA. Hand Function (Physical and occupational therapy vs no intervention)
6 HOA. Hand Function (splint vs no splint)
7 HOA. Hand Function (splint vs no splint) Medium Term
8 HOA. Hand Function Arnica Gel
9 HOA. Hand Function platelet-rich plasma (PRP)
10 HOA. Hand Function Pulsed Electromagnetic Field (PEMF) Therapy
11 HOA. Hand Grip Strength Paraffin bath therapy
12 HOA. Hand Pain (carpometacarpal) intra-articular (IA) Corticosteroids
13 HOA. Hand Pain (carpometacarpal) intra-articular (IA) Corticosteroids Medium-Term
14 HOA. Hand Pain (carpometacarpal) intra-articular (IA) Corticosteroids Short-Term
15 HOA. Hand Pain (carpometacarpal) intra-articular (IA) Hyaluronic Acid
16 HOA. Hand Pain (interphalageal) intra-articular (IA) Corticosteroids
17 HOA. Hand Pain (splint vs no splint)
18 HOA. Hand Pain Arnica Gel
19 HOA. Hand Pain Joint protection (JP)
20 HOA. Hand Pain (splint vs no splint) Medium Term
21 HOA. Hand Pain Paraffin bath therapy
22 HOA. Hand Pain platelet-rich plasma (PRP)
23 HOA. Hand Pain Pulsed Electromagnetic Field (PEMF) Therapy
24 HOA. Hand Pain Short Term (<3 months) (splint vs no splint)

25 HOA. QoL (exercise vs no exercise)
26 HOA. Thumb Function Intra-articular Corticosteroids
27 HOA. Thumb Function Intra-articular Corticosteroids (24 weeks)
28 HOA. Thumb Function Intra-articular Saline vs Corticosteroids
29 HOA. Thumb Function Mobilisation
30 HOA. Thumb Grip Strength Manual Therapy (MT)
31 HOA. Thumb Pain Intra-articular Corticosteroids
32 HOA. Thumb Pain Intra-articular Corticosteroids Medium-term
33 HOA. Thumb Pain Intra-articular Hyaluronic Acid
34 HOA. Thumb Pain Intra-articular Platelet-Rich Plasma
35 HOA. Thumb Pain Intra-articular Platelet-Rich Plasma Medium-term
36 HOA. Thumb Pain Intra-articular Saline vs Corticosteroids
37 HOA. Thumb Pain Manual Therapy (MT)
38 HOA. Thumb Pain Mobilisation
39 HOA. Thumb Pinch Strength Mobilisation
40 HOA. Thumb Pulp Pinch Force Intra-articular Hyaluronica Acid (24 weeks)

ANEXO 5

Análisis de Rdl con te discordantes de mas de un nivel en el Nivel O (N=15)

En este documento vamos a analizar de forma detallada los tamaños de efecto con *d* de Cohen que presentan discordancias entre URT y TU® de más de 1 nivel. Según se puede ver en la **Tabla 10** que sigue:

URT - TE	TU - TE				
	1.Trivial	2.Pequeño	3.Moderado	4.Grande	
1. Trivial	55.6% (10)	13.3% (4)	-	-	21.1% (14)
2.Pequeño	38.9% (7)	23.3% (7)	6.3% (1)	-	22.7% (15)
3.Moderado	-	16.7% (5)	6.3% (1)	50.0% (1)	10.6% (7)
4.Grande	5.6% (1)	46.7% (14)	87.5% (14)	50.0% (1)	45.5% (30)
	27.3% (18)	45.5% (30)	24.2% (16)	3.0% (2)	(66)
$\chi^2 = 37.85^{**}$ (p < 0.001), V = 0.437 (Moderado)					
Notas ^{ns} No significativo, † Casi significativo (p<.10), * Significativo (p<.05) y ** Altamente significativo (p<.01) Test estadístico/contraste: χ^2 - Chi Cuadrado. Tamaño o magnitud del efecto: V – V de Cramer.					
Concordancia Plena	Concordancia Parcial (Discrepancia de 1 nivel)		Discordancia (Discrepancia de > 1 nivel)		

Existen en total 15 outcomes con TE dispares entre URT/TU® de mas de 1 nivel (Discordancia):

- 1 outcome considerado trivial por TU® es considerado grande por URT
- 14 outcomes considerados pequeño por TU® son considerados grandes por URT.

Resumen de resultados del análisis de las diferencias que encontramos entre los TE Trivial /Grande y Pequeño/Grande entre TU®/URT.

TRIVIAL/GRANDE (N=1):

1. F181: NO PODEMOS ASUMIR QUE LA DISCREPANCIA SEA CIERTA
- a. Outcome: Fasting Blood Glucose.
 - b. No es correcto que el WMD de 1.083 sea GRANDE, al menos clínicamente no es un efecto grande, por lo que no podemos darle un valor estadístico grande. Quizás tampoco trivial.
 - c. Utiliza URT un valor WMD*.

PEQUEÑO/GRANDE (N=14):

1. F015: NO PODEMOS ASUMIR QUE LA DISCREPANCIA SEA CIERTA
- a. Outcome: Respiratory cancer risk.
 - b. el valor utilizado por el URT no figura en ninguna de las referencias del trabajo, por lo que debemos asumir que no es cierto.
 - c. Valor utilizado por URT es OR.
2. F016: NO PODEMOS ASUMIR QUE LA DISCREPANCIA SEA CIERTA
- a. Outcome: Pancreatic Cancer
 - b. el valor grande procede de un valor equivocado.
 - c. Valor utilizado por URT es RR.
3. F028: ASUMIMOS QUE LA DISCREPANCIA SEA CIERTA
- a. Outcome: Esophageal Cancer
 - b. es correcta la diferencia.
 - c. Valor utilizado OR.
4. F099: ASUMIMOS QUE LA DISCREPANCIA SEA CIERTA
- a. Outcome F099: Dynamic Balance
 - b. es correcta la diferencia, aunque con el problema con el significado de 1.76 segundos como algo grande...
 - c. valor que utiliza URT es MD.
5. F184: ASUMIMOS QUE LA DISCREPANCIA SEA CIERTA
- a. Outcome F184: Vit D supplementation on Insulin levels
 - b. es correcta la diferencia.
 - c. Valor utilizado por URT es WMD.
6. F185: ASUMIMOS QUE LA DISCREPANCIA SEA CIERTA
- a. Outcome: F185: VDD. Vit D supplement fasting insulin (FINS)
 - b. es correcta la diferencia.
 - c. Valor utilizado por URT es SMD.
7. F330: ASUMIMOS QUE LA DISCREPANCIA SEA CIERTA
- a. Outcome F330: LRI. Mindfulness-based intervention (MBI)
 - b. es correcta la diferencia.
 - c. Valor utilizado por URT es MD.
8. F331: ASUMIMOS QUE LA DISCREPANCIA SEA CIERTA
- a. Outcome F331: Animal therapy / robopets /
 - b. es correcta la diferencia.
 - c. Valor utilizado por URT es SMD.
9. F333: NO PODEMOS ASUMIR QUE LA DISCREPANCIA SEA CIERTA
- a. Outcome F333: social Support
 - b. no debe ser correcta la diferencia. URT no puede ser GRANDE
 - c. Utiliza una SMD y la transformación a *d* de Cohen resulta difícil de comprender.

10. F380: NO PODEMOS ASUMIR QUE LA DISCREPANCIA SEA CIERTA
- Outcome F380: Hand pain / Resistant Training vs no Resistant Training
 - no debe ser correcta la diferencia. URT no puede ser GRANDE
 - utiliza una SMD que viene bien recogida.
 - Autores de ambas referencias dicen que el efecto es Pequeño.
11. F381: NO PODEMOS ASUMIR QUE LA DISCREPANCIA SEA CIERTA
- Outcome F381: Finger joint stiffness / Physical exercise vs no exercise
 - no debe ser correcta la diferencia. URT no puede ser GRANDE
 - utiliza una SMD que viene bien recogida.
 - Autores de ambas referencias dicen que el efecto es Pequeño
12. F378: NO PODEMOS ASUMIR QUE LA DISCREPANCIA SEA CIERTA
- Outcome F378: Function (follow-up 3-12 m) in thumb carpometacarpal OA / splints vs no splint
 - no debe ser correcta la diferencia. URT no puede ser GRANDE
 - utiliza una SMD que viene bien recogida.
 - Autores de ambas referencias dicen que el efecto es Pequeño
13. F382: NO PODEMOS ASUMIR QUE LA DISCREPANCIA SEA CIERTA
- Outcome F382: Pain in hand OA / Physical exercise vs no exercise
 - no debe ser correcta la diferencia. URT no puede ser GRANDE
 - utiliza una SMD que viene bien recogida.
 - Autores de ambas referencias dicen que el efecto es Pequeño
14. F377: ASUMIMOS QUE LA DISCREPANCIA SEA CIERTA
- Outcome F377: Pain (follow-up 3-12 m) in thumb carpometacarpal OA / splint vs no splint
 - es correcta la diferencia
 - Valor que utiliza URT es SMD.

ANALISIS DETALLADO DEL Rdi TRIVIAL EN TU® y GRANDE EN URT:

- 1. F184 Outcome Trivial de Proyecto 2.** (176)(Marx, 2021). Outcome: Fasting Blood Glucose: F032
- URT WMD: WMD 1.083 (0.100 to 2.065). Cohen *d*: 1.002 (Grande)
 - En TU®: R_{TU} : 0.0174 (0.0174 to 0.0174) (trivial)
 - Outcome: Fasting Blood Glucose.
 - El URT utiliza la referencia 50 (325)(Farhangi, 2020)
 - Analizando el meta-análisis de Farhangi, se encuentra el WMD para Fasting Blood Glucose de 1.083 (IC: 0.099-2.068) (pequeño error en el tercer decimal de intervalo mayor), esa es a la diferencia que se observa entre una dieta de la mayor categoría de índice inflamatorio frente a una dieta con índice bajo, es decir 1.08 mg de glucosa en sangre en ayunas.

- Aunque es un valor significativo, $p=0.031$, no parece un efecto muy grande a nivel clínico.
- TU® utiliza la misma referencia, al ser una WMD solo se pudo utilizar el valor de la p que se trasladó al valor R_{TU} sin intervalo de confianza.
- ANALISIS DETALLADO DE LOS 14 Rdi CLASIFICADOS COMO TE PEQUEÑO EN TU® Y GRANDE EN URT.
- 2. F015 Outcome Pequeño de Proyecto 2,** URT (176)(Marx, 2021). Outcome: Respiratory cancer risk
- Es del proyecto 2 (clasificación en Registro Estudio Base de Datos TU® con etiqueta F015)
- En URT: OR 2.274 (1.894 to 2.729) (grande)
 - En TU®: R_{TU} 0.1257 (0.04 to 0.194) (pequeño)
 - Se corresponde con el outcome: DII Respiratory Cancer Hsc Risk
 - Según el URT el outcome “respiratory cancer” es derivado de la referencia 28, que se corresponde con (326)(Hua, 2020):
 - Si se analiza este metaanálisis se observa que entre sus resultados no aparece el outcome “Respiratory Cancer”. Figuran como se detalla en el título cáncer del tracto aerodigestivos, que incluye: cáncer de esófago, cáncer cavidad oral, cáncer faringe y cáncer laringe.
 - Es curioso que, en el URT de Marx et al., 2021 si mencionan al detalle que los resultados son agregados de 18 estudios con 17,514 participantes.
 - Mientras que en el Metaanálisis de Hua hay 9 estudios casos control con 13,714 participantes.
 - Dentro del URT de Marx et al., 2021 solo existe otra referencia la 36 (327)(Zahedi 2020), donde se menciona el riesgo de cáncer de pulmón y DII.
 - En concreto Mortalidad en la tabla 2 en pacientes normales (no diabetes) HR 1.4 (0.79-2.47), prediabetes HR2.01 (0.93-4.34) y diabetes HR 0.55 (0.09-3.36) todos del mismo estudio referencia 52: (Deng, 2016).
 - y luego en la tabla 4, se mencionó la relación de DII con morbilidad de Cáncer de tracto respiratorios, de 4 estudios con un OR 1.64 (1.11-2.17)
 - ASI QUE, el outcome de Respiratory cancer de Marx et al. (176)(Marx, 2021):
 - no sale de la referencia que dice que utilizan para ello, la referencia 28 (326)(Hua, 2020).
 - No sale de las referencias 29 (328)(Jayedi, 2018) ni 30 (Li, 2018) que se refieren a riesgo de cáncer en general, pero no especifican datos sobre respiratorio.
 - Ni parece que salgan de la referencia 36 (327)(Zahedi 2020), que, si hace referencia a cáncer pulmón y respiratorio, pero con métricas diferentes (HR) o con cifras diferentes OR1.64 (1.11-2.17).
 - PODEMOS ASUMIR QUE EL VALOR DE OR 2.274 (1.894 to 2.729) (grande) PARA RIESGO DE RESPIRATORY CANCER NO ES CIERTO, PERO TAMPOCO PODEMOS

SABER QUE VALOR ES, AL NO SER QUE ASUMAMOS EL OR 1.64 (1.11-2.17). DE Zahedi, 2018, o que hayan agregado datos de Zahedi, algo muy poco probable.

3. F016 Outcome Pequeño de Proyecto 2. URT (176)(Marx, 2021). Outcome F016: DII Pancreatic Cancer

- URT RR 2.524 (1.941 to 3.281). Cohen *d*: 0.925 (Grande)
- TU® R_{TU} 0.1790 (0.046 to 0.302) (Pequeño)
- El URT utiliza la referencia 29 (Jayedi, 2018)
 - En dicha referencia, sin embargo, el RR que aparece es: pancreatic cancer, 1.16 (95% CI: 1.05, 1.28; I2 = 61.6%; n = 2).
- En TU® se utilizan 2 artículos para llegar a los resultados el mismo de Jayedi y otro más reciente de (329)(Guo, 2021)
- DEBEMOS ASUMIR QUE EL VALOR RR 2.524 (1.941 to 3.281). Cohen *d*: 0.925 (Grande) es incorrecto, y debe ser RR 1.16 (95% CI: 1.05, 1.28; I2 = 61.6%; n = 2).

4. F028 Outcome Pequeño de Proyecto 2. (Marx, 2021). Outcome F028: Esophageal Cancer

- URT OR 2.530 (1.738 to 3.682). Cohen *d*: 0.928 (Grande)
- TU® R_{TU} 0.2061 (0.13 to 0.279) (Pequeño)
- El URT utiliza la referencia 28 (326)(Hua, 2020)
 - Analizando la referencia parece correcta la magnitud del OR.
- TU® utiliza para el mismo outcome 3 referencias:
 - (328)(Jayedi, 2018):
 - Existen dos RR:
 - 1.24 (1.10-1.38)
 - 1.26 (1.13-1.39)
 - (326)(Hua, 2020):
 - OR 2.530 (1.738 to 3.682).
 - (330)(Chen, 2020):
 - OR 2.54 (1.90-3.40)
- AQUÍ NOS ENCONTRAMOS CON UN PROBLEMA FRECUENTE EN SINTESIS DE LA EVIDENCIA, y es que cuando se agregan datos de varios estudios, necesariamente los valores individuales con la agregación de magnitudes se van a modificar. De tal manera que el valor final de la media será menor que el valor del valor máximo. Así el URT utiliza solo un dato de una referencia (326)(Hua, 2020): que es uno de las cifras altas, mientras que TU® agrega datos de otros estudios incorporando en la media cifras mas bajas que lógicamente reducirán el valor final.
- POR LO QUE, aunque la cifra del URT parece correcta, la cifra de TU® también lo es.
- ASUMIMOS QUE LA DISCORDANCIA DE CATEGORÍAS ES CIERTA.

5. F099 Outcome Pequeño de Proyecto 3. (167)(Shen, 2022). Outcome F099: Dynamic Balance

- URT MD -1.76 (-2.24 to -1.28). Cohen *d*: 1.718 (Grande)
- TU® R_{TU} 0.1 (0.1 to 0.1) (Pequeño)
- El URT utiliza la referencia 34 (331)(Wu, 2020), diciendo que dice: improved dynamic balance assessed by timed up-and-go test and 8-foot up-and-go test (-1.76 s, 95% CI: -2.24, -1.28).
 - El estudio de (331)(Wu, 2020), dice lo mismo
- TU® utiliza la misma referencia (331)(Wu, 2020)
 - Como es una MD, y solo refiere que es significativo, nos tuvimos que quedar con la p de menor valor.
- TENDREMOS QUE ASUMIR QUE ES CORRECTO, aunque una modificación de 1.76 s en los test que se utilizan habría que saber si eso es GRANDE.

6. F184 Outcome Pequeño de Proyecto 6. (170)(Musazadeh, 2023). Outcome F184: Vit D supplementation on Insulin levels WMD / VDD. Vit D supplement fasting insulin (FINS)

- URT WMD -2.62 (-4.11, -1.13). Cohen *d*: 1,72 (Grande)
- TU® R_{TU} 0.207868(0.151, 0.245). Cohen *d*: 0,429 (Pequeño)
- URT menciona que el resultado proviene de 9 estudios, pero no detalla cual de ellos son, la magnitud es la correcta. No especifica las unidades de niveles de insulina. Las unidades habituales son entre 5-25 Unidades / mililitro. Por lo que la reducción de un 2.62 es entre el 50% y 10%, así que si que parece una reducción grande.
- TU® utiliza 6 referencias:
 - (332)(Emadzadeh, 2020):
 - MD: -2.937; 95 CI: -4.695 to -1.178
 - (333)(Zhang, 2021):
 - SMD = 0.18; 95CI: -0.26, -0.09)
 - (334)(Tang, 2018):
 - fasting insulin by 1.47 mIU/L,
 - (335)(Li, 2018):
 - SMD = 0.18; 95CI: -0.26, -0.09)
 - (336)(Milajerdi, 2019):
 - no effect on insulin
 - (337)(Qi, 2022):
 - fasting insulin (FI) (p = 0.006)
- DEBEMOS asumir que esta correcto esta diferencia Pequeño TU® / Grande URT.
- **7. F185 Outcome Pequeño de Proyecto 6.** (170)(Musazadeh, 2023). Outcome: F185: VDD. Vit D supplement fasting insulin (FINS)

- URT: SMD -0.33 (-0.56 , -0.11). Cohen d : 1.44 (Grande)
- TU[®]: R_{TU} 0.207868(0.151, 0.245). Cohen d : 0.429 (Pequeño)
- URT detalla que viene de 12 estudios sin especificar, la cifra es correcta.
- TU[®] emplea las mismas referencias.
- DEBEMOS asumir que este es correcto.

8. F330 Outcome Pequeño de Proyecto 7. (171)(Veronese 2021(1)). Outcome F330: LRI. Mindfulness-based intervention (MBI)

- URT MD -6.03 (-9.33 to -2.73). Cohen d : 1.791 (Grande)
- TU[®] R_{TU} 0.1 (0.1, 0.1) (Pequeño)
- El URT para este outcome emplea dos artículos:
 - (Hagan, 2014)
 - (Quan, 2019)
- TU[®] por su lado utiliza otra referencia:
 - (338)(Teoh, 2021):
 - MD of -6.33 (CI: -9.39 , -3.26)
 - MD of -6.05 (95 CI: -9.53 , 2.58)

- DEBEMOS asumir que no hay error en esta diferencia.

9. F331 Outcome Pequeño de Proyecto 7. (171)(Veronese 2021(1)). Outcome F331: Animal therapy / robopets / LRI. Animal-Assisted Interventions.

- URT SMD -0.38 (-0.80 to 0.05). Cohen d : 0.876 (Grande)
- TU[®] R_{TU} 0.1142 (0.104, 0.118) (Pequeño)
- URT emplea 3 pero no especifica cual de las 7 RS/MA incluidas son.
- TU[®] emplea 2 artículos:
 - (339)(Hoang, 2022)
 - -1.86 ; 95 CI, -3.14 to -0.59 ; $I^2 = 86$ no figura métrica.
 - (340)(Orr, 2023)
 - No figura métrica solo refiere significación, por lo qué utilizamos el p mas alto
- DEBEMOS ASUMIR como correcta la diferencia entre las d de Cohen.

10. F333 Outcome Pequeño de Proyecto 7 (171)(Veronese 2021(1)). Outcome F333: social Support

- URT SMD -0.13 (-0.25 to -0.01). Cohen d : 1.066 (Grande)
- TU[®] R_{TU} 0.229000(0.359, 0.090). Cohen d : 0.471 (Pequeño)
- URT utiliza 1 referencia (Masi, 2011) que utilizando 9 ECA agrega ese resultado.
- TU[®] utiliza 2 referencias distintas:

- (341)(Grillich, 2023):
 - SMD: 0.00; 95 CI: -0.11 to 0.12 (Valor no significativo)
- (342)(Fu, 2022):
 - SMD = -0.41 [95 CI, -0.70 to -0.13]

- Me cuesta creer que un SMD de -0.13 pueda ser un Cohen GRANDE. Hay que revisar esta transformación. Porque por lo demás parecen resultados muy parecidos.

- Y luego una SMD de TU[®] de -0.41 , aunque le hagamos la media con un SMD de 0.00 sea menor, no es posible que -0.13 sea GRANDE y -0.41 (media con 0.00) sea pequeño.
- POR LO QUE, creo que debemos mirar esa transformación para estar seguros de que está bien.

11. F380 Outcome Pequeño de Proyecto 8 (172)(Veronese, 2021 (2)). Outcome F380: Hand pain / Resistant Training vs no Resistant Training / HOA. Hand Pain (resistance training vs no resistance training)

- URT SMD $0-0.27$ (-0.47 , -0.07). Cohen d : 1.32 (Grande)
- TU[®] R_{TU} 0.1140 (0.02, 0.206) (Pequeño)
- URT esta correcta la cifra de -0.27 . VER SI ESE $0-0.27$ CAUSA ALGUN PROBLEMA. Utiliza dos referencias:
 - (343)(Magni 2017)
 - (344)(Kjeken, 2011):
 - results from single trials indicated that hand exercises may reduce pain
 - limited evidence for the effects of hand exercises and a combination of hand exercises and splints in hand
- TU[®] utiliza uno de las dos referencias que utiliza en URT (343)(Magni 2017). Lo que dice Magni es relevante:
- Low-quality evidence showed significant improvements in joint pain (standardised mean difference (SMD) = -0.23 ; 95 CI = -0.42 , -0.04 ; $I^2 = 0$; $p = 0.02$) which were not clinically relevant.
- Low-quality evidence suggests it has a small, clinically unimportant pain-relieving effect.
- AQUI TENEMOS UN ASUNTO INTERESANTE, las dos referencias utilizadas por el URT, una de ellas también por TU, hablan de un efecto pequeño.
- NO ENTIENDO porque nuestra Cohen d del URT es GRANDE.

12. F381 Outcome Pequeño de Proyecto 8 (172)(Veronese, 2021 (2)). Outcome F381: Finger joint stiffness / Physical exercise vs no exercise / HOA. Finger Joint Stiffness (exercise vs no exercise)

- URT SMD -0.36 (-0.58 , -0.15). Cohen d : 1.64 (Grande)
- TU[®] R_{TU} 0.1170 (0.075, 0.279) (Pequeño)
- URT utiliza 2 referencias:

- $i(344)$ (Kjeken, 2011):
 - results from single trials indicated that hand exercises may reduce pain and increase range of motion and strength, while a combination of splints and daily exercises may reduce pain and stiffness and improve function.
- $i(345)$ (Østerås, 2017):
 - finger joint stiffness (4 trials, SMD -0.36, 95% CI -0.58 to -0.15)
 - Pooled results from 5 studies with low risk of bias showed low-quality evidence for small to moderate beneficial effects of exercise on hand pain, function, and finger joint stiffness postintervention
- TU[®] utiliza 2 referencias (pero es el mismo estudio, una publicada en Cochrane y otra en otra revista)
 - $i(345)$ (Østerås, 2017):
 - $i(346)$ (Østerås, 2017):
 - Four studies (369 participants) indicated that exercise reduced finger joint stiffness (SMD -0.36, 95 CI -0.58 to -0.15) post intervention.
- AQUI NOS VOLVEMOS A ENCONTRAR CON EL PROBLEMA DEL OUTCOME F380, los resultados de los estudios originales refieren que el efecto es pequeño.
- HAY QUE VER PORQUE, el valor de Cohen d sale GRANDE.
- 13. F378 Outcome Pequeño de Proyecto 8** (172)(Veronese, 2021 (2)). Outcome F378: Function (follow-up 3-12 m) in thumb carpometacarpal OA / splints vs no splint / HOA. Thumb Function (splint vs no splint)
 - URT SMD 0.42 (0.08–0.77). Cohen d : 1.19 (Grande)
 - TU[®] R_{TU} 0.1335 (0.093, 0.17) (Pequeño)
 - URT utiliza una referencia:
 - (347)(Buhler, 2019)
 - TU[®] utiliza 2 referencias distintas:
 - (348)(Hamasaki, 2020):
 - No hay cifras, solo refiere significación, por lo que utilizaríamos la p de menos valor.
 - (349)(Meireles, 2019):
 - Effect size -0.44, Confidence Interval 95 -0.72 to -0.15, $p = 0.002$, $I^2 = 0$].
 - AQUI NOS VOLVEMOS A ENCONTRAR CON EL PROBLEMA DEL OUTCOME F380, los resultados de los estudios originales refieren que el efecto es pequeño.
 - HAY QUE VER PORQUE, el valor de Cohen d sale GRANDE.
 - 14. F382 Outcome Pequeño de Proyecto 8** (172)(Veronese, 2021 (2)). Outocme F382: Pain in hand OA / Physical exercise vs no exercise / HOA. Pain (exercise vs no exercise)

- URT SMD -0.23 (-0.42, -0.04). Cohen d : 1.190 (Grande)
- TU[®] R_{TU} 0.134 (0.035, 0.229) (Pequeño)
- URT utiliza 2 referencias:
 - (344)(Kjeken, 2011):
 - results from single trials indicated that hand exercises may reduce pain
 - (346)(Østerås, 2017):
 - low-quality evidence indicated that exercise may improve hand pain [5 trials, standardized mean difference (SMD) -0.27, 95 CI -0.47 to -0.07]
 - Pooled results from 5 studies with low risk of bias showed low-quality evidence for small to moderate beneficial effects of exercise on hand pain
- TU[®] utiliza dos referencias:
 - (345)(Østerås, 2017):
 - (346)(Østerås, 2017):
- AQUI NOS VOLVEMOS A ENCONTRAR CON EL PROBLEMA DEL OUTCOME F380, los resultados de los estudios originales refieren que el efecto es pequeño.
- HAY QUE VER PORQUE, el valor de Cohen d sale GRANDE.
- 15. F377 Outcome Pequeño de Proyecto 8** (172)(Veronese, 2021 (2)). Outcome F377: Pain (follow-up 3-12 m) in thumb carpometacarpal OA / splint vs no splint / HOA. Thumb Pain (splint vs no splint)
 - URT SMD -0.70 (-1.05 to -0.35). Cohen d : 1.960 (Grande)
 - TU[®] R_{TU} 0.1443 (0.195, 0.087) (Pequeño)
 - URT utiliza 1 referencia:
 - 347)(Buhler, 2019)
 - low quality evidence showed that splints cause a moderate to large reduction in pain (SMD 0.7 [95% confidence interval (CI) 1.04, 0.35], $P < 0.0001$)
 - TU[®] utiliza 2 referencias:
 - (348)(Hamasaki, 2020):
 - No hay cifras, solo significación, por lo que usamos la p de menor valor.
 - (349)(Meireles, 2019):
 - Effect size -0.52, Confidence Interval 95 -0.94 to -0.11, $p = 0.01$, $I^2 = 50$
 - AQUI NOS ENCONTRAMOS QUE, el URT si justifica un TE moderado a grande y TU[®] justifica un TE pequeño. EL PROBLEMA es que TU[®] no encontró la referencia utilizada por URT.
- POR LO QUE ACEPTAMOS LA DIFERENCIA.

ANEXO 6

Análisis de dirección y significación estadística del efecto de Rdl por proyectos de forma individualizada, re-análisis URT con Rdl discordantes.

PROYECTO 1. ARTOSIS DE RODILLA Y EVENTOS DE SALUD ADVERSOS

En el proyecto sobre “Artrosis de rodilla y eventos de adversos de salud”, el URT (165) (Veronese, 2023) / código TU® KOA, hay 5 resultados de interés (Rdl) en que las dos metodologías coinciden en la identificación del Rdl. TU® encuentra que la artrosis de rodilla aumenta el riesgo y estadísticamente significativo de todos los Rdl identificados, mientras que el URT solo encuentra aumento de riesgo y significación estadística en 3 de los Rdl (Cardiovascular Mortality, Falls Risk y Subclinical Atherosclerosis Risk), mientras que desconoce el sentido del riesgo en las otras 2 Rdl (All-cause Mortality y Cardiovascular Disease Risk) por obtener un valor no-significativo en el valor de la métrica utilizada para estimar el tamaño de efecto de dicha relación. Dejando el nivel de concordancia de TU® frente a URT en el 60% (3/5) (Tabla 71).

Tabla 71

P1. (Veronese, 2023/(KAO). Concordancia de dirección y significación estadística del efecto del Rdl entre TU® y URT (N=5).

Resultado de Interés	URT dirección	TU dirección	Concordancia TU/URT	URT significación	TU Significación	Concordancia TU/URT
All-cause Mortality	Desconocido	↑ riesgo	TU/URT	No-significativo	Significativo	
Cardiovascular Mortality	↑ riesgo	↑ riesgo		Significativo	Significativo	
Cardiovascular Disease Risk	ORIGINAL: Unknown REVISADO: Higher Risk	↑ riesgo	ORIGINAL: 3/5 (60%) REVISADO: 4/5 (80%)	ORIGINAL: No-significativo REVISADO: Significativo	Significativo	ORIGINAL: 3/5 (60%) REVISADO: 4/5 (80%)
Falls Risk	↑ riesgo	↑ riesgo		Significativo	Significativo	
Subclinical Atherosclerosis Risk	↑ riesgo	↑ riesgo		Significativo	Significativo	

Sin embargo, tras el reanálisis del URT, estamos seguros que el Rdl Cardiovascular Disease Risk esta incorrectamente reflejado en el URT, y es, analizando las dos RS/MA que lo avalan, mostrando un riesgo aumentado y estadísticamente significativo de Cardiovacular Disease Risk en relación con artrosis de rodilla.

De esta manera queda que existe concordancia en la dirección y significación estadística del efecto en 4 de las 5 Rdl (80%), no coincidiendo en “Mortalidad de todas las causas” donde el URT desconoce la dirección del efecto, mientras que TU® refiere que la artrosis de rodilla aumenta el riesgo.

Re-análisis de Veronese, 2023 centrado en los Rdl donde existen discrepancias en el sentido del efecto entre URT y TU®

Haciendo un reanálisis de los resultados del URT de Veronese, 2023, analizando las RS/MA incluidos para los Rdl en discordia, nos encontramos los siguientes hallazgos:

Para calcular el efecto de la artrosis de rodilla en la “Mortalidad de todas las causas”, para este Rdl dos referencias, (350)Leyland, 2021 y (351)Veronese, 2016:

- Leyland, 2021: Esta RS/MA agrega los datos de 7 (6 útiles para analizar) cohortes comunitarias de pacientes con knee OA-related pain (POA), Knee OA radiográfico (ROA) y combinación de ambos (dolor + datos radiográficos) (PROA) y su relación con el tiempo de muerte. Modelos multivariantes muestran los siguientes HR:
 - a. ROA: HR = 1.03; IC95%: 0.83 – 1.28) (no significativo)
 - b. POA: HR = 1.35; IC95%: 1.12 – 1.63) (significativo)
 - c. PROA: HR = 1.37; IC95%: 1.22 – 1.54) (significativo)
- Veronese, 2016: En este MA se analiza la relación entre Osteoartrosis y mortalidad. En cuanto a All-cause Mortality exponen 3 Hazard Ratios que relacionan OA con All-cause Mortality:

- a. Los datos del estudio PRO.V.A; HR = 0.95; IC95%: 0.77 a 1.15 (no significativo)
- b. Análisis cuantitativo (MA) de 7 estudios, HR = 1.10; IC95%: 0.97 – 1.25 (no significativo), retirando OA de mano HR = 1.18; IC95%: 1.08 a 1.28 (significativo).
- c. En la discusión los autores dicen: *There was a significant higher risk of overall mortality for (1) studies conducted in Europe, (2) patients with multi-joint OA; and (3) a radiological diagnosis of OA y The relationship with overall mortality is less clear and may be moderated by the presence of hand OA.*

En los resultados del URT de Veronese 2023, el Rdl All-cause Mortality queda con un HR = 1.18; IC95%: 0.95 a 1.46), por lo que no es significativo. (EN ESTE CASO TENEMOS 3 HR SIGNIFICATIVOS Y 3 NO-SIGNIFICATIVOS; PERO EL VALOR FINAL DEL URT ES NO-SIGNIFICATIVO).

Para calcular el efecto de la artrosis de rodilla en el “Riesgo de enfermedad cardiovascular”, para este Rdl el URT emplea los datos de las RS/MA (352)Macedo, 2022 y (353)Wang, 2016.

- Macedo, 2022: Describe un aumento de riesgo de Hip and Knee OA con Cardiovascular Disease: OR = 1.13; IC95%: 1.05 a 1.22 (significativo).
- Wang, 2016: Describe un riesgo de Cardiovascular Disease aumentado por 24, con un RR = 1.24; IC95%: 1.12 a 1.37 (significativo).
- Es decir que ambos son significativos y en el mismo sentido de aumentar el riesgo.

Con este reanálisis de los RS/MA empleados en el URT de Veronese, 2023, nos encontramos con que el resultado del sentido del efecto y su significación para el Rdl Cardiovascular Disease Risk, no es correcto como viene expresado en el URT. Por lo que corregimos este Rdl, y con ello aumenta el grado de coincidencia entre TU® y el URT a 4/5.

Sin embargo, tras el reanálisis del URT, estamos seguros que el Rdl Cardiovascular Disease Risk esta incorrectamente reflejado en el URT, y es, analizando las dos RS/MA que lo avalan, mostrando un riesgo aumentado y estadísticamente significativo de Cardiovascular Disease Risk en relación con artrosis de rodilla.

De esta manera queda que existe concordancia en la dirección y significación estadística del efecto en 4 de las 5 Rdl (80%), no coincidiendo en “Mortalidad de todas las causas” donde el URT desconoce la dirección del efecto, mientras que TU® refiere que la artrosis de rodilla aumenta el riesgo.

Re-analisis de Veronese, 2023 centrado en los Rdl donde existen discrepancias en el sentido del efecto entre URT y TU®

Haciendo un reanálisis de los resultados del URT de Veronese, 2023, analizando las RS/MA incluidos para los Rdl en discordia, nos encontramos los siguientes hallazgos:

Para calcular el efecto de la artrosis de rodilla en la “Mortalidad de todas las causas”, para este Rdl dos referencias, (350)Leyland, 2021 y (351)Veronese, 2016:

- Leyland, 2021: Esta RS/MA agrega los datos de 7 (6 útiles para analizar) cohortes comunitarias de pacientes con knee OA-related pain (POA), Knee OA radiográfico (ROA) y combinación de ambos (dolor + datos radiográficos) (PROA) y su relación con el tiempo de muerte. Modelos multivariantes muestran los siguientes HR:
 - a. ROA: HR = 1.03; IC95%: 0.83 – 1.28) (no significativo)
 - b. POA: HR = 1.35; IC95%: 1.12 – 1.63) (significativo)
 - c. PROA: HR = 1.37; IC95%: 1.22 – 1.54) (significativo)
- Veronese, 2016: En este MA se analiza la relación entre Osteoartritis y mortalidad. En cuanto a All-cause Mortality exponen 3 Hazard Ratios que relacionan OA con All-cause Mortality:
 - a. Los datos del estudio PRO.V.A; HR = 0.95; IC95%: 0.77 a 1.15 (no significativo)
 - b. Análisis cuantitativo (MA) de 7 estudios, HR = 1.10; IC95%: 0.97 – 1.25 (no significativo), retirando OA de mano HR = 1.18; IC95%: 1.08 a 1.28 (significativo).
 - c. En la discusión los autores dicen: *There was a significant higher risk of overall mortality for (1) studies conducted in Europe, (2) patients with multi-joint OA; and (3) a radiological diagnosis of OA y The relationship with overall mortality is less clear and may be moderated by the presence of hand OA.*

En los resultados del URT de Veronese 2023, el Rdl All-cause Mortality queda con un HR = 1.18; IC95%: 0.95 a 1.46), por lo que no es significativo. (EN ESTE CASO TENEMOS 3 HR SIGNIFICATIVOS Y 3 NO-SIGNIFICATIVOS; PERO EL VALOR FINAL DEL URT ES NO-SIGNIFICATIVO).

Para calcular el efecto de la artrosis de rodilla en el “Riesgo de enfermedad cardiovascular”, para este Rdl el URT emplea los datos de las RS/MA (352)Macedo, 2022 y (353)Wang, 2016.

- Macedo, 2022: Describe un aumento de riesgo de Hip and Knee OA con Cardiovascular Disease: OR = 1.13; IC95%: 1.05 a 1.22 (significativo).
- Wang, 2016: Describe un riesgo de Cardiovascular Disease aumentado por 24, con un RR = 1.24; IC95%: 1.12 a 1.37 (significativo).
- Es decir que ambos son significativos y en el mismo sentido de aumentar el riesgo.

Con este reanálisis de los RS/MA empleados en el URT de Veronese, 2023, nos encontramos con que el resultado del sentido del efecto y su significación para el Rdl Cardiovascular Disease Risk, no es correcto como viene expresado en el URT. Por lo que corregimos este Rdl, y con ello aumenta el grado de coincidencia entre TU y el URT a 4/5.

PROYECTO 2. ÍNDICE DE DIETA INFLAMATORIA Y SALUD HUMANA

En el proyecto sobre el efecto del “Índice de Dieta Inflamatoria sobre la salud humana” (176)(Marx, 2021 / DII) hay en total 30 Rdl en que las dos metodologías coinciden en la identificación de Rdl. En cuanto al sentido del efecto y la significación estadística coinciden en la dirección del sentido en 26/30 (86.67%). Los 4 Rdl en que no coinciden el sentido ni significación estadística del efecto son: Central Obesity, Hyperglycemia, Metabolic Syndrome y Urothelial Cancer.(.).

Tabla 72

P2. (Marx, 2021/(DII). Concordancia de la dirección y significación estadística del efecto de Rdl entre TU® y URT (N=30).

	Resultado de Interés	URT dirección	TU dirección	Concordancia TU/URT	URT significación	TU Significación	Concordancia TU/URT
1	Oral Cancer	↑ riesgo	↑ riesgo	26/30 86.67%)	No-significativo	Significativo	26/30 (86.67%)
2	Respiratory Cancer	↑ riesgo	↑ riesgo		Significativo	Significativo	
3	Pancreatic Cancer	↑ riesgo	↑ riesgo		Significativo	Significativo	
					Significativo	Significativo	
4	Colorectal Cancer	↑ riesgo	↑ riesgo		Significativo	Significativo	
5	Overall Cancer	↑ riesgo	↑ riesgo		Significativo	Significativo	
6	All-cause Mortality	↑ riesgo	↑ riesgo		Significativo	Significativo	
7	Waist Circumference	↑ riesgo	↑ riesgo		Significativo		

8	HbA1C	↑ riesgo	↑ riesgo	26/30 86.67%)	Significativo	Significativo	26/30 (86.67%)
9	Depression	↑ riesgo	↑ riesgo		Significativo	Significativo	
10	Prostate Cancer	↑ riesgo	↑ riesgo		Significativo	Significativo	
11	Pharyngeal Cancer	↑ riesgo	↑ riesgo		Significativo	Significativo	
12	Ovarian Cancer	↑ riesgo	↑ riesgo		Significativo	Significativo	
13	Breast Cancer	↑ riesgo	↑ riesgo		Significativo	Significativo	
14	Lung Cancer	↑ riesgo	↑ riesgo		Significativo	Significativo	
15	Esophageal Cancer	↑ riesgo	↑ riesgo		Significativo	Significativo	
16	Systolic Blood Pressure	↑ riesgo	↑ riesgo		Significativo	Significativo	
17	HOMA-IR	↑ riesgo	↑ riesgo		Significativo	Significativo	
18	Insulin	↑ riesgo	↑ riesgo		Significativo	Significativo	
19	Fasting Blood Glucose	↑ riesgo	↑ riesgo		Significativo	Significativo	
20	Cardiovascular	↑ riesgo	↑ riesgo		Significativo	Significativo	
21	Hypertension	↑ riesgo	↑ riesgo		Significativo	Significativo	
22	Kidney Cancer	↑ riesgo	↑ riesgo		Significativo	Significativo	
23	Gastric Cancer	↑ riesgo	↑ riesgo		Significativo	Significativo	
24	CVD Mortality	↑ riesgo	↑ riesgo		Significativo	Significativo	
25	Cancer Mortality	↑ riesgo	↑ riesgo		Significativo	Significativo	
26	Central Obesity	Desconocido	↑ riesgo		No-significa.	Significativo	
27	Hyperglycemia	Desconocido	↑ riesgo		No-significa.	Significativo	
28	Metabolic Syndrome	Desconocido	↑ riesgo		No-significa.	Significativo	
29	Urothelial Cancer	Desconocido	↑ riesgo		No-significa.	Significativo	
30	Laryngeal Cancer	Desconocido	↑ riesgo		No-significa.	No-significa.	

Re-análisis de Marx, 2021 centrado en los Rdl donde existen discrepancias en el sentido del efecto de URT y TU®

En esta URT, Marx, 2021, hay 4 Rdl en que TU difiere sobre la dirección y significación estadística del efecto:

- Central obesity: URT (unknown) / TU (higher risk)
- Hyperglycemia: URT (unknown) / TU (higher risk)
- Metabolic Syndrome: URT (unknown) / TU (higher risk)
- Urothelial Cancer: URT (unknown) / TU (higher risk)

Central Obesity

Para este Rdl Marx, 2021 utiliza la siguiente referencia:

- Farhangi, 2020 (325).

En esta RS/MA los autores observan lo siguiente con respecto a la obesidad central:

- Meta-analysis of observational studies showed that higher DII® scores were associated with
 - 1.81 cm increase in WC (Pooled weighted mean difference (WMD) = 1.813; CI: 0.785–2.841; p = 0.001). (SIGNIFICATIVO)
 - Also, a non-significant increase in the odds of having higher WC (OR = 1.162; CI: 0.95–1.43; p = 0.154) in the highest DII category was also observed. (NO SIGNIFICATIVO)

Y en conclusiones dicen:

- The findings proposed that adherence to diets with high DII® scores was associated with increased WC.

Nuestra conclusion:

- Marx, 2021 concluye que la DII tiene una relación no significativa con la obesidad central apoyándose en una RS/MA (Farhangi, 2020) que tiene al respecto dos valoraciones; una significativa (una WMD) y otra no significativa (una OR).
- Queda en duda el criterio utilizado: ¿¿¿puede ser que hayan priorizado la métrica de riesgo (OR) a la de valor de variación de medida (WMD)???

Hyperglycemia

Para este Rdl Marx. 2021 utiliza la siguiente referencia:

- Farhangi 2020 (325)

En esta RS/MA los autores observan:

- Que no existe una relación significativa entre DII e Hiperglucemia en el metaanálisis de 9 estudios (OR = 1·13; 95 % CI 0·95, 1·35; P = 0·17), ni siquiera al retirar a uno de los estudios por ser exclusivamente en mujeres, (OR = 1·180; 95 % CI 0·997, 1·395; P = 0·054; I 2 = 49·4 %).

- Sin embargo, en la descripción narrativa de resultados de los estudios primarios incluidos mencionan que incluyen 21 estudios. De los cuales 14 si encuentran una asociación entre DII e hiperglucemia, mientras que 6 no la encuentran y 1 estudio encuentra una asociación inversa.

Por tanto, nuestra valoración es:

- No existe, según datos de agregación en metaanálisis relación entre DII e hiperglicemia. Aunque según datos de revisión narrativa, existen mas estudios que si encuentran una relación entre DII e hiperglicemia.

Metabolic Syndrome

Para este Rdl Marx. 2021 utiliza la siguiente referencia:

- Namazi, 2018 (354)

En esta RS/MA los autores observan:

- no significant association was found between the DII® and the risk for MetS (pooled RR: 1.01; 95% CI: 0.82, 1.24; I2: 32.6%, p=0.20)

Por tanto nuestra valoración:

- es correcta la estimación del URT de Marx, 2021 de que no existe relación significativa entre DII y síndrome metabólico

Urothelial Cancer

Para este Rdl Marx. 2021 utiliza la siguiente referencia:

- Zahedi, 2020 (327)

En esta RS/MA los autores observan:

- Que niveles elevados del DII están asociados a una mayor incidencia de cánceres, incluido el cáncer urotelial urothelial cancers (OR: 1.36; 95% CI: 1.01-1.73). Esta cifra la repiten en el abstract y en el texto
- Sin embargo, en Marx, 2021 el valor del OR dicen que es 1.526 (0.972 – 2.397).
- En la figura 3, aparece el subtotal de 1.50 (0.81 – 2.18) de la agregación de dos estudios (Pierre-Antoine Dugue, 2016: 1.24 (0.90 – 1.70; peso 3.39%) y (Nitin Shivappa, 2016: 1.97 (1.28 – 3.03; peso 1.12%).

Por tanto, nuestra valoración es:

- Parece que la apreciación de Marx, 2021 es correcta, y no existe una relación entre la DII y el cáncer urotelial.
- El error está en el texto de la RS/MA, pero la figura 3 no deja lugar a dudas.

Nuestra valoración sobre los Rdl discordantes es que lo que viene recogido en el URT se sustenta en los resultados presentados por las RS/MA analizados.

PROYECTO 3. EFECTOS DE EJERCICIO EN ANCIANOS CON SARCOPENIA

En el proyecto 3, sobre el “efecto del ejercicio en ancianos con sarcopenia”, incluye el URT de Shen, 2022 (167), y el proyecto TU® con código EXS, se observa que ambas metodologías coinciden en identificar 5 Rdl. El URT identifica el sentido del efecto de 3 de ellas (Knee Extension Strength, Timed Up and Go Test y Dynamic Balance) como que el efecto del ejercicio es favorable y estadísticamente significativo, coincidiendo TU® con esta apreciación; mientras que en 2 de ellas (Grip Strength y Timed Five Chair Stand) el URT desconoce el sentido del efecto por ser no ser estadísticamente significativo, mientras que el TU® reconoce que el sentido del efecto del ejercicio sobre estos Rdl es favorable y estadísticamente significativo (**Tabla 73**).

Tabla 73

P3. (Shen, 2022/(EXS). Coincidencia de sentido y significación de efecto entre TU® y URT en Rdl identificados por ambos métodos (n=5).

	Resultado de Interés	URT Sentido del Efecto	TU Sentido del Efecto	Concordancia TU/URT	URT Significación TE	TU Significación TE	Concordancia TU/URT
1	Grip Strength	TU/URT	Favor Intervención	ORIGINAL: 3/5 (60%) REVISADO: 5/5 (100%)	TU/URT	Significativo	ORIGINAL: 3/5 (60%) REVISADO: 5/5 (100%)
2	Knee Extension Strength	In Favor Intervention	Favor Intervención		Significativo	Significativo	
3	Timed Up and Go Test	Favor Intervención	Favor Intervención		Significativo	Significativo	
4	Timed Five Chair Stand	ORIGINAL: Desconocido REVISADO: Favor intervención	Favor Intervención		ORIGINAL: No significativo REVISADO: Significativo	Significativo	
5	Dynamic Balance	Favor Intervención	Favor Intervención		Significativo	Significativo	

Si se aceptan los valores según vienen reflejados originalmente en el URT el grado de concordancia, en cuanto a la dirección y significación estadística del efecto de los Rdl, es del 60% (3/5). Pero, haciendo una revisión en profundidad de los resultados de Shen, 2022, accediendo a las RS/MA incluidas, tenemos que corregir el sentido y significación en los dos Rdl discordantes según vienen reflejados en el URT. Ambas en una primera evaluación figuran como de sentido de efecto “unknown”, pero la realidad es que son “in favor intervention”, similar a la valoración dada por TU®. El grado de concordancia de la dirección y significación estadística del efecto del Rdl de TU® con URT en este proyecto queda en 5/5 (100%).

Adjunto texto explicativo del análisis en profundidad del URT de Shen, 2022:

En esta URT hay 2 Rdl en que TU® difiere sobre la dirección y significación estadística del efecto:

- Grip Strength: URT (unknown / no-signification) / TU® (in favor intervention / significativo)
- Timed Five Chair Stand: URT (unknown / no-significativo) / TU® (in favor intervention significativo)

Grip Strength

Para este Rdl Shen, 2022 utiliza las siguientes referencias:

- Bao 2020 (355)
- Yoshimura, 2017 (356)
- Yin, 2020 (357)

Bao, 2020 dice respecto a grip strength:

- Que el ejercicio tiene un efecto positivo sobre grip strength [SMD 0.57, 95 % CI 0.42 to 0.73, P <0.00001]. Mientras que Shen, 2022 dice que el resultado es MD= 1.98 (1.18 to 2.78). EN CUALQUIER CASO, AMBOS SON SIGNIFICATIVOS. Aunque no queda claro de dónde sacan el MD = 1.98

Yoshima, 2017 dice al respecto de muscle strength:

- En el abstract no da cifras, pero refiere que el ejercicio puede mejorar fuerza muscular
- Mientras que Shen, 2022 da la cifra de MD = 0.54 (-2.90 to 3.99)
- En el texto Shen, 2022 dice que el ejercicio físico si tiene un efecto positivo sobre la fuerza muscular (MD = 1.98; 95% CI: 1.18 to 2.78 vs. 5 kg, high certainty.

Yin, 2020 dice al respecto de grip strength: (aunque referido a sarcopenia obesidad)

- significant effects on grip strength (exercise: mean difference (MD): 1.63 kg, 95% confidence interval (CI): 0.94, 2.32, P< 0.00001)
- en el texto de Shen, 2022 dice al respecto de grip strength:
 - Exercise intervention increased the grip strength with statistically significant differences but did not exceed the MID threshold (MD: 1.70; 95% CI: 0.36–3.04 vs. 5 Kg). Significativo.

Nuestra conclusion:

- El resultado global de Shen, 2022 con respecto a grip strength y ejercicio físico es que es A FAVOR DE LA INTERVENCIÓN (significativo)..

Five Chair Stand Timed

Para este Rdl Shen, 2022 utiliza la siguiente referencia:

- Bao, 2020 (355)

Bao, 2020 dice al respecto a Timed Five Chair Stand:

- timed five chair stands [SMD -0.56, 95 % CI -0.85 to -0.28, P < 0.0001]. SIGNIFICATIVO
- EN EL URT DE Shen, 2022 dice en la tabla de resultados que Bao, 2020 dice:
- MD = -1.92 (-3.87 to 0.04)
- Accediendo al artículo completo de Bao, 2020 NO FIGURA EL VALOR DE -1.92 (-3.87 TO 0.04) PARA NINGUN OUTCOME, MIENTRAS QUE SI FIGURA EL VALOR DE SMD=-0.56 PARA TIMED FIVE CHAIR STAND
- Nuestra conclusión:
- Para este Rdl (Timed Five Chair Stand), si que existe evidencia significativa de que el ejercicio en ancianos con sarcopenia mejora el test de Timed Five Chair Stand, en las RS/MA utilizados por el URT de Shen, 2022.

Conclusión final: del grado de coincidencia inicial 3/5, pasamos a 5/5. Por un error debido a que el URT presenta datos confusos entre figuras y texto. Y corrección de Timed Five Chair Stand, que el URt dice que es unknown, pero en la RS/MA original figura que si es significativo.

PROYECTO 4. INTERVENCIONES PARA REDUCIR EFECTOS ADVERSOS EN ANCIANOS DESPUES DE ACUDIR A URGENCIAS

En el proyecto sobre “Intervenciones para reducir efectos adversos en ancianos después de acudir a urgencias”, (161)(Conneely, 2022/EDI), hay 7 Rdl en que las dos metodologías coinciden en su identificación. Inicialmente, según figura en la publicación original del URT, el URT y TU® coinciden en el sentido del efecto en 4/7 (57.14%), no coincidiendo en 3/7 (42.85%) que son los siguientes Rdl: ED Revisits, Hospital Admissions y Number of Fallers (Tabla 74).

Tabla 74

P4. (Conneely, 2022/EDI). Coincidencia de sentido y significación de efecto entre TU® y URT en Rdl identificados por ambos métodos (n=7).

	Resultado de Interés	URT Sentido del Efecto	TU Sentido del Efecto	Coincidencia TU/URT	URT Significación TE	TU Significación TE	Coincidencia TU/URT
1	Quality of Life	Desconocido	Desconocido		No-significativo	No-significativo	
2	Mortality	Desconocido	Desconocido		No-significativo	No-significativo	

3	Patient Experience or Satisfaction	Favor intervención	Favor intervención	ORIGINAL: 4/7 (57.14%) REVISADO: 5/7 (71.42%)	Significativo	Significativo	ORIGINAL: 4/7 (57.14%) REVISADO: 5/7 (71.42%)
4	ED Revisits	Desconocido	Favor intervención		No-significativo	No-significativo	
5	Hospital Admissions	Desconocido	Favor intervención		No-significativo	No-significativo	
6	Rate of Falls	Desconocido	Desconocido		No-significativo	No-significativo	
7	Number of Fallers	ORIGINAL: Favor intervención REVISADO: Desconocido	Desconocido		ORIGINAL: Significativo REVISADO: No-Significa	No-significativo	

Si se aceptan los valores según vienen reflejados originalmente en el URT el grado de coincidencia es del 57.14% (4/7). Pero, haciendo una revisión en profundidad de los resultados de Conneely, 2022, accediendo a las RS/MA incluidas, tenemos que corregir el sentido y significación de uno de los Rdl discordantes según vienen reflejados en el URT, “Number of Fallers”. Que pasa de ser a favor de la intervención y estadísticamente significativo a tener un efecto desconocido y no estadísticamente significativo. El grado de concordancia en la dirección y significación estadística del efecto de Rdl de TU® con URT en este proyecto sería de 5/7 (71.42%).

En la reanálisis de los resultados del URT, nos encontramos que el Rdl “Number of Fallers” en la RS/MA original (Morello, 2019)¹, los resultados refieren que no existe una relación significativa entre las intervenciones realizadas en urgencia para reducir el “Number of Fallers”. Por lo que la coincidencia en el sentido de efecto aumenta a 5/7 (71.42%), y en los que no hay coincidencia a 2/7 (28.57%).

Re-analisis de Conneely, 2022 centrado en los Rdl donde existen discrepancias en el sentido del efecto de URT y TU®

En esta URT hay 3 resultados de interés en que TU® difiere sobre la dirección y significación estadística del efecto de Rdl:

- ED revisits: URT (unknown / no-significativo) / TU® (in favor intervention / significativo)
- Hospital Admissions: URT (unknown / no-significativ) / TU® (in favor intervention / significativo)
- Number of fallers: URT (in favor intervention / significativo) / TU® (unknown / no-significativo)

ED revisits

Conneely, 2022 dice que según las 7 referencias incluidas, no hay ningún efecto de ninguna intervención sobre ED revisits. Las referencias son:

- Conroy, 2011(187)
 - CGA interventions readmissions [RR 0.95 (95% CI 0.83-1.08)] (NO SIGNIFICATIVO)
- Fealy, 2009 (358)
 - Nurse intervention failed to demonstrate statistically significant effects on predicted patient and/or health systems outcomes. (NO SIGNIFIATIVO)
- Karam 2015 (359)
 - Resultados escasamente representados, y no significativos (NO SIGNIFICATIVO)
- Hastings, 2005 (360)
 - Results of trials to decrease health service utilization rates following an ED visit were mixed. (NO SIGNIFICATIVO)
- Hughes, 2019 (361)
 - no effects on QOL, patient experience, hospitalization at or after the initial ED index visit, or ED return visit. (NO SIGNIFICATIVO)
- Lowthian, 2015 (362)
 - the evidence indicates no appreciable benefit for ED-CTS for unplanned ED re-attendance up to 30 days (odds ratio (OR) 1.32, 95% confidence interval (CI) 0.99-1.76; n = 1,389) (NO SIGNIFICATIVO)
- Malik, 2018 (363)
 - Geriatric focused nursing assessment and interventions did not have a statistical impact on hospitalization, readmissions, LOHS and ED revisits. (NO SIGNIFICATIVO)

Hospital Admissions

El URT de Conneely,2022 utiliza 5 referencias:

- Karam, 2015 (359)
- Hastings, 2005 (360)
 - Results of trials to decrease health service utilization rates following an ED visit were mixed. (NO SIGNIFICATIVO)
- Hughes, 2019 (361)
 - no effects on QOL, patient experience, hospitalization at or after the initial ED index visit, or ED return visit. (NO SIGNIFICATIVO)

- Lowthian, 2015 (362)
 - the evidence indicates no appreciable benefit for ED-CTS for unplanned ED re-attendance up to 30 days (odds ratio (OR) 1.32, 95% confidence interval (CI) 0.99-1.76; n = 1,389) (NO SIGNIFICATIVO)
- Malik, 2018 (363)
 - Geriatric focused nursing assessment and interventions did not have a statistical impact on hospitalization, readmissions, LOHS and ED revisits. (NO SIGNIFICATIVO)

Number of Fallers

El URT de Conneely, 2022 utiliza una única RS/MA para analizar este Rdl:

- Morello, 2019 (364)
- Aunque Conneely dice que esta RS/MA dice: There was moderate quality evidence for reducing the number of fallers in this systematic review, which was of critically low quality. Es decir que hay un efecto significativo favorable a la intervención, pero de baja calidad...
- La realidad es que Morello, 2019 dice:
- Meta-analyses demonstrated no reduction in falls (rate ratio = 0.78; 95% CI: 0.58 to 1.05), number of fallers (risk ratio = 1.02; 95% CI: 0.88 to 1.18). NO-SIGNIFICATIVO.
- Nuestra conclusión es que el URT diga que es significativo a favor de la intervención, no está ajustado a la realidad publicada en la RS/MA de Morello, 2019. PASARIA DE SER "IN FAVOR INTERVENTION / SIGNIFICATIVO" A "UNKNOWN / NO-SIGNIFICATIVO".

PROYECTO 5. EFECTO DE DIGOXINA SOBRE MORTALIDAD EN FIBRILACION AURICULAR

En el proyecto sobre el "Efecto de digoxina sobre la mortalidad en fibrilación auricular", (169)(Gazzaniga, 2023/DGM), hay 4 Rdl en que las dos metodologías coinciden en su identificación. Además, existe coincidencia plena 4/4 (100%) en el sentido y significación estadística del efecto entre las dos metodologías (**Tabla 75**).

Tabla 75

P5. (Gazzaniga, 2023/DGM). Coincidencia de sentido y significación de efecto entre TU® y URT en Rdl identificados por ambos métodos (n=4).

	Resultado de Interés	URT dirección efecto	TU dirección Efecto	Concordancia TU/URT	URT Significación	TU Significación	Concordancia TU/URT
1	All-cause Mortality in whole population	↑ riesgo	↑ riesgo	4/4 (100%)	Significativo	Significativo	4/4 (100%)
2	Cardiovascular Mortality	↑ riesgo	↑ riesgo		Significativo	Significativo	
3	All-cause Mortality patients AF only	↑ riesgo	↑ riesgo		Significativo	Significativo	
4	All-cause Mortality patients AF + HF	↑ riesgo	↑ riesgo		Significativo	Significativo	

PROYECTO 6. EFECTOS DE SUPLEMENTOS DE VITAMINA D SOBRE BIOMARCADORES DE DIABETES MELLITUS TIPO 2

En el proyecto sobre el "Efectos de suplementos de vitamina D sobre biomarcadores de diabetes mellitus tipo 2", (170)(Musazadeh, 2023/VDD), hay 8 Rdl en que las dos metodologías coinciden en su identificación. Además, existe coincidencia plena 8/8 (100%) en el sentido y significación estadística del efecto entre las dos metodologías (**Tabla 76**).

Tabla 76

P6. (Musazadeh, 2023/VDD). Coincidencia de sentido y significación de efecto entre TU® y URT en Rdl identificados por ambos métodos (n=8).

	Resultado de Interés	URT dirección efecto	TU dirección efecto	Coincidencia TU/URT	URT Significación TE	TU Significación TE	Coincidencia TU/URT
1	Vit D supplementation on FBG levels WMD	Favor intervención	Favor intervención	8/8 (100%)	Significativo	Significativo	8/8 (100%)
2	SMD Vit D supplementation on FBS levels	Favor intervención	Favor intervención		Significativo	Significativo	
3	Vit D supplementation on HbA1c levels WMD	Favor intervención	Favor intervención		Significativo	Significativo	
4	SMD Vit D supplementation on HbA1c levels	Favor intervención	Favor intervención		Significativo	Significativo	
5	Vit D supplementation on HOMA-IR levels WMD	Favor intervención	Favor intervención		Significativo	Significativo	

6	SMD Vit D supplementation on HOMA-IR levels	Favor intervención	Favor intervención	8/8 (100%)	Significativo	Significativo	8/8 (100%)
7	Vit D supplementation on Insulin levels WMD	Favor intervención	Favor intervención		Significativo	Significativo	
8	SMD Vit D supplementation on Insulin levels	Favor intervención	Favor intervención		Significativo	Significativo	

PROYECTO 7. INTERVENCIONES PARA REDUCIR LA SOLEDAD

En el proyecto sobre “Intervenciones para reducir la soledad”, (171)(Veronese, 2021(1) / LRI), hay 6 Rdl en que las dos metodologías coinciden en su identificación. Inicialmente el URT y TU® coinciden en el sentido y significación estadística del efecto en 4/6 (66.67%), no coincidiendo en 2/6 (33.33%) que son los siguientes Rdl: Technological interventions y Animal therapy / robopets (Tabla 77).

Tabla 77

P7. (Veronese, 2021(1) /LRI). Coincidencia de sentido y significación de efecto entre TU® y URT en Rdl identificados por ambos métodos (n=6).

	Resultado de Interés	URT Sentido del Efecto	TU Sentido del Efecto	Coincidencia TU/URT	URT Significación TE	TU Significación TE	Coincidencia TU/URT
1	Befriending Interventions	Desconocido	Desconocido	4/6 (66.67%)	No-significa.	No-significa.	8/8 (4/6 (66.67%)%)
2	Technological interventions	Desconocido	Favor Intervención		No-significa.	Significativo	
3	Meditation / Mindfulness	Favor Intervención	Favor Intervención		Significativo	Significativo	
4	Animal therapy / robopets	Desconocido	Favor Intervención		No-significa.	Significativo	
5	Social cognitive training	Favor Intervención	Favor Intervención		Significativo	Significativo	
6	Social support	Favor Intervención	Favor Intervención		Significativo	Significativo	
	Social support						

En el análisis detallado del URT, no objetivamos datos que sugieran que los resultados del URT no estén en línea con los RS/MA analizados por el URT.

Re-análisis de Veronese, 2021(1) centrado en los Rdl donde existen discrepancias en el sentido del efecto de URT y TU

En el URT de Veronese, 2021 (1) existen dos Rdl en que no coinciden el sentido del efecto:

- Technological interventions: URT (unknown) / TU® (in favor intervention)
- Animal therapy / robopets: URT (unknown) / TU® (in favor intervention)

Technological Interventions

Hay que ir analizando uno a uno las RS/MA incluidas:

Abbott 2019 (365)

- Efecto de Robopets: para soledad: SMD: -0.51 (-1.24 a 0.22) (NO SIGNIFICATIVO) Choi, 2012 (366)
- ANALIZAN interventions tecnolgicas, y para soledad obtienen un efecto significativo de tamaño de efecto medio ponderado de g de Hedges de 0.546 (0.033 a 1.059) (SIGNIFICATIVO) Cohen-Mansfield, 2015 (367)
- ANALIZAN educational interventions focused on social networks maintenance (NO HAY DISCORDANCIA ENTRE TU Y URT EN ESTE RDI) Hagan, 2014 (368)
- interventions tecnolgicas En sus resultados dice: Three studies reporting on new technologies and one on a group work intervention identified significant reductions in loneliness. (SIGNIFICATIVO) Masi, 2011 (369)
- Intervenciones tecnologicas en grupos de edad mas avanzada, todas las relaciones son NO SIGNIFICATIVAS Quan, 2019 (370)
- Solo acceso a abstract, refiere efecto pero no de las intervenciones donde hay discrepancias entre TU® y URT Siette, 2017 (371)
- Se refiere solo a befriending, Rdl sin conflicto entre TU® y URT
- There was moderate quality evidence to support the use of befriending for the treatment of individuals with different physical and mental health conditions.

Nuestra conclusión:

Para Robopets URT dice que no es significativo, análisis en profundidad de la RS/MA corrobora ese dato, pero solo de 1 de los 3 RS/MA a los que se refiere.

Para intervenciones tecnológicas:

- Hay dos RS/MA a favor
- Otra no significativa
- Falta la 4ª RS/MA que hace referencia en el URT

Por lo que no cambiamos el sentido de la intervención en technological intervnetions ni robopets de la dirección ni significación estadística del efecto de Rdl en este URT.

PROYECTO 8. EFICACIA DE TRATAMIENTOS CONSERVADORES PARA ARTROSIS DE MANOS

En el proyecto sobre el “Eficacia de tratamientos conservadores para artrosis de manos”, (172)(Veronese, 2021(2) /HOA), hay 8 Rdl en que las dos metodologías coinciden en su identificación. Además, existe concordancia plena 8/8 (100%) en el sentido y significación estadística del efecto entre las dos metodologías (Tabla 78).

Tabla 78

P8. (Veronese, 2021(2) /HOA). Coincidencia de sentido y significación de efecto entre TU® y URT en Rdl identificados por ambos métodos (n=8).

	Resultado de Interés	URT dirección efecto	TU dirección efecto	Concordancia TU/URT	URT Significación	TU Significación	Concordancia TU/URT
1	Pain (follow-up 3-12 m) in thumb carpometacarpal OA / splint vs no splint	Favor intervención	Favor intervención	8/8 (100%)	Significativo	Significativo	8/8 (100%)
2	Function (follow-up 3-12 m) in thumb carpometacarpal OA / splints vs no splint	Favor intervención	Favor intervención		Significativo	Significativo	
3	Pain (long follow-up) in symptomatic hand OA / splints vs no splint	Favor intervención	Favor intervención		Significativo	Significativo	
4	Hand pain / Resistant Training vs no Resistant Training	Favor intervención	Favor intervención		Significativo	Significativo	

5	Finger joint stiffness / Physical exercise vs no exercise	Favor intervención	Favor intervención	8/8 (100%)	Significativo	Significativo	8/8 (100%)
6	Pain in hand OA / Physical exercise vs no exercise	Favor intervención	Favor intervención		Significativo	Significativo	
7	Pain in trapeziometacarpal OA / Physical and occupational therapy vs no intervention	Favor intervención	Favor intervención		Significativo	Significativo	
8	Function in thumb OA / IAHA (intra-articular hyaluronic acid) vs placebo	Favor intervención	Favor intervención		Significativo	Significativo	

ANEXO 7

Análisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU® según las dos categorías profesionales participantes (médico / enfermero).

Análisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU® según las dos categorías profesionales participantes (médico / enfermero).

Presentamos los datos para las dos categorías profesionales sobre frecuencia de valoración (1 – 5) recibidas por cada ítem de forma individualizado por ítem y de forma global por valoración. En la categoría Médico (N=104) se observan de manera global 571 valoraciones 5 (54.9%) y 329 valoraciones 4 (31.63%) (**Tabla 79**) (**Figura 45**). En la categoría Enfermeros (N=17) se observan 77 valoraciones 5 (45,29%) y 66 valoraciones 4 (38.82%) (**Tabla 80**) (**Figura 46**). Lo que supone valoraciones de alta puntuación (4 y 5) en el 86.53% de la categoría Médico y 84.11% en la categoría Enfermero. Por otro, lado las puntuaciones bajas supusieron 25 valoraciones 1 (2.4%) y 7 valoraciones 2 (0.67%) en la categoría Médico, y 0 valoraciones 1 (0.00%) y 1 valoración 2 (0.28%) en la categoría Enfermero.

En el análisis de los ítems mejor valorados por las dos categorías profesionales, basándonos en aquellos ítems con frecuencias más altas en valoración 5, encontramos que la categoría Médico el ítem P. 14 (Evolución histórica es útil) y el P. 16 (Satisfacción con metodología TU) son los mejor valorados con valoraciones 5 en 63 ocasiones, representando el 60,57% para ambos ítems. Mientras que en la categoría profesional Enfermero, los ítems P. 11 (Interpretación más rápida de resultados), P. 12 (Es útil en toma decisiones) y P. 13 (Facilita la exploración de datos) recibieron 9 valoraciones 5 que representa el 52,94% dentro de los tres ítems.

Tabla 79

Frecuencia de valoración a cada ítem de la encuesta de profesionales Médicos (N=104).

Pregunta (N.º)	Valoración 1 (%)	Valoración 2 (%)	Valoración 3 (%)	Valoración 4 (%)	Valoración 5 (%)
P. 7	3 (2.88)	-	11 (10.57)	40 (38.46)	50 (48.07)
P. 8	2 (1.92)	1 (0.96)	8 (7.69)	35 (33.65)	58 (55.76)
P. 9	2 (1.92)	1 (0.96)	11 (10.57)	38 36.53)	52 (50.00)
P. 10	2 (1.92)	1 (0.96)	9 (8.65)	37 (35.57)	55 (52.88)
P. 11	3 (2.88)	-	10 (9.61)	34 (32.69)	57 (54.80)
P. 12	3 (2.88)	-	11 (10.57)	32 (30.76)	58 (55.76)
P. 13	3 (2.88)	-	10 (9.61)	36 (34.61)	55 (52.88)
P. 14	2 (1.92)	1 (0.96)	15 (14.42)	23 (22.11)	63 (60.57)
P. 15	2 (1.92)	3 (2.88)	14 (13.46)	25 (24.03)	60 (57.69)
P. 16	3 (2.88)	-	9 (8.65)	29 (27.88)	63 (60.57)
Total	25 (2.4%)	7 (0.67%)	108 (10.38%)	329 (31.63%)	571 (54.9%)

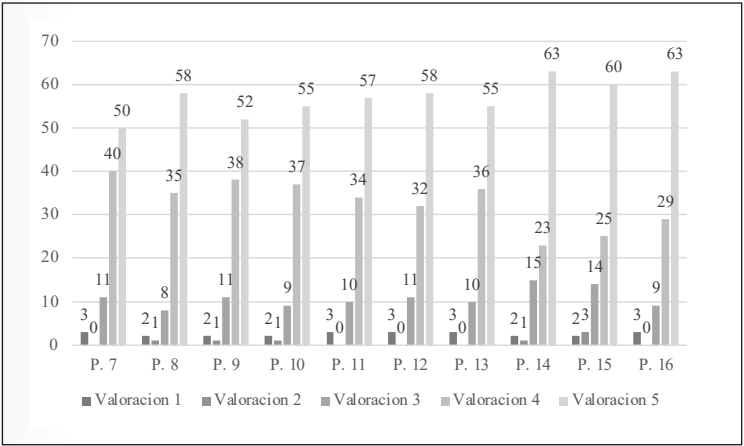


Figura 45

Frecuencia de los valores desglosados de cada ítem de la encuesta de profesionales Médicos (N=104).

Tabla 80

Frecuencia de valoración a cada ítem de la encuesta de profesionales Enfermeros (N=17).

Pregunta (N.º)	Valoración 1 (%)	Valoración 2 (%)	Valoración 3 (%)	Valoración 4 (%)	Valoración 5 (%)
P. 7	-	-	5 (29.41)	4 (23.52)	8 (47.05)
P. 8	-	-	1 (5.88)	11 (64.70)	5 (29.41)
P. 9	-	-	1 (5.88)	9 (52.94)	7 (41.17)
P. 10	-	-	1 (5.88)	9 (52.94)	7 (41.17)
P. 11	-	-	4 (23.52)	4 (23.52)	9 (52.94)
P. 12	-	-	3 (17.64)	5 (29.41)	9 (52.94)
P. 13	-	-	2 (11.76)	6 (35.29)	9 (52.94)
P. 14	-	1 (5.88)	3 (17.64)	5 (29.41)	8 (47.05)
P. 15	-	-	4 (23.52)	6 (35.29)	7 (41.17)
P. 16	-	-	2 (11.76)	7 (41.17)	8 (47.05)
Total	0 (0%)	1 (0.58%)	26 (15.29%)	66 (38.82%)	77 (45.29%)

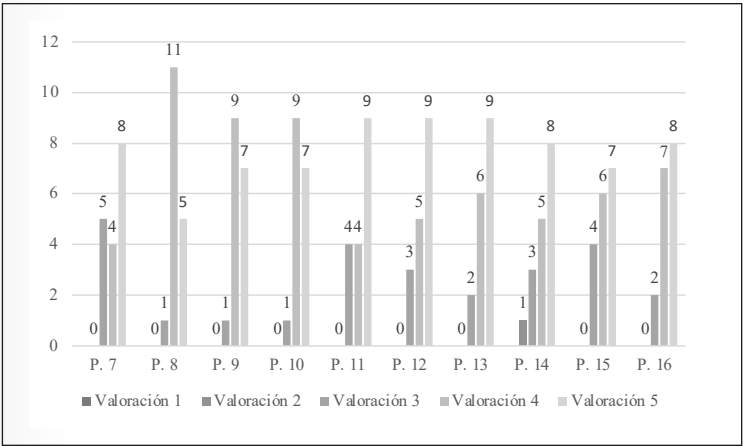


Figura 46

Frecuencia de los valores desglosados de cada ítem de la encuesta de profesionales Enfermeros (N=17).

ANEXO 8

Análisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU® según género (mujer / hombre).

Análisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU® según género (mujer / hombre).

Presentamos los datos para las dos categorías de género (mujer y hombre) sobre frecuencia de valoración (1 – 5) recibidas por cada ítem de forma individualizada por ítem y de forma global por valoración.

En la categoría mujer (N=97) (**Tabla 81**) (**Figura 47**) se observan de manera global 487 valoraciones 5 (50.20%) y 346 valoraciones 4 (35.67%); mientras que en la categoría hombre (N=24) (**Tabla 82**) (**Figura 48**) se observan 161 valoraciones 5 (67.08%) y 49 valoraciones 4 (20.41%). Lo que supone valoraciones de alta puntuación (4 y 5) en el 85.87% de las mujeres y 87.49% en los hombres.

Las puntuaciones bajas supusieron 16 valoraciones 1 (1.64%) y 7 valoraciones 2 (0.72%) en mujeres, y 9 valoraciones 1 (3.75%) y 1 valoración 2 (0.41%) en hombres. En el análisis de los ítems mejor valorados por los dos sexos, basándonos en aquellos ítems con frecuencias más altas en valoración 5, encontramos que, en la mujer, el ítem P. 14 (Evolución histórica es útil) y el ítem P. 16 (Satisfacción con la metodología) son los mejor valorados con valoraciones 5 en ambos ítems de 53 ocasiones, representando el 54.63%. De la misma manera, en los hombres, los ítems P. 14 (Evolución histórica es útil) y P. 16 (Satisfacción con la metodología) recibieron 18 valoraciones 5, representando el 75.0% en ambos ítems.

Tabla 81

Frecuencia de valores a cada pregunta de la encuesta en mujeres (N=97).

Pregunta (N.º)	Valoración 1 (%)	Valoración 2 (%)	Valoración 3 (%)	Valoración 4 (%)	Valoración 5 (%)
P. 7	2 (2.06%)	-	14 (14.43%)	37 (38.14%)	44 (45.36%)
P. 8	1 (1.03%)	1 (1.03%)	7 (7.21%)	40 (41.23%)	48 (49.48%)
P. 9	1 (1.03%)	1 (1.03%)	9 (9.27%)	42 (43.29%)	44 (45.36%)
P. 10	1 (1.03%)	1 (1.03%)	9 (9.27%)	40 (41.23%)	46 (47.42%)
P. 11	2 (2.06%)	-	12 (12.37%)	34 (35.05%)	49 (50.51%)
P. 12	2 (2.06%)	-	12 (12.37%)	32 (32.98%)	51 (52.57%)
P. 13	2 (2.06%)	-	10 (10.31%)	38 (39.17%)	47 (48.45%)
P. 14	1 (1.03%)	2 (2.06%)	15 (15.46%)	26 (26.80%)	53 (54.63%)
P. 15	2 (2.06%)	2 (2.06%)	15 (15.46%)	26 (26.80%)	52 (53.61%)
P. 16	2 (2.06%)	-	11 (11.34%)	31 (31.95%)	53 (54.63%)
Total	16 (1.64%)	7 (0.72%)	114 (11.75%)	346 (35.67%)	487 (50.20%)

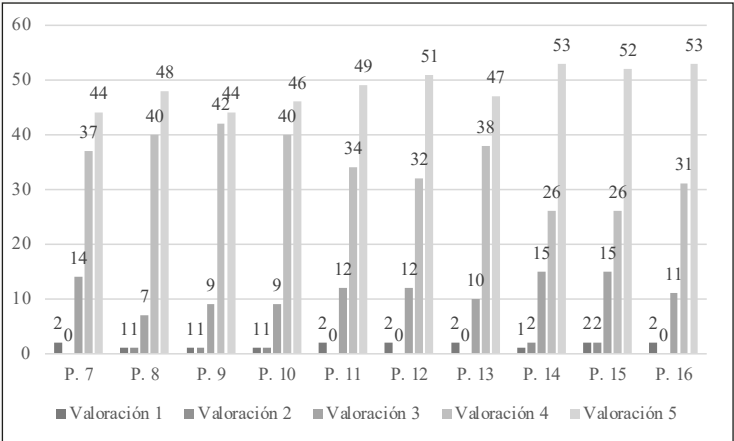


Figura 47

Frecuencia de los valores desglosados de cada ítem de la encuesta en mujeres (N=96).

Tabla 82

Frecuencia de valores a cada pregunta de la encuesta en hombres (N=24).

Pregunta (N.º)	Valoración 1 (%)	Valoración 2 (%)	Valoración 3 (%)	Valoración 4 (%)	Valoración 5 (%)
P. 7	1 (4.16%)	-	2 (8.33%)	7 (29.16%)	14 (58.33%)
P. 8	1 (4.16%)	-	2 (8.33%)	6 (25.0%)	15 (62.5%)
P. 9	1 (4.16%)	-	3 (12.5%)	5 (20.83%)	15 (62.5%)
P. 10	1 (4.16%)	-	1 (4.16%)	6 (25.0%)	16 (66.66%)
P. 11	1 (4.16%)	-	2 (8.33%)	4 (16.66%)	17 (70.83%)
P. 12	1 (4.16%)	-	2 (8.33%)	5 (20.83%)	16 (66.66%)
P. 13	1 (4.16%)	-	2 (8.33%)	4 (16.66%)	17 (70.83%)
P. 14	1 (4.16%)	-	3 (12.5%)	2 (8.33%)	18 (75.0%)
P. 15	-	1 (4.16%)	3 (12.5%)	5 (20.83%)	15 (62.5%)
P. 16	1 (4.16%)	-	-	5 (20.83%)	18 (75.0%)
Total	9 (3.75%)	1 (0.41%)	20 (8.33%)	49 (20.41%)	161 (67.08%)

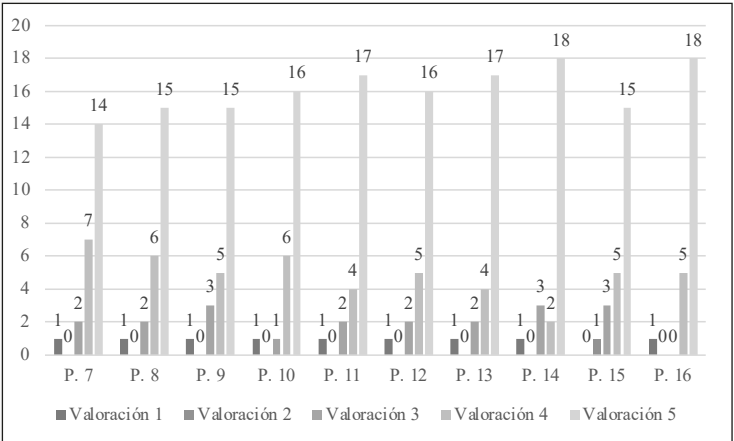


Figura 48

Frecuencia de los valores desglosados de cada ítem de la encuesta en hombres (N=24).

ANEXO 9

Análisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU® por grupos de edad.

Análisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU® por grupos de edad.

Presentamos los datos de frecuencia de valoración a cada ítem de la encuesta, diferenciados por grupos de edad.

En el grupo de edad menor de 33 años (N=68), (Tabla 83) (Figura 49) las frecuencias de valoración más altas se concentraron en las categorías 4 y 5, con un total de 215 (31.61%) y 354 (52.05%) valoraciones respectivamente. Las valoraciones 1 y 2 fueron mínimas, con 25 (3.67%) y 8 (1.17%) respectivamente. Destacaron los ítems P. 12 (Utilidad para tomar decisiones) y P. 16 (Satisfacción con la metodología), con 40 valoraciones 5 (58.82%) y 39 valoraciones 5 (57.35%), respectivamente. El ítem P. 10 (Facilidad más la exploración de los datos) también obtuvo una alta valoración con 34 valoraciones 5 (50.00%). En este grupo, las valoraciones 4 y 5 suman un total del 83.66%.

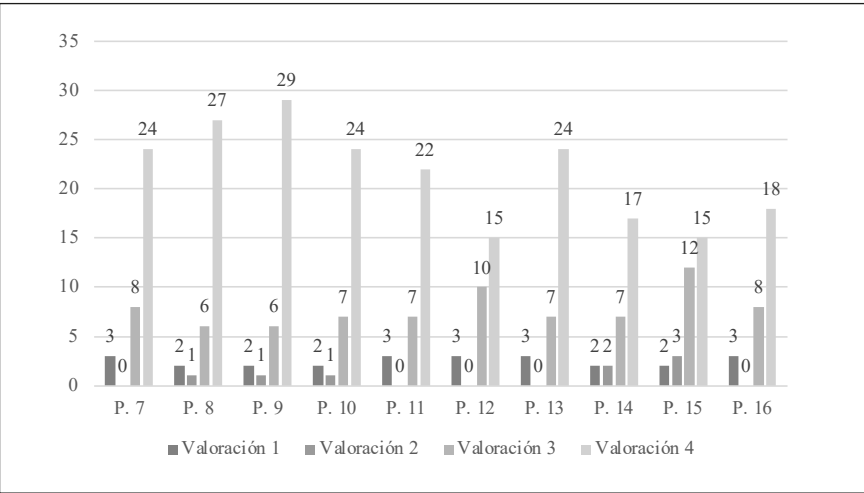
Tabla 83

Frecuencia de valores a cada ítem de la encuesta de utilidad de TU® en el grupo de edad < 33 años (N=68).

Pregunta (N.º)	Valoración 1 (%)	Valoración 2 (%)	Valoración 3 (%)	Valoración 4 (%)	Valoración 5 (%)
P. 7	3 (4.41)	-	8 (11.76)	24 (35.29)	33 (48.52)
P. 8	2 (2.94)	1 (1.47)	6 (8.82)	27 (39.70)	32 (47.05)
P. 9	2 (2.94)	1 (1.47)	6 (8.82)	29 (42.64)	30 (44.11)
P. 10	2 (2.94)	1 (1.47)	7 (10.29)	24 (35.29)	34 (50.00)
P. 11	3 (4.41)	-	7 (10.29)	22 (32.35)	36 (52.94)
P. 12	3 (4.41)	-	10 (14.70)	15 (22.05)	40 (58.82)
P. 13	3 (4.41)	-	7 (10.29)	24 (35.29)	34 (50.00)
P. 14	2 (2.94)	2 (2.94)	7 (10.29)	17 (25.00)	40 (58.82)
P. 15	2 (2.94)	3 (4.41)	12 (17.64)	15 (22.05)	36 (52.94)
P. 16	3 (4.41)	-	8 (11.76)	18 (26.47)	39 (57.35)
Total	25 (3.67)	8 (1.17)	78 (11.47)	215 (31.61)	354 (52.05)

Figura 49

Frecuencia de los valores a cada ítem de la encuest de utilidad de TU® en el grupo de edad < 33 años (N=68).



En el grupo de edad de 33 a 43 años (N=26), (Tabla 84) (Figura 50) predominaron las valoraciones 4 y 5, con 116 (44.61%) y 102 (39.23%) valoraciones respectivamente. Las valoraciones 1 y 2 estuvieron ausentes. Este grupo mostró altas valoraciones en los ítems P. 8 (Bubble plots ayuda identificar resultados) y P. 15 (Se requieren menos conocimientos), con 13 (50.00%) y 15 (57.69%) valoraciones 5 respectivamente. El ítem P. 7 (Facilidad comprensión) también obtuvo una valoración destacada con 13 valoraciones 4 (50.00%). Las valoraciones 4 y 5 representan el 83.84% de las respuestas.

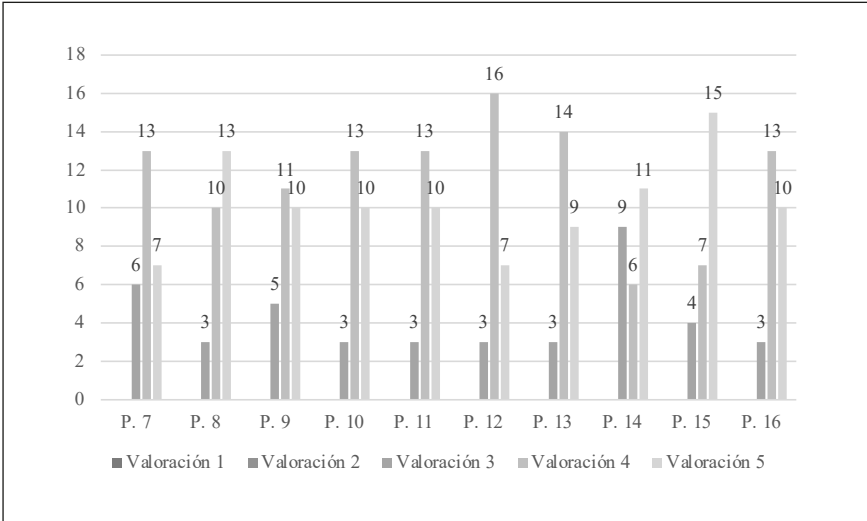
Tabla 84

Frecuencia de valores a cada ítem de la encuesta de utilidad de TU® en el grupo de edad 33-43 años (N=26).

Pregunta (N.º)	Valoración 1 (%)	Valoración 2 (%)	Valoración 3 (%)	Valoración 4 (%)	Valoración 5 (%)
P. 7	-	-	6 (23.07)	13 (50.00)	7 (26.92)
P. 8	-	-	3 (11.53)	10 (38.46)	13 (50.00)
P. 9	-	-	5 (19.23)	11 (42.30)	10 (38.46)
P. 10	-	-	3 (11.53)	13 (50.00)	10 (38.46)
P. 11	-	-	3 (11.53)	13 (50.00)	10 (38.46)
P. 12	-	-	3 (11.53)	16 (61.53)	7 (26.92)
P. 13	-	-	3 (11.53)	14 (53.84)	9 (34.61)
P. 14	-	-	9 (34.61)	6 (23.07)	11 (42.30)
P. 15	-	-	4 (15.38)	7 (26.92)	15 (57.69)
P. 16	-	-	3 (11.53)	13 (50.00)	10 (38.46)
Total	0 (0.00%)	0 (0.00%)	42 (16.15)	116 (44.61)	102 (39.23)

Figura 50

Frecuencia de los valores a cada ítem de la encuesta de utilidad de TU® en el grupo de edad 33 a 43 años (N=26)



El grupo de edad de 44 a 54 años (N=11) (**Tabla 85**) (**Figura 51**) mostró una clara preferencia por las valoraciones más altas, con 26 (23.63%) valoraciones 4 y 83 (75.45%) valoraciones 5. No se registraron valoraciones 1 o 2 en este grupo. Los ítems P. 16 (Satisfacción con la metodología) recibió 10 valoraciones 5 (90.90%), y P. 13 (Facilita más la exploración de los datos) y P. 14 (Evolución histórica) recibieron 9 valoraciones 5 (81.81%). El 99.08% de las respuestas se encuentran en las categorías 4 y 5.

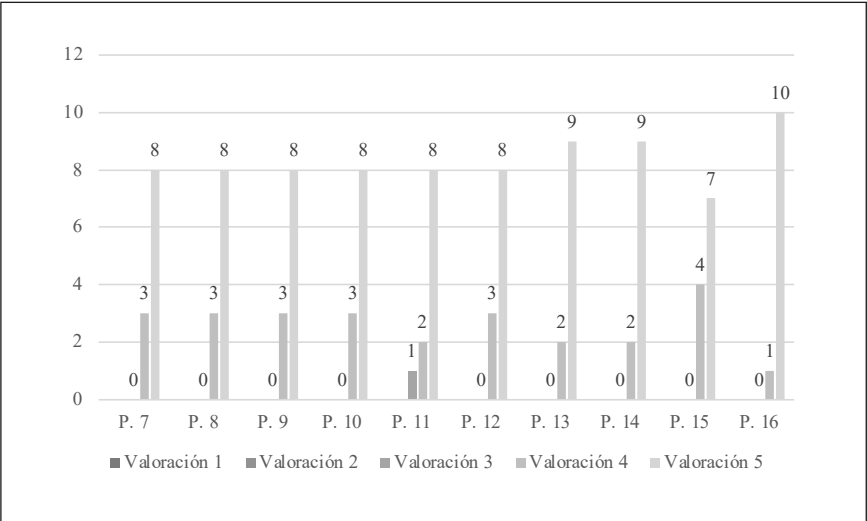
Tabla 85

Frecuencia de valores a cada ítem de la encuesta de utilidad de TU® en el grupo de edad 43 a 54 años (N=11).

Pregunta (N.º)	Valoración 1 (%)	Valoración 2 (%)	Valoración 3 (%)	Valoración 4 (%)	Valoración 5 (%)
P. 7	-	-	-	3 (27.27)	8 (72.72)
P. 8	-	-	-	3 (27.27)	8 (72.72)
P. 9	-	-	-	3 (27.27)	8 (72.72)
P. 10	-	-	-	3 (27.27)	8 (72.72)
P. 11	-	-	1 (9.09)	2 (18.18)	8 (72.72)
P. 12	-	-	-	3 (27.27)	8 (72.72)
P. 13	-	-	-	2 (18.18)	9 (81.81)
P. 14	-	-	-	2 (18.18)	9 (81.81)
P. 15	-	-	-	4 (36.36)	7 (63.63)
P. 16	-	-	-	1 (9.09)	10 (90.90)
Total	0 (0.00%)	0 (0.00%)	1 (0.90%)	26 (23.63)	83 (75.45)

Figura 55

Frecuencia de los valores a cada ítem de la encuesta de utilidad de TU® en el grupo de edad 43 a 54 años (N=11).



En el grupo de edad de 55 a 65 años (N=16), (**Tabla 86**) (**Figura 52**) las valoraciones 4 y 5 también fueron predominantes, con 36 (22.50%) y 109 (68.12%) respectivamente. No se registraron valoraciones 1 o 2. Los ítems P. 16 (Satisfacción con la metodología), P. 13 (Facilita más la exploración de los datos) y P. 11 (Interpretación rápida) recibieron 12 valoraciones 5 (75.00%). En este grupo, el 90.62% de las respuestas están en las categorías 4 y 5, indicando una percepción muy positiva de los ítems evaluados.

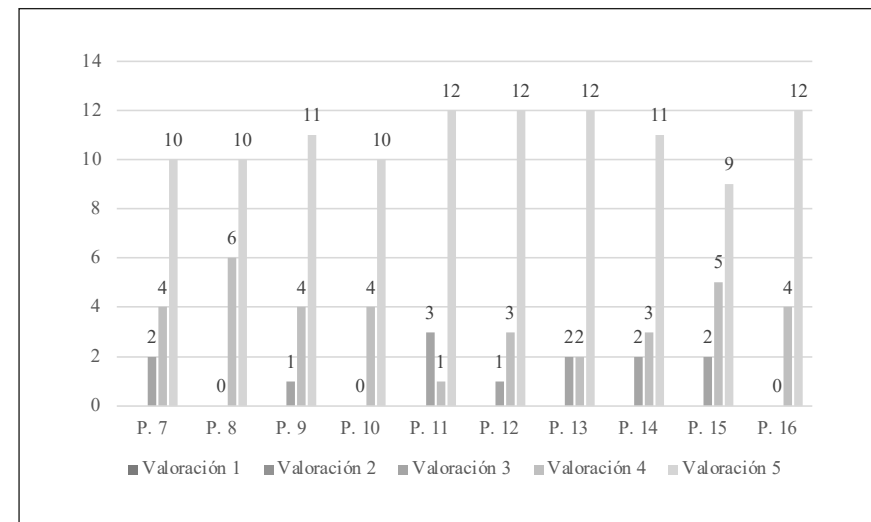
Tabla 86

Frecuencia de valores a cada pregunta de la encuesta de utilidad de TU® en el grupo de edad 54 a 65 años (N=16).

Pregunta (N.º)	Valoración 1 (%)	Valoración 2 (%)	Valoración 3 (%)	Valoración 4 (%)	Valoración 5 (%)
P. 7	-	-	2 (12.50)	4 (25.00)	10 (62.50)
P. 8	-	-	-	6 (37.50)	10 (62.50)
P. 9	-	-	1 (6.25)	4 (25.00)	11 (68.75)
P. 10	-	-	-	4 (25.00)	10 (62.50)
P. 11	-	-	3 (18.75)	1 (6.25)	12 (75.00)
P. 12	-	-	1 (6.25)	3 (18.75)	12 (75.00)
P. 13	-	-	2 (12.50)	2 (12.50)	12 (75.00)
P. 14	-	-	2 (12.50)	3 (18.75)	11 (68.75)
P. 15	-	-	2 (12.50)	5 (31.25)	9 (56.25)
P. 16	-	-	-	4 (25.00)	12 (75.00)
Total	0 (0.00)	0 (0.00)	13 (8.12)	36 (22.50)	109 (68.12)

Figura 52

Frecuencia de los valores a cada ítem de la encuesta de utilidad de TU® en el grupo de edad 55 a 65 años (N=11).



En conclusión, los datos reflejan una fuerte y uniforme percepción de la utilidad de TU®, con valoraciones predominantes en las categorías más altas en todos los grupos de edad, y con una variabilidad mínima en las respuestas. Estos hallazgos subrayan la efectividad y aceptación de la colaboración en la comprensión y toma de decisiones basada en datos.

ANEXO 10

Subanálisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU® por categorías participantes jóvenes (< 44 años) y participantes mayores (≥ 44 años).

Subanálisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU® por categorías participantes jóvenes (< 44 años) y participantes mayores (≥ 44 años).

Analizamos a continuación las diferencias en la frecuencia de valor a cada ítem de la encuesta diferenciando entre los profesionales jóvenes frente a los mayores (Tabla 87) (Figura 53) (Tabla 88) (Figura 54).

Al analizar las diferencias en las frecuencias de valoración de cada ítem de la encuesta sobre la utilidad de TU® entre los profesionales jóvenes (menores de 44 años) y los mayores (mayores de 44 años), se observan algunas diferencias notables en la distribución de las valoraciones:

- En el ítem P.7, los profesionales mayores calificaron más positivamente este ítem, con un 66.66% otorgando una valoración de 5, mientras que los jóvenes mostraron un 42.55% en esta categoría. Las valoraciones combinadas de 4 y 5 suman un 92.58% para los mayores y un 81.91% para los jóvenes. Las valoraciones bajas (1 y 2) fueron inexistentes entre los mayores y sumaron solo un 3.19% entre los jóvenes. La diferencia en las valoraciones no resultó estadísticamente significativa (p=0.149).
- El ítem P.8, el 66.66% de los mayores dieron una valoración de 5, en comparación con el 47.87% de los jóvenes. Las valoraciones combinadas de 4 y 5 alcanzaron el 100% en los mayores y el 87.23% en los jóvenes. Las valoraciones bajas fueron mínimas, con un 1.06% en los jóvenes y ninguna en los mayores. Esta diferencia no fue estadísticamente significativa (p=0.280).
- El ítem P.9, el 70.37% de los mayores dieron una valoración de 5, mientras que el 42.55% de los jóvenes dieron la misma puntuación. Sumando las valoraciones de 4 y 5, los mayores alcanzaron el 96.29% frente al 85.10% de los jóvenes. Las valoraciones bajas fueron de nuevo mínimas, con solo un 1.06% entre los jóvenes. La diferencia no fue significativa (p=0.134).
- El ítem P.10, el 66.66% de los mayores valoraron este ítem con un 5, comparado con el 46.80% de los jóvenes. Las valoraciones combinadas de 4 y 5 alcanzaron el 100% en los mayores y el 86.16% en los jóvenes. Las valoraciones bajas fueron mínimas, con un 1.06% entre los jóvenes y ninguna entre los mayores. Esta diferencia no alcanzó significación estadística (p=0.235).
- En el ítem P.11, este ítem mostró una diferencia significativa (p=0.043), con el 74.07% de los mayores otorgando una valoración de 5, frente al 48.93% de los jóvenes. La suma de las valoraciones de 4 y 5, en los mayores alcanzaron el 85.18% en comparación con el 86.16% de los jóvenes. Las valoraciones bajas fueron inexistentes entre los mayores y muy bajas entre los jóvenes (3.19%).
- El ítem P.12, el 74.07% de los mayores dieron una valoración de 5, mientras que el 50.00% de los jóvenes hicieron lo mismo. Las valoraciones combinadas de 4 y 5 suman un 96.29% para los mayores y un 82.97% para los jóvenes. Las valoraciones bajas

fueron mínimas, con solo un 3.19% entre los jóvenes y ninguna entre los mayores. La diferencia no fue significativa (p=0.127).

- En el ítem P.13, mostró una diferencia significativa (p=0.028), con el 77.77% de los mayores valorándolo con un 5, en comparación con el 45.74% de los jóvenes. Las valoraciones combinadas de 4 y 5 suman un 92.58% para los mayores y un 86.16% para los jóvenes. Las valoraciones bajas fueron mínimas en ambos grupos.
- El ítem P.14, el 74.07% de los mayores dieron una valoración de 5, frente al 54.25% de los jóvenes. Las valoraciones combinadas de 4 y 5 alcanzaron el 92.58% en los mayores y el 78.71% en los jóvenes. Las valoraciones bajas fueron mínimas, con solo un 4.24% entre los jóvenes y ninguna entre los mayores. La diferencia no fue significativa (p=0.383).
- El ítem P.15, el 59.25% de los mayores dieron una valoración de 5, en comparación con el 54.25% de los jóvenes. Las valoraciones combinadas de 4 y 5 suman un 92.58% para los mayores y un 77.65% para los jóvenes. Las valoraciones bajas fueron mínimas, con solo un 5.31% entre los jóvenes y ninguna entre los mayores. La diferencia no fue significativa (p=0.457).
- En el ítem P.16, mostró una diferencia significativa (p=0.036), con el 81.48% de los mayores otorgando una valoración de 5, frente al 52.12% de los jóvenes. Las valoraciones combinadas de 4 y 5 suman un 100% para los mayores y un 85.10% para los jóvenes. Las valoraciones bajas fueron inexistentes en ambos grupos.

En general, los profesionales mayores tienden a valorar más positivamente los ítems de la metodología TU® en comparación con los jóvenes. Las diferencias en las valoraciones de los ítems relacionados con la rapidez en la interpretación de datos (P. 11), la facilidad para explorar los datos (P. 13) y la satisfacción general P. 16) con la metodología resultaron estadísticamente significativas. Además, las valoraciones bajas (1 y 2) fueron muy poco frecuentes en ambos grupos, lo que indica una percepción generalmente positiva de la metodología. Esto sugiere una mayor apreciación de estos aspectos por parte de los mayores, posiblemente influenciada por su experiencia y familiaridad con metodologías de análisis de datos.

Tabla 87
Frecuencia de valores a cada pregunta de la encuesta en el grupo de edad MAYORES (N=27).

Pregunta (N.º)	Valoración 1 (%)	Valoración 2 (%)	Valoración 3 (%)	Valoración 4 (%)	Valoración 5 (%)	Valor p
P. 7	-	-	2 (7.40)	7 (25.92)	18 (66.66)	0.149
P. 8	-	-	-	9 (33.33)	18 (66.66)	0.280
P. 9	-	-	1 (3.70)	7 (25.92)	19 (70.37)	0.134
P. 10	-	-	-	9 (33.33)	18 (66.66)	0.235
P. 11	-	-	4 (14.81)	3 (11.11)	20 (74.07)	0.043

P. 12	-	-	1 (3.70)	6 (22.22)	20 (74.07)	0.127
P. 13	-	-	2 (7.40)	4 (14.81)	21 (77.77)	0.028
P. 14	-	-	2 (7.40)	5 (18.51)	20 (74.07)	0.383
P. 15	-	-	2 (7.40)	9 (33.33)	16 (59.25)	0.457
P. 16	-	-	-	5 (18.51)	22 (81.48)	0.036
Total	0 (0.00)	0 (0.00)	14 (5.18)	64 (23,70)	192 (71.11)	

Figura 53

Frecuencia de valores a cada pregunta de la encuesta en el grupo de edad MAYORES

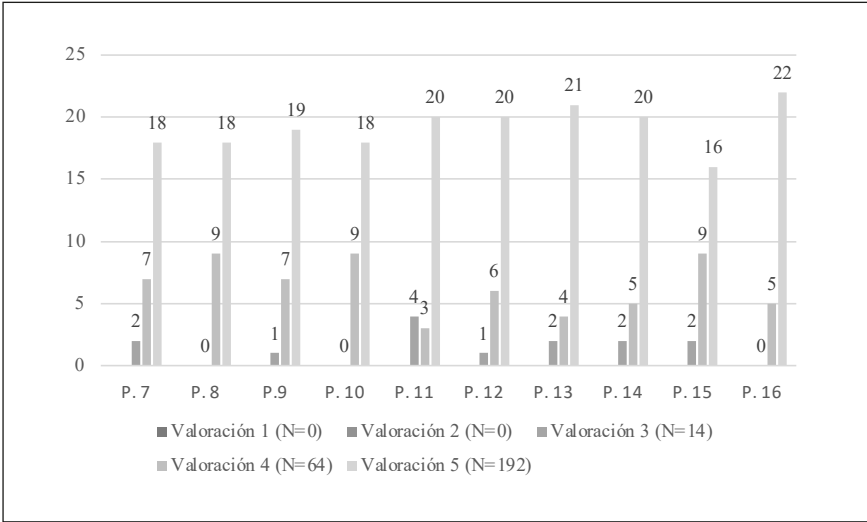


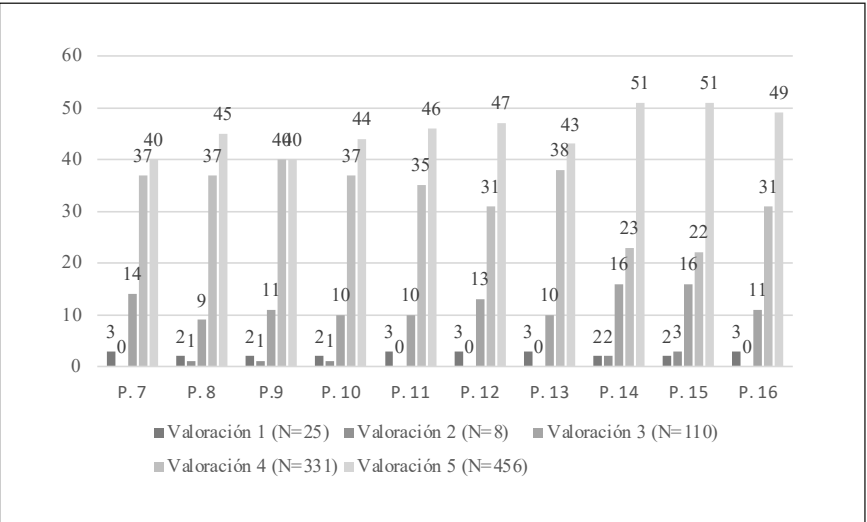
Tabla 88

Frecuencia de valores a cada pregunta de la encuesta en el grupo de edad MENORES (N=94).

Pregunta (N.º)	Valoración 1 (%)	Valoración 2 (%)	Valoración 3 (%)	Valoración 4 (%)	Valoración 5 (%)	Valor p
P. 7	3 (3.19)	-	14 (14.89)	37 (39.36)	40 (42.55)	0.149
P. 8	2 (2.12)	1 (1.06)	9 (9.57)	37 (39.36)	45 (47.87)	0.280
P. 9	2 (2.12)	1 (1.06)	11 (11.70)	40 (42.55)	40 (42.55)	0.134
P. 10	2 (2.12)	1 (1.06)	10 (10.63)	37 (39.36)	44 (46.80)	0.235
P. 11	3 (3.19)	-	10 (10.63)	35 (37.23)	46 (48.93)	0.043
P. 12	3 (3.19)	-	13 (13.82)	31 (32.97)	47 (50.00)	0.127
P. 13	3 (3.19)	-	10 (10.63)	38 (40.42)	43 (45.74)	0.028
P. 14	2 (2.12)	2 (2.12)	16 (17.02)	23 (24.46)	51 (54.25)	0.383
P. 15	2 (2.12)	3 (3.19)	16 (17.02)	22 (23.40)	51 (54.25)	0.457
P. 16	3 (3.19)	-	11 (11.70)	31 (32.97)	49 (52.12)	0.036
Total	25 (2.65)	8 (0.85)	120 (12.76)	331 (35.21)	456 (48.51)	

Figura 54

Frecuencia de valores a cada pregunta de la encuesta en el grupo de edad MENORES (N=94).



ANEXO 11

Análisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU® por Hospitales.

Análisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU® por Hospitales.

El análisis de la frecuencia de valoración por ítem en la encuesta de utilidad de TU®, que incluye a 121 encuestados y un total de 1210 valoraciones, revela que las valoraciones más altas (5) son predominantes en todos los ítems.

En el Hospital Universitario Puerta de Hierro-Majadahonda (N=13) (Tabla 89) (Figura 55), las valoraciones más altas (5) predominaron en los ítems P. 13 (Facilidad para explorar datos) y P. 12 (Utilidad para tomar decisiones), con 7 respuestas (53.84%) cada una. Los ítems con menor frecuencia de valoraciones 5 fueron P. 7 (Facilidad comprensión) y P. 8 (Bubble plots), cada una con 4 respuestas (30.76%). Las valoraciones 5 y 4 supusieron el 43.07% y 36.92% de las valoraciones emitidas respectivamente, mientras que las valoraciones más bajas (1 y 2) fueron mínimas en todas las preguntas, suponiendo las valoraciones 1 el 4.61% y las valoraciones 2 el 3.07% de las 130 valoraciones emitidas.

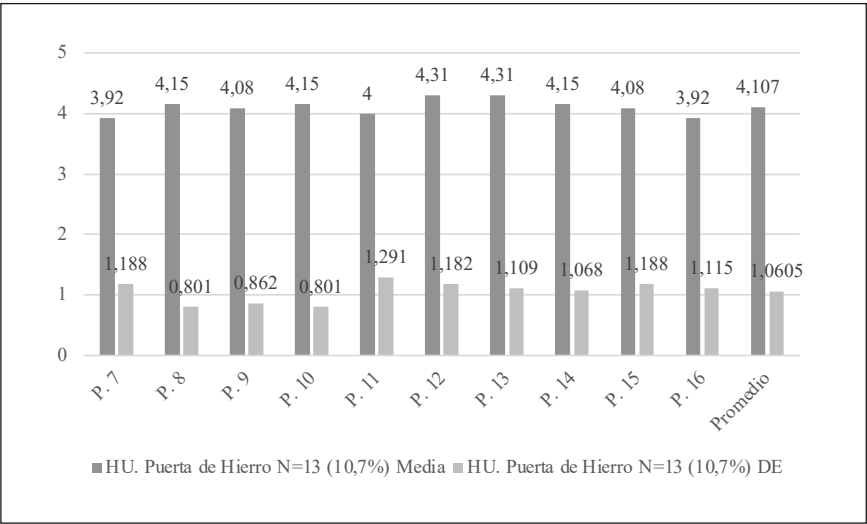
Tabla 89

Frecuencia de valores a cada ítem de la encuesta en HU Puerta de Hierro-Majadahonda (N=13).

Pregunta (N.º)	Valoración 1 (%)	Valoración 2 (%)	Valoración 3 (%)	Valoración 4 (%)	Valoración 5 (%)	Valor p
P. 7	1 (7.69)	-	3 (23.07)	4 (30.76)	5 (38.46)	0.277
P. 8	-	1 (7.69)	-	8 (61.53)	4 (30.76)	0.275
P. 9	-	1 (7.69)	1 (7.69)	7 (53.84)	4 (30.76)	0.112
P. 10	-	1 (7.69)	-	8 (61.53)	4 (30.76)	0.077
P. 11	1 (7.69)	-	4 (30.76)	1 (7.69)	7 (53.84)	0.086
P. 12	1 (7.69)	-	1 (7.69)	3 (23.07)	8 (61.53)	0.394
P. 13	1 (7.69)	-	-	5 (38.46)	7 (53.84)	0.645
P. 14	-	1 (7.69)	3 (23.07)	2 (15.38)	7 (53.84)	0.822
P. 15	1 (7.69)	-	2 (15.38)	4 (30.76)	6 (46.15)	0.913
P. 16	1 (7.69)	-	2 (15.38)	6 (46.15)	4 (30.76)	0.559
Total	6 (4.61)	4 (3.07)	16 (12.30)	48 (36.92)	56 (43.07)	

Figura 55

Frecuencia de valores en cada ítem en el HU Puerta de Hierro-Majadahonda.



En el Hospital Universitario Clínico San Carlos (N=25) (Tabla 90) (Figura 56), los ítems P. 16 (Satisfacción) y P. 15 (Menos conocimientos) recibieron la mayor cantidad de valoraciones 5, con 18 respuestas (60.00%) cada una. Los ítems con menor frecuencia de valoraciones 5 fueron P. 10 (Acceso a datos) y P. 8 (Bubble plots), con 10 respuestas (40.00%) y 14 respuestas (56.00%) respectivamente. Las valoraciones 5 y 4 supusieron el 55.20% y 34,40% de las valoraciones emitidas respectivamente, mientras que las valoraciones más bajas (1 y 2) fueron mínimas en todas las preguntas, suponiendo las valoraciones 1 el 4.00% y las valoraciones 2 el 0.40% de las 250 valoraciones emitidas.

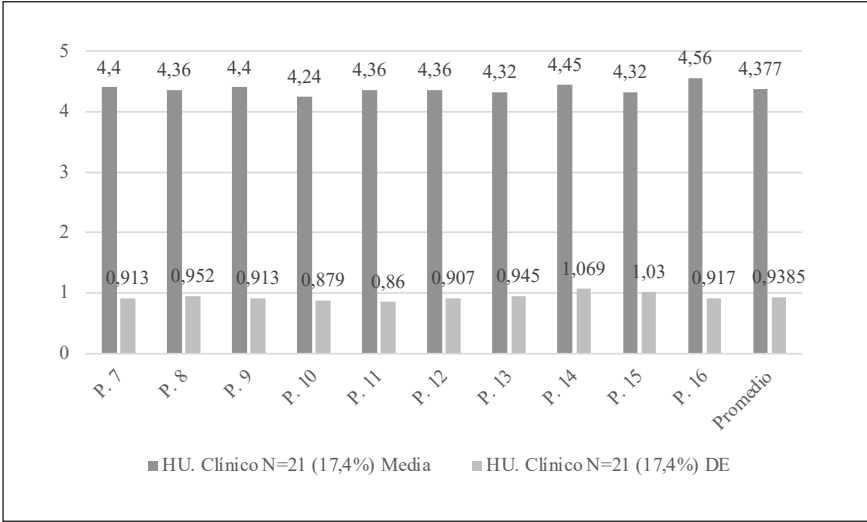
Tabla 90

Frecuencia de valores a cada pregunta de la encuesta en HU Clínico San Carlos (N=25)..

Pregunta (N.º)	Valoración 1 (%)	Valoración 2 (%)	Valoración 3 (%)	Valoración 4 (%)	Valoración 5 (%)	Valor p
P. 7	1 (4.00)	-	1 (4.00)	9 (36.00)	14 (56.00)	0.277
P. 8	1 (4.00)	-	2 (8.00)	8 (32.00)	14 (56.00)	0.275
P. 9	1 (4.00)	-	1 (4.00)	9 (36.00)	14 (56.00)	0.112
P. 10	1 (4.00)	-	1 (4.00)	13 (52.00)	10 (40.00)	0.077
P. 11	1 (4.00)	-	-	12 (48.00)	12 (48.00)	0.086
P. 12	1 (4.00)	-	1 (4.00)	10 (40.00)	13 (52.00)	0.394
P. 13	1 (4.00)	-	2 (8.00)	9 (36.00)	13 (52.00)	0.645
P. 14	1 (4.00)	1 (4.00)	2 (8.00)	6 (24.00)	15 (60.00)	0.822
P. 15	1 (4.00)	-	4 (16.00)	5 (20.00)	15 (60.00)	0.913
P. 16	1 (4.00)	-	1 (4.00)	5 (20.00)	18 (72.00)	0.559
Total	10 (4.00)	1 (0.40)	15 (6.00)	86 (34.40)	138 (55.20)	

Figura 56

Frecuencia de valores en cada ítem en el H U Clínico San Carlos (N=25).



En el Hospital Universitario La Paz (N=21) (**Tabla 91**) (**Figura 57**), las valoraciones más altas (5) fueron predominantes en el ítem P. 11 (Interpretación rápida), con 13 respuestas (61.90%), seguida de cerca por los ítems P. 8 (Bubble plots) y P. 14 (Evolución histórica), con 12 respuestas (57.14%) cada una. Los ítems con menor frecuencia de valoraciones 5 fueron P. 7 (Facilita comprensión) y P. 15 (Menos conocimientos), con 9 respuestas (42.85%) y 11 respuestas (52.38%) respectivamente. Las valoraciones 5 y 4 supusieron el 54,28% y 29.52% de las valoraciones emitidas respectivamente, mientras que las valoraciones más bajas (1 y 2) fueron mínimas en todas las preguntas, suponiendo las valoraciones 1 el 4.28% y las valoraciones 2 el 0.47% de las 210 valoraciones emitidas.

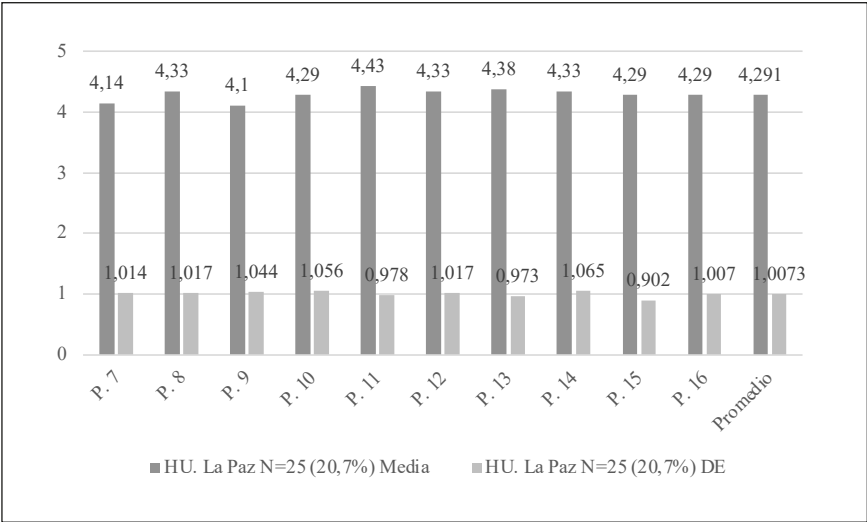
Tabla 91

Frecuencia de valores a cada ítem de la encuesta en HU La Paz (N=21).

Pregunta (N.º)	Valoración 1 (%)	Valoración 2 (%)	Valoración 3 (%)	Valoración 4 (%)	Valoración 5 (%)	Valor p
P. 7	1 (4.76)	-	3 (14.28)	8 (38.09)	9 (42.85)	0.277
P. 8	1 (4.76)	-	2 (9.52)	6 (28.57)	12 (57.14)	0.275
P. 9	1 (4.76)	-	4 (19.04)	7 (33.33)	9(42.85)	0.112
P. 10	1 (4.76)	-	3 (14.28)	5 (23.80)	12 (57.14)	0.077
P. 11	1 (4.76)	-	1 (4.76)	6 (28.57)	13 (61.90)	0.086
P. 12	1 (4.76)	-	2 (9.52)	6 (28.57)	12 (57.14)	0.394
P. 13	1 (4.76)	-	1 (4.76)	7 (33.33)	12 (57.14)	0.645
P. 14	1 (4.76)	-	3 (14.28)	4 (19.04)	13 (61.90)	0.822
P. 15	-	1 (4.76)	3 (14.28)	6 (28.57)	11 (52.38)	0.913
P. 16	1 (4.76)	-	2 (9.52)	7 (33.33)	11 (52.38)	0.559
Total	9 (4.28)	1 (0.47)	24 (11.42)	62 (29.52)	114 (54.28)	

Figura 57

Frecuencia de valores en cada ítem en el H U La Paz (N=21).



En el Hospital Universitario Ramón y Cajal (N=22), (**Tabla 92**) (**Figura 58**), los ítems P. 14 (Evolución histórica) y P. 15 (Menos conocimientos) recibieron la mayor cantidad de valoraciones 5, con 13 respuestas (59.09%) cada una. Los ítems con menor frecuencia de valoraciones 5 fueron P. 10 (Acceso a datos) y P. 11 (Interpretación rápida), con 11 respuestas (50.00%) cada una. Las valoraciones 5 y 4 supusieron el 45.90% y 43.18% de las valoraciones emitidas respectivamente, mientras que las valoraciones más bajas (1 y 2) fueron mínimas en todas las preguntas, suponiendo ambas el 0,00% de las 220 valoraciones emitidas.

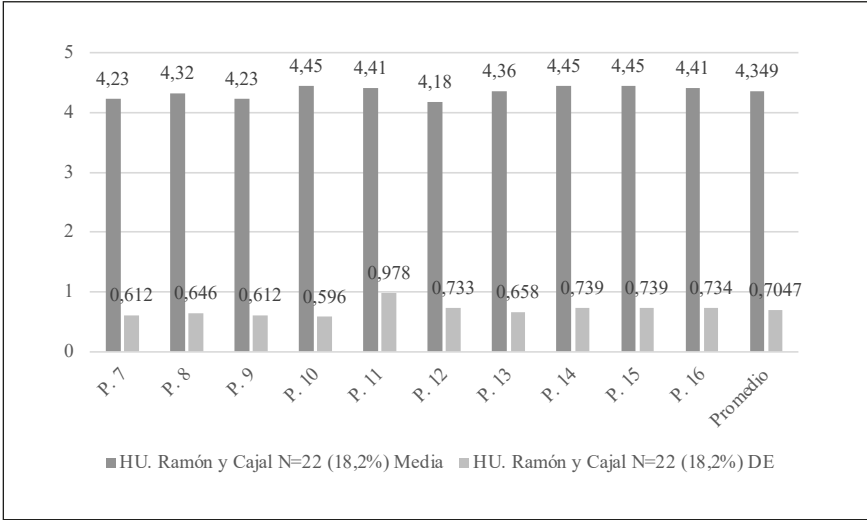
Tabla 92

Frecuencia de valores a cada ítem de la encuesta en HU Ramón y Cajal (N=22)..

Pregunta (N.º)	Valoración 1 (%)	Valoración 2 (%)	Valoración 3 (%)	Valoración 4 (%)	Valoración 5 (%)	Valor p
P. 7	-	-	2 (9.09)	13 (59.09)	7 (31.81)	0.277
P. 8	-	-	2 (9.09)	11 (50.00)	9(40.90)	0.275
P. 9	-	-	2 (9.09)	13 (59.09)	7 (31.81)	0.112
P. 10	-	-	1 (4.54)	10 (45.45)	11 (50.00)	0.077
P. 11	-	-	2 (9.09)	9 (40.90)	11 (50.00)	0.086
P. 12	-	-	4 (18.18)	10 (45.45)	8 (36.36)	0.394
P. 13	-	-	2 (9.09)	10 (45.45)	10 (45.45)	0.645
P. 14	-	-	3 (13.63)	6 (27.27)	13 (59.09)	0.822
P. 15	-	-	3 (13.63)	6 (27.27)	13 (59.09)	0.913
P. 16	-	-	3 (13.63)	7 (31.81)	12 (54.54)	0.559
Total	0 (0.00)	0 (0.00)	24 (10.90)	95 (43.18)	101 (45.90)	

Figura 58

Frecuencia de valores en cada ítem en el HU Ramón y Cajal (N=22).



En el Hospital Universitario Central de la Cruz Roja (N=40) (**Tabla 93**) (**Figura 59**) los ítems P. 12 (Utilidad para tomar decisiones) y P. 16 (Satisfacción) destacaron con la mayor cantidad de valoraciones 5, con 26 respuestas (65.00%) cada una. Los ítems con menor frecuencia de valoraciones 5 fueron P. 7 (Facilita comprensión) y P. 15 (Menos conocimientos), con 23 respuestas (57.50%) cada una. Las valoraciones 5 y 4 supusieron el 59,75% y 26.00% de las valoraciones emitidas respectivamente, mientras que las valoraciones más bajas (1 y 2) fueron mínimas en todas las preguntas, suponiendo las valoraciones 1 el 0.00% y las valoraciones 2 el 0.50% de las 400 valoraciones emitidas.

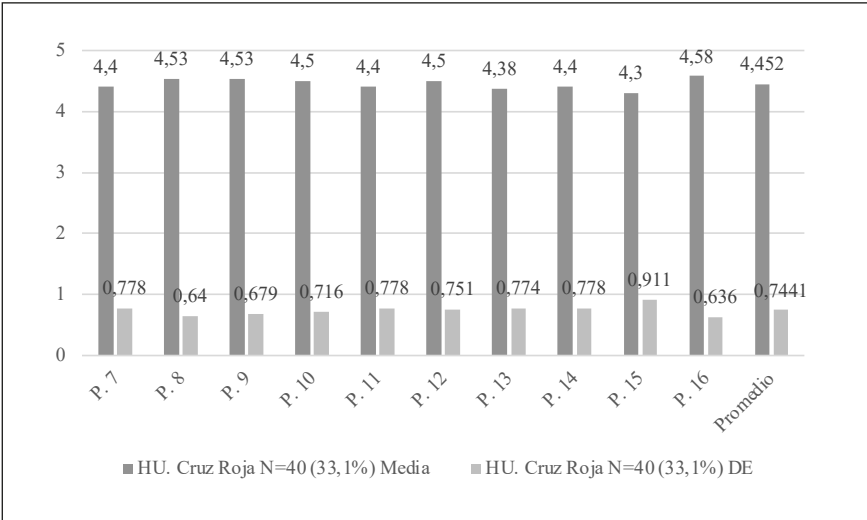
Tabla 93

Frecuencia de valores a cada ítem de la encuesta en HU Central de la Cruz Roja (N=40).

Pregunta (N.º)	Valoración 1 (%)	Valoración 2 (%)	Valoración 3 (%)	Valoración 4 (%)	Valoración 5 (%)	Valor p
P. 7	-	-	7 (17.50)	10 (25.00)	23 (57.50)	0.277
P. 8	-	-	3 (7.50)	13 (32.50)	24 (60.00)	0.275
P. 9	-	-	4 (10.00)	11 (27.50)	25 (62.50)	0.112
P. 10	-	-	5 (12.50)	10 (25.00)	25 (62.50)	0.077
P. 11	-	-	7 (17.50)	10 (25.00)	23 (57.50)	0.086
P. 12	-	-	6 (15.00)	8 (20.00)	26 (65.00)	0.394
P. 13	-	-	7 (17.50)	11 (27.50)	22 (55.00)	0.645
P. 14	-	-	7 (17.50)	10 (25.00)	23 (57.50)	0.822
P. 15	-	2 (5.00)	6 (15.00)	10 (25.00)	22 (55.00)	0.913
P. 16	-	-	3 (7.50)	11 (27.50)	26 (65.00)	0.559
Total	0 (0.00)	2 (0.50)	55 (13.75)	104 (26.00)	239 (59.75)	

Figura 59

Frecuencia de valores en cada ítem en el HU Cruz Roja (N=40).



ANEXO 12

Análisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU® entre Geriatras y no Geriatras.

Análisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU® entre Geriatras y no Geriatras.

Se observa que la mayoría de los ítems reciben valoraciones altas (4 o 5) tanto por parte de geriatras como de no-geriatras, presentando un porcentaje de valoraciones altas (entre 4 o 5) entre los geriatras de un 85.9%, y en los médicos no-geriatras de un 91.81%. De la misma manera, las frecuencias de valoraciones máximas (5) son discretamente más pronunciadas en los médicos no-geriatras (59.09%) en comparación con los geriatras (54.40%). No se han encontrado diferencias estadísticamente significativas entre la valoración de los diferentes ítems entre los dos grupos, Geriatras y no Geriatras (Tabla 94).

En el ítem P. 9 (Representación gráfica), el 72.72% de los no-geriatras asignaron una valoración de 5 representando el valor más alto en este grupo, mientras que, entre los geriatras, este porcentaje fue del 47.30%, representando el valor más bajo en este grupo, aun sin representar una diferencia estadísticamente significativa (p=0.594).

En tres ítems, P. 8 (Bubble plots), P. 10 (Acceso a datos) y P. 11 (Interpretación rápida) la categoría de geriatras valoró con mayor frecuencia con la puntuación máxima de 5 comparado con los no-geriatras; 55.91% / 54.54% (P. 8); 53.76% / 45.45% (P. 10); 54.83% / 54.54% (P.11). Igualmente, las diferencias no observan significación estadística según la prueba de la chi-cuadrada, con valores de p de 0.981, 0.593 y 0.542 respectivamente. En los restantes ítems lo médicos no-geriatras valoraron con mayor frecuencia con la puntuación máxima comparado con los médicos geriatras, pero sin ser una diferencia estadísticamente significativa.

Tabla 94

Análisis de las diferencias en las frecuencias de valoración de cada ítem entre Geriatras (N=93) y no Geriatras (N=11).

Preguntas %	Valoración 1 (%)		Valoración 2 (%)		Valoración 3 (%)		Valoración 4 (%)		Valoración 5 (%)		Valor-p
	Geriatra	No Geriatra	Geriatra	No Geriatra	Geriatra	No Geriatra	Geriatra	No Geriatra	Geriatra	No Geriatra	
P. 7	3 (3.22)	-	-	-	11 (11.82)	-	35 (37.63)	5 (45.45)	44 (47.31)	6 (54.54)	0,59
P. 8	2 (2.15)	-	1 (1.07)	-	7 (7.52)	1 (9.09)	31 (33.33)	4 (36.36)	52 (55.91)	6 (54.54)	0,98
P. 9	2 (2.15)	-	1 (1.07)	-	10 (10.75)	1 (9.09)	36 (38.70)	2 (18.18)	44 (47.31)	8 (72.72)	0,59
P. 10	2 (2.15)	-	1 (1.07)	-	9 (9.67)	-	31 (33.33)	6 (54.54)	50 (53.76)	5 (45.45)	0,59

P. 11	3 (3.22)	-	-	-	10 (10.75)	-	29 (31.18)	5 (45.45)	51 (54.83)	6 (54.54)	0.54
P. 12	3 (3.22)	-	-	-	10 (10.75)	1 (9.09)	29 (31.18)	3 (27.27)	51 (54.83)	7 (63.63)	0.9
P. 13	3 (3.22)	-	-	-	8 (8.60)	2 (18.18)	33 (35.48)	3 (27.27)	49 (52.68)	6 (54.54)	0.68
P. 14	2 (2.15)	-	1 (1.07)	-	14 (15.05)	1 (9.09)	20 (21.50)	3 (27.27)	56 (60.21)	7 (63.63)	0.94
P. 15	2 (2.15)	-	3 (3.22)	-	12 (12.90)	2 (18.18)	23 (24.73)	2 (18.18)	53 (56.98)	7 (63.63)	0.9
P. 16	3 (3.22)	-	-	-	8(8.60)	1 (9.09)	26 (27.95)	3 (27.27)	56 (60.21)	7 (63.63)	0.95
Total (%)	25 (2.68)	0 (0.00)	7 (0.75)	0 (0,00)	99 (10.64)	9 (8.18)	293 (31.50)	36 (32.72)	506 (54.40)	65 (59.09)	

ANEXO 13

Análisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU® en el grupo de Geriatras.

Análisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU® en el grupo de Geriatras.

Se muestran los datos de respuesta a la encuesta geriatras (N=93) (Tabla 95) (Figura 60).

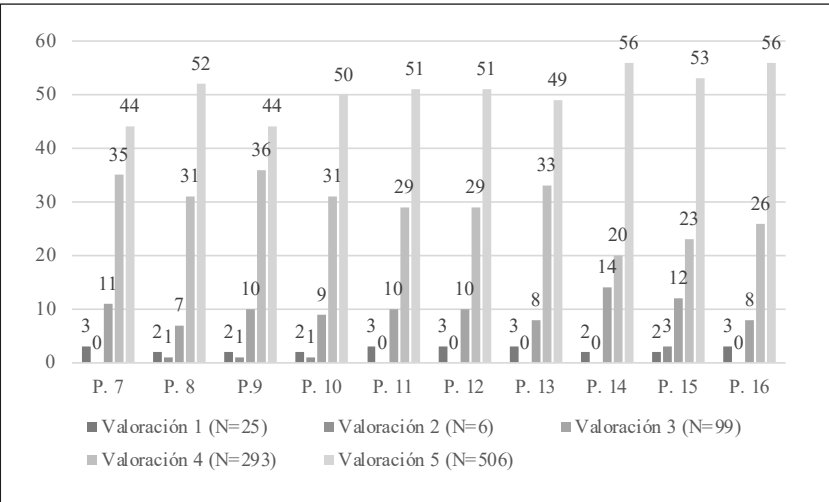
Tabla 95

Frecuencia de valores a cada pregunta de la encuesta en los Geriatras (N=93).

Pregunta (N.º)	Valoración 1 (%)	Valoración 2 (%)	Valoración 3 (%)	Valoración 4 (%)	Valoración 5 (%)
P. 7	3 (3.22)	-	11 (11.82)	35 (37.63)	44 (47.31)
P. 8	2 (2.15)	1 (1.07)	7 (7.52)	31 (33.33)	52 (55.91)
P. 9	2 (2.15)	1 (1.07)	10 (10.75)	36 (38.70)	44 (47.31)
P. 10	2 (2.15)	1 (1.07)	9 (9.67)	31 (33.33)	50 (53.76)
P. 11	3 (3.22)	-	10 (10.75)	29 (31.18)	51 (54.83)
P. 12	3 (3.22)	-	10 (10.75)	29 (31.18)	51 (54.83)
P. 13	3 (3.22)	-	8 (8.60)	33 (35.48)	49 (52.68)
P. 14	2 (2.15)	1 (1.07)	14 (15.05)	20 (21.50)	56 (60.21)
P. 15	2 (2.15)	3 (3.22)	12 (12.90)	23 (24.73)	53 (56.98)
P. 16	3 (3.22)	-	8(8.60)	26 (27.95)	56 (60.21)
Total	25 (2.68)	7 (0.75)	99 (10.64)	293 (31.50)	506 (54.40)

Figura 60

Frecuencia de valores a cada ítem de la encuesta en los Geriatras (N=93).



El ítem P. 7, relativa a la facilidad de comprensión, muestra que la mayoría de los geriatras otorgaron valoraciones altas, con 44 respuestas (47.31%) para la valoración 5 y 35 respuestas (37.63%) para la valoración 4. Las valoraciones bajas fueron mínimas, con solo 3 respuestas (3.22%) para la valoración 1 y ninguna para la valoración 2.

Para el ítem P. 8, referente a los "Bubble plots", las valoraciones 5 y 4 fueron las más frecuentes, con 52 (55.91%) y 31 (33.33%) respuestas respectivamente. Las valoraciones 3 y menores fueron menos comunes, con 7 respuestas (7.52%) para la valoración 3, 1 respuesta (1.07%) para la valoración 2 y 2 respuestas (2.15%) para la valoración 1.

El ítem P. 9, sobre la representación gráfica en general, también obtuvo valoraciones altas, con 44 respuestas (47.31%) para la valoración 5 y 36 respuestas (38.70%) para la valoración 4. Las valoraciones bajas fueron escasas, con 2 respuestas (2.15%) para la valoración 1 y 1 respuesta (1.07%) para la valoración 2.

En cuanto al ítem P. 10, que evalúa el acceso a los datos, las valoraciones 5 y 4 fueron predominantes, con 50 respuestas (53.76%) y 31 respuestas (33.33%) respectivamente. Las valoraciones más bajas fueron limitadas, con 2 respuestas (2.15%) para la valoración 1 y 1 respuesta (1.07%) para la valoración 2.

El ítem P. 11, que trata sobre la interpretación rápida, recibió 51 respuestas (54.83%) para la valoración 5 y 29 respuestas (31.18%) para la valoración 4, mientras que las valoraciones bajas fueron mínimas, con 3 respuestas (3.22%) para la valoración 1 y ninguna para la valoración 2.

El ítem P. 12, relacionada con la utilidad para la toma de decisiones, mostró un patrón comparable con 51 respuestas (54.83%) para la valoración 5 y 29 respuestas (31.18%) para la valoración 4.

Para el ítem P. 13, sobre la facilidad para explorar datos, se registraron 49 respuestas (52,68%) para la valoración 5 y 33 respuestas (35.48%) para la valoración 4, con solo 3 respuestas (3.22%) para la valoración 1 y ninguna para la valoración 2.

El ítem P. 14, referente a la evolución histórica, destacó con la mayor cantidad de valoraciones 5, acumulando 56 respuestas (60.21%), y 20 respuestas (21.50%) para la valoración 4. Las valoraciones bajas fueron muy bajas, con 2 respuestas (2.15%) para la valoración 1 y 1 respuesta (1.07%) para la valoración 2.

El ítem P. 15, que evalúa la utilidad de la metodología para quienes tienen menos conocimientos previos, 53 respuestas (56.98%) fueron para la valoración 5 y 23 respuestas (24.73%) para la valoración 4. Las valoraciones bajas fueron limitadas, con 2 respuestas (2.15%) para la valoración 1 y 3 respuestas (3.22%) para la valoración 2.

El ítem P. 16, relativa a la satisfacción general, obtuvo la mayor cantidad de valoraciones 5 junto con P. 14, con 56 respuestas (60.21%), y 26 respuestas (27.95%) para la valoración 4. Las valoraciones bajas fueron mínimas, con 3 respuestas (3.22%) para la valoración 1 y ninguna para la valoración 2.

En total, las valoraciones 5 sumaron 506 respuestas (54,40%), las valoraciones 4 alcanzaron 293 respuestas (31.50%), y las valoraciones 3, 2 y 1 fueron menores, con 99 (10.64%), 7 (0.75%) y 25 (2.8%) respuestas respectivamente. Estos resultados indican una clara tendencia hacia las valoraciones altas de los distintos ítems de la encuesta sobre la utilidad de la herramienta TU®, subrayando una percepción muy positiva y una satisfacción general con la metodología TU® entre los geriatras encuestados.

ANEXO 14

Análisis descriptivo de la frecuencia de valoración de valoración por ítem en la encuesta de utilidad de TU® por grupos de edad de los Geriatras.

Análisis descriptivo de la frecuencia de valoración de valoración por ítem en la encuesta de utilidad de TU® por grupos de edad de los Geriatras.

En el grupo geriatras de edad menor de 33 años (N=57) (Tabla 96) (Figura 61) las valoraciones más altas (5) fueron predominantes en todos los ítems. El ítem P. 16 (Satisfacción) obtuvo la mayor frecuencia de valoraciones 5, con 35 respuestas (61.40%), seguida por los ítems P. 14 (Evolución histórica) y P. 12 (Utilidad para tomar decisiones), cada una con 35 respuestas (61.40%). Las valoraciones 4 también fueron comunes, con la mayor frecuencia en el ítem P. 10 (Acceso a datos) con 16 respuestas (28,07%). Las valoraciones más bajas (1 y 2) fueron mínimas representando el 4.48% y el 1.22% en este grupo etario de geriatras.

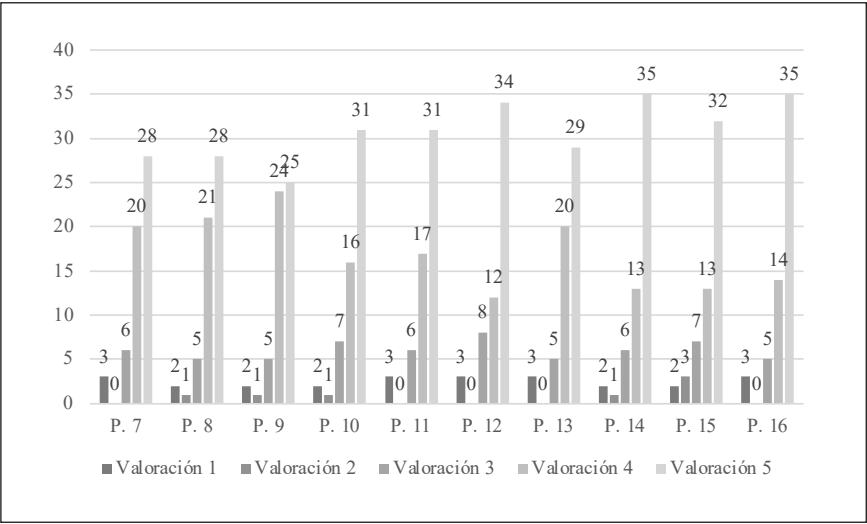
Tabla 96

Frecuencia de valoración de cada ítem en grupo de edad menor de 33 años (N=57).

Pregunta (N.º)	Valoración 1 (%)	Valoración 2 (%)	Valoración 3 (%)	Valoración 4 (%)	Valoración 5 (%)
P. 7	3 (5.26)	-	6 (10.52)	20 (35.08)	28 (49.12)
P. 8	2 (3.50)	1 (1.75)	5 (8.77)	21 (36.84)	28 (49.12)
P. 9	2 (3.50)	1 (1.75)	5 (8.77)	24 (42.10)	25 (43.85)
P. 10	2 (3.50)	1 (1.75)	7 (12.28)	16 (28.07)	31 (54.38)
P. 11	3 (5.26)	-	6 (10.52)	17 (29.82)	31 (54.38)
P. 12	3 (5.26)	-	8 (14.03)	12 (21.05)	34 (59.64)
P. 13	3 (5.26)	-	5 (8.77)	20 (35.08)	29 (50.87)
P. 14	2 (3.50)	1 (1.75)	6 (10.52)	13 (22.80)	35 (61.40)
P. 15	2 (3.50)	3 (5.26)	7 (12.28)	13 (22.80)	32 (56.14)
P. 16	3 (5.26)	-	5 (8.77)	14 (24.56)	35 (61.40)
Total	25 (4.38)	7 (1.22)	60 (10.52)	170 (29.82)	308 (54.03)

Figura 61

Frecuencias de valoración por ítem en geriatras < 33 años (N=57).



En el grupo geriatras de edad entre 33 a 44 años (N=22) (Tabla 97) (Figura 62), las valoraciones 5 fueron frecuentes, especialmente en los ítems P. 8 (Bubble plots) y P. 15 (Menos conocimientos), con 12 respuestas (54.54%) cada una. Las valoraciones 4 tuvieron una alta frecuencia en los ítems P. 12 (Utilidad para tomar decisiones) y P. 13 (Facilidad para explorar datos), con 14 (63.63%) y 12 respuestas (54.54%) respectivamente. Las valoraciones bajas (1 y 2) estuvieron ausentes en este grupo.

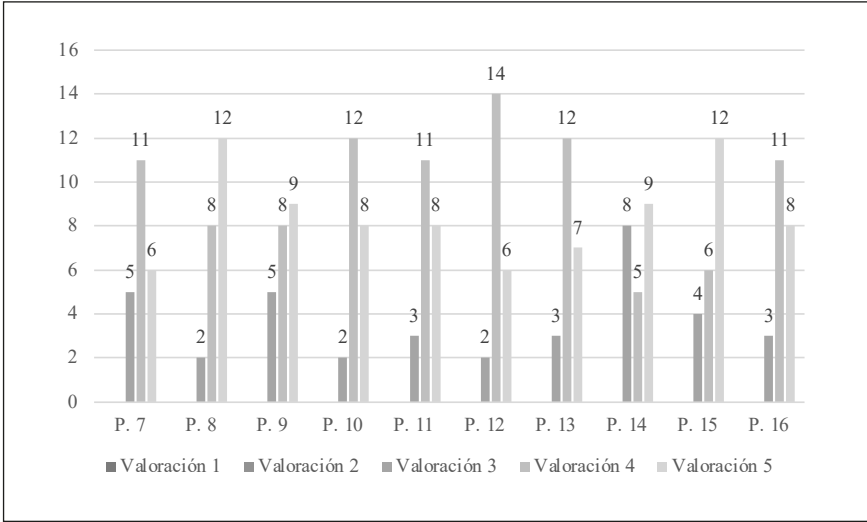
Tabla 97

Frecuencia de valoración de cada ítem en el grupo de edad 33 a 44 años (N=22).

Pregunta (N.º)	Valoración 1 (%)	Valoración 2 (%)	Valoración 3 (%)	Valoración 4 (%)	Valoración 5 (%)
P. 7	-	-	5 (22.72)	11 (50.00)	6 (27.27)
P. 8	-	-	2 (9.09)	8 (36.36)	12 (54.54)
P. 9	-	-	5 (22.72)	8 (36.36)	9 (40.90)
P. 10	-	-	2 (9.09)	12 (54.54)	8 (36.36)
P. 11	-	-	3 (13.63)	11 (50.00)	8 (36.36)
P. 12	-	-	2 (9.09)	14 (63.63)	6 (27.27)
P. 13	-	-	3 (13.63)	12 (54.54)	7 (31.81)
P. 14	-	-	8 (36.36)	5 (22.72)	9 (40.90)
P. 15	-	-	4 (18.18)	6 (27.27)	12 (54.54)
P. 16	-	-	3 (13.63)	11 (50.00)	8 (36.36)
Total	0 (0.00)	0 (0.00)	37 (16.81)	98 (44.54)	85 (38.63)

Figura 62

Frecuencias de la valoración por ítem geriatras entre 33 - 44 años (N=22).



El grupo de geriatras de edad entre 44 a 54 años (N=7) (**Tabla 98**) (**Figura 63**) mostró una tendencia hacia las valoraciones más altas. Los ítems P. 8 (Bubble plots) y P. 13 (Facilidad para explorar datos) obtuvieron la mayor frecuencia de valoraciones 5, con 6 respuestas (85.71%) cada una. Las valoraciones 4 fueron menos frecuentes, siendo más comunes en los ítems P. 7 (Facilita comprensión) y P. 9 (Representación gráfica), cada una con 3 respuestas (42.85%). No se registraron valoraciones 1 o 2.

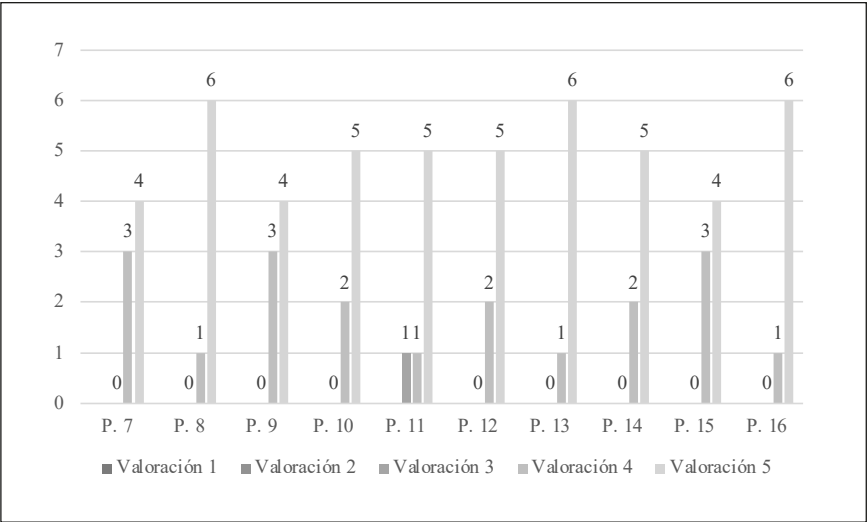
Tabla 98

Frecuencia de valoración de cada ítem en el grupo geriatras de edad entre 45 a 54 años (N=7).

Pregunta (N.º)	Valoración 1 (%)	Valoración 2 (%)	Valoración 3 (%)	Valoración 4 (%)	Valoración 5 (%)
P. 7	-	-	-	3 (42.85)	4 (57.14)
P. 8	-	-	-	1 (14.28)	6 (85.71)
P. 9	-	-	-	3 (42.85)	4 (57.14)
P. 10	-	-	-	2 (28.57)	5 (71.42)
P. 11	-	-	1 (14.28)	1 (14.28)	5 (71.42)
P. 12	-	-	-	2 (28.57)	5 (71.42)
P. 13	-	-	-	1 (14.28)	6 (85.71)
P. 14	-	-	-	2 (28.57)	5 (71.42)
P. 15	-	-	-	3 (42.85)	4 (57.14)
P. 16	-	-	-	1 (14.28)	6 (85.71)
Total	0 (0.00)	0 (0.00)	1 (1.42)	19 (27.14)	50 (71.42)

Figura 63

Frecuencias de valoración por ítem en geriatras entre 45 - 54 años (N=11).



En el grupo de edad de 55 a 65 años (N=7) (**Tabla 99**) (**Figura 64**) las valoraciones 5 fueron predominantes en varios ítems. Los ítems P. 11 (Interpretación rápida), P. 13 (Facilidad para explorar datos), P. 14 (Evolución histórica) y P. 16 (Satisfacción) obtuvieron 7 respuestas (100%) para la valoración 5. Las valoraciones 4 fueron presentes en menor medida, con un máximo de 1 respuesta (14.28%) en varios ítems. No se registraron valoraciones 1 o 2, y solo una valoración 3 fue dada a los ítems P. 10 (Acceso a datos) y P. 15 (Menos conocimientos).

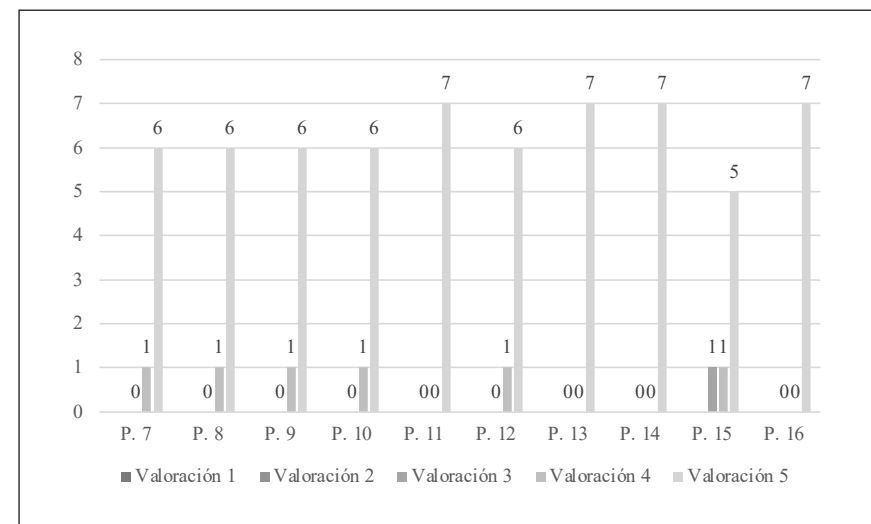
Tabla 99

Frecuencias valoración de cada ítem en grupo de geriatras de edad entre 55 a 65 años (N=7).

Pregunta (N.º)	Valoración 1 (%)	Valoración 2 (%)	Valoración 3 (%)	Valoración 4 (%)	Valoración 5 (%)
P. 7	-	-	-	1 (14.28)	6 (85.71)
P. 8	-	-	-	1 (14.28)	6 (85.71)
P. 9	-	-	-	1 (14.28)	6 (85.71)
P. 10	-	-	-	1 (14.28)	6 (85.71)
P. 11	-	-	-	-	7 (100.00)
P. 12	-	-	-	1 (14.28)	6 (85.71)
P. 13	-	-	-	-	7 (100.00)
P. 14	-	-	-	-	7 (100.00)
P. 15	-	-	1 (14.28)	1 (14.28)	5 (71.42)
P. 16	-	-	-	-	7 (100.00)
Total	0 (0.00)	0 (0.00)	1 (1.42)	6 (8.57)	63 (90.00)

Figura 64

Frecuencia de valoración por ítem en geriatras entre 55 - 65 años (N=7).



ANEXO 15

Subanálisis descriptivo de la frecuencia de valoración por ítem en la encuesta de utilidad de TU® por categorías participantes jóvenes (< 44 años) y participantes mayores (≥ 44 años).

Analizamos a continuación las diferencias en la frecuencia de valor a cada ítem de la encuesta diferenciando entre los profesionales jóvenes frente a los mayores (Tabla 101) (Figura 66) (Tabla 102) (Figura 67).

Al analizar las diferencias en las frecuencias de valoración de cada ítem de la encuesta sobre la utilidad de TU® entre los profesionales jóvenes (menores de 44 años) y los mayores (mayores de 44 años), se observan algunas diferencias notables en la distribución de las valoraciones:

- En el ítem P.7, los geriatras mayores calificaron más positivamente este ítem, con un 71.43% otorgando una valoración de 5, mientras que los jóvenes mostraron un 43.04% en esta categoría. Las valoraciones combinadas de 4 y 5 suman un 100% para los mayores y un 82.28% para los jóvenes. Las valoraciones bajas (1 y 2) fueron inexistentes entre los mayores y sumaron solo un 3.8% entre los jóvenes.
- El ítem P.8, el 85.71% de los mayores dieron una valoración de 5, en comparación con el 50.63% de los jóvenes. Las valoraciones combinadas de 4 y 5 alcanzaron el 100% en los mayores y el 87.34% en los jóvenes. Las valoraciones bajas fueron mínimas, con un 3.80% en los jóvenes y ninguna en los mayores.
- El ítem P.9, el 71.43% de los mayores dieron una valoración de 5, mientras que el 43.04% de los jóvenes dieron la misma puntuación. Sumando las valoraciones de 4 y 5, los mayores alcanzaron el 100% frente al 83.55% de los jóvenes. Las valoraciones bajas fueron de nuevo mínimas, con solo un 3.80% entre los jóvenes.
- El ítem P.10, el 78.57% de los mayores valoraron este ítem con un 5, comparado con el 49.37% de los jóvenes. Las valoraciones combinadas de 4 y 5 alcanzaron el 100% en los mayores y el 84.81% en los jóvenes. Las valoraciones bajas fueron mínimas, con un 3.80% entre los jóvenes y ninguna entre los mayores. Esta diferencia no alcanzó significación estadística (p=0.235).
- En el ítem P.11, con el 85.71% de los mayores otorgando una valoración de 5, frente al 49.37% de los jóvenes. La suma de las valoraciones de 4 y 5, en los mayores alcanzaron el 92.85% en comparación con el 84.81% de los jóvenes. Las valoraciones bajas fueron inexistentes entre los mayores y muy bajas entre los jóvenes (3.80%).
- El ítem P.12, el 78.57% de los mayores dieron una valoración de 5, mientras que el 50.63% de los jóvenes hicieron lo mismo. Las valoraciones combinadas de 4 y 5 suman un 100% para los mayores y un 83.54% para los jóvenes. Las valoraciones bajas fueron mínimas, con solo un 3.80% entre los jóvenes y ninguna entre los mayores.
- En el ítem P.13, con el 92.86% de los mayores valorándolo con un 5, en comparación con el 45.57% de los jóvenes. Las valoraciones combinadas de 4 y 5 suman un 100% para los mayores y un 85.72% para los jóvenes. Las valoraciones bajas fueron mínimas, con solo un 3.80% entre los jóvenes y ninguna entre los mayores.

- El ítem P.14, el 85.71% de los mayores dieron una valoración de 5, frente al 55.70% de los jóvenes. Las valoraciones combinadas de 4 y 5 alcanzaron el 100% en los mayores y el 78.48% en los jóvenes. Las valoraciones bajas fueron mínimas, con solo un 3.80% entre los jóvenes y ninguna entre los mayores.
- El ítem P.15, el 64.29% de los mayores dieron una valoración de 5, en comparación con el 55.70% de los jóvenes. Las valoraciones combinadas de 4 y 5 suman un 92.58% para los mayores y un 78.48% para los jóvenes. Las valoraciones bajas fueron mínimas, con solo un 6.33% entre los jóvenes y ninguna entre los mayores.
- En el ítem P.16, con el 92.86% de los mayores otorgando una valoración de 5, frente al 54.43% de los jóvenes. Las valoraciones combinadas de 4 y 5 suman un 100% para los mayores y un 86.08% para los jóvenes. Las valoraciones bajas fueron mínimas, con solo un 3.80% entre los jóvenes y ninguna entre los mayores.

En general, los profesionales mayores tienden a valorar más positivamente los ítems de la metodología TU® en comparación con los jóvenes. Las diferencias en las valoraciones de los ítems relacionados con la rapidez en la interpretación de datos (P. 11), la facilidad para explorar los datos (P. 13) y la satisfacción general P. 16) con la metodología resultaron estadísticamente significativas. Además, las valoraciones bajas (1 y 2) fueron muy poco frecuentes en ambos grupos, lo que indica una percepción generalmente positiva de la metodología. Esto sugiere una mayor apreciación de estos aspectos por parte de los mayores, posiblemente influenciada por su experiencia y familiaridad con metodologías de análisis de datos.

Tabla 100
Frecuencia de valores a cada pregunta de la encuesta en el grupo de edad Geriatras MAYORES (N=14).

Pregunta (N.º)	Valoración 1 (%)	Valoración 2 (%)	Valoración 3 (%)	Valoración 4 (%)	Valoración 5 (%)
P. 7	-	-	-	4 (28.57)	10 (71.43)
P. 8	-	-	-	2 (14.29)	12 (85.71)
P. 9	-	-	-	4 (28.57)	10 (71.43)
P. 10	-	-	-	3 (21.43)	11 (78.57)
P. 11	-	-	1 (7.14)	1 (7.14)	12 (85.71)
P. 12	-	-	-	3 (21.43)	11 (78.57)
P. 13	-	-	-	1 (7.14)	13 (92.86)
P. 14	-	-	-	2 (14.29)	12 (85.71)
P. 15	-	-	1 (7.14)	4 (28.57)	9 (64.29)
P. 16	-	-	-	1 (7.14)	13 (92.86)
Total	-	-	2 (1.42)	25 (17.85)	113 (80.71)

Figura 66

Frecuencia de valores a cada pregunta de la encuesta en el grupo de edad Geriatras MAYORES (N=14).

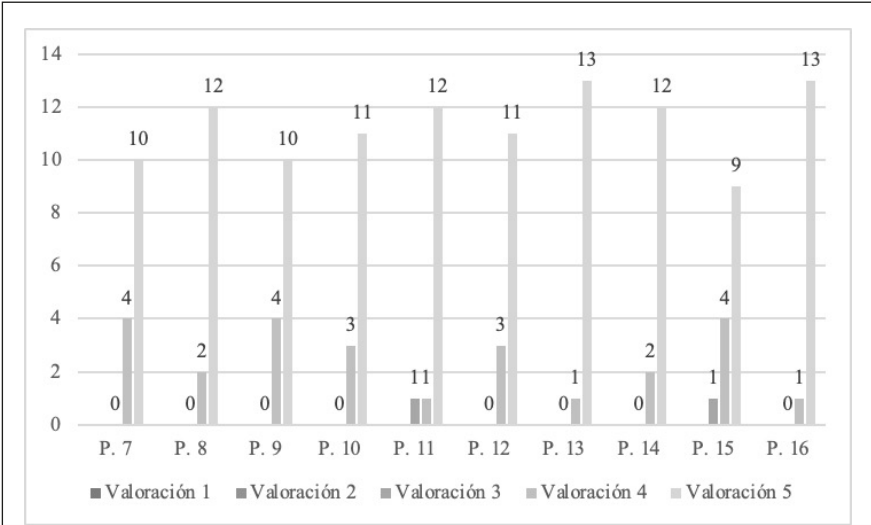


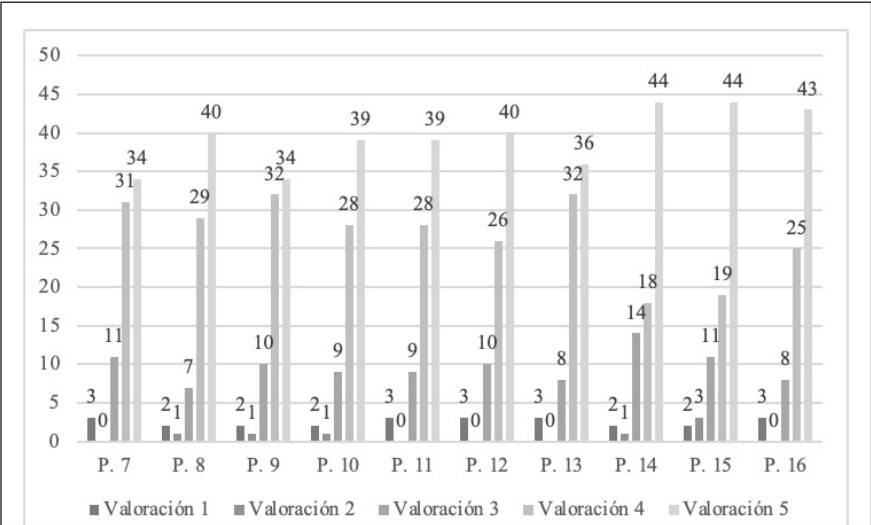
Tabla 101

Frecuencia de valores a cada pregunta de la encuesta en el grupo de edad Geriatras MENORES (N=79).

Pregunta (N.º)	Valoración 1 (%)	Valoración 2 (%)	Valoración 3 (%)	Valoración 4 (%)	Valoración 5 (%)
P. 7	3 (3,8)	-	11 (13,92)	31 (39,24)	34 (43,04)
P. 8	2 (2,53)	1 (1,27)	7 (8,86)	29 (36,71)	40 (50,63)
P. 9	2 (2,53)	1 (1,27)	10 (12,66)	32 (40,51)	34 (43,04)
P. 10	2 (2,53)	1 (1,27)	9 (11,39)	28 (35,44)	39 (49,37)
P. 11	3 (3,8)	-	9 (11,39)	28 (35,44)	39 (49,37)
P. 12	3 (3,8)	-	10 (12,66)	26 (32,91)	40 (50,63)
P. 13	3 (3,8)	-	8 (10,13)	32 (40,51)	36 (45,57)
P. 14	2 (2,53)	1 (1,27)	14 (17,72)	18 (22,78)	44 (55,7)
P. 15	2 (2,53)	3 (3,8)	11 (13,92)	19 (24,05)	44 (55,7)
P. 16	3 (3,8)	-	8 (10,13)	25 (31,65)	43 (54,43)
Total	25 (3,16)	7 (0,89)	97 (12,28)	268 (33,92)	393 (49,75)

Figura 67

Frecuencia de valores a cada pregunta de la encuesta en el grupo de edad Geriatras MENORES (N=79).



ANEXO 16

Análisis descriptivo de la frecuencia de valoración de valoración por ítem en la encuesta de utilidad de TU® entre Geriatras por hospitales.

Análisis descriptivo de la frecuencia de valoración de valoración por ítem en la encuesta de utilidad de TU® entre Geriatras por hospitales.

En el Hospital Universitario Puerta de Hierro Majadahonda (N= 7; 7.5%) (**Tabla 102**) (**Figura 68**), las valoraciones 5 se presentaron en el 41.42% de las 70 opiniones registradas, siendo la P. 14 (Evolución histórica) la más valorada con un 71.42% de valoraciones 5. Las valoraciones 3 (14.28%) se presentaron en varios ítems, lo que sugiere una percepción más variada en este grupo. Las valoraciones más negativas (1 y 2) supusieron de forma conjunta un 12.85%, un porcentaje superior al de los otros hospitales que tuvieron HU. Clínico 5.50%, HU. La Paz 4.75%, HU. Cruz Roja 0.68% y HU. Ramón y Cajal 0.00%.

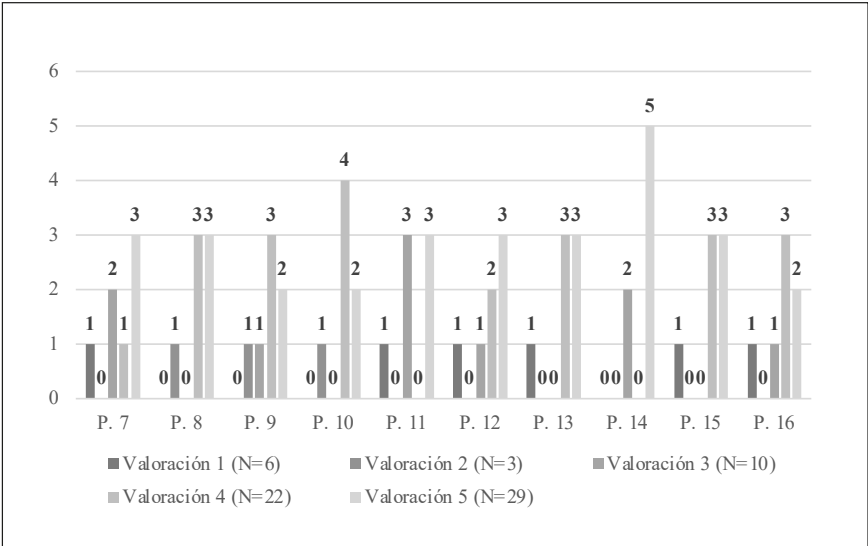
Tabla 102

Frecuencias de valoración de cada ítem en geriatras del HU Puerta de Hierro Majadahonda (N=7; 7.5%).

Pregunta (N.º)	Valoración 1 (%)	Valoración 2 (%)	Valoración 3 (%)	Valoración 4 (%)	Valoración 5 (%)	Valor p
P. 7	1 (14.28)	-	2 (28.57)	1 (14.28)	3 (42.85)	0.358
P. 8	-	1 (14.28)	-	3 (42.85)	3 (42.85)	0.284
P. 9	-	1 (14.28)	1 (14.28)	3 (42.85)	2 (28.57)	0.121
P. 10	-	1 (14.28)	-	4 (57.14)	2 (28.57)	0.100
P. 11	1 (14.28)	-	3 (42.85)	-	3 (42.85)	0.054
P. 12	1 (14.28)	-	1 (14.28)	2 (28.57)	3 (42.85)	0.679
P. 13	1 (14.28)	-	-	3 (42.85)	3 (42.85)	0.578
P. 14	-	-	2 (28.57)	-	5 (71.42)	0.866
P. 15	1 (14.28)	-	-	3 (42.85)	3 (42.85)	0.552
P. 16	1 (14.28)	-	1 (14.28)	3 (42.85)	2 (28.57)	0.601
Total	6 (8.57)	3 (4.28)	10 (14.28)	22 (31.42)	29 (41.42)	

Figura 68

Frecuencias de valoración por ítem en geriatras del HU Puerta de Hierro Majadahonda (N=7; 7.5%).



El Hospital Universitario Clínico San Carlos (N=20; 21.5%) (**Tabla 103**) (**Figura 69**) mostró una predominancia de valoraciones 5 con un porcentaje sobre el total de 200 opiniones recogidas del 58.00%, especialmente en el ítem P. 16 (Satisfacción) con un 75%. Las valoraciones más bajas fueron marginales, con 5.00% de valoraciones 1 y un 0.50% de valoraciones 2. Este hospital presentó una alta consistencia en las valoraciones más positivas (4 y 5) sumando entre ambas el 88.00% de las 200 opiniones registradas.

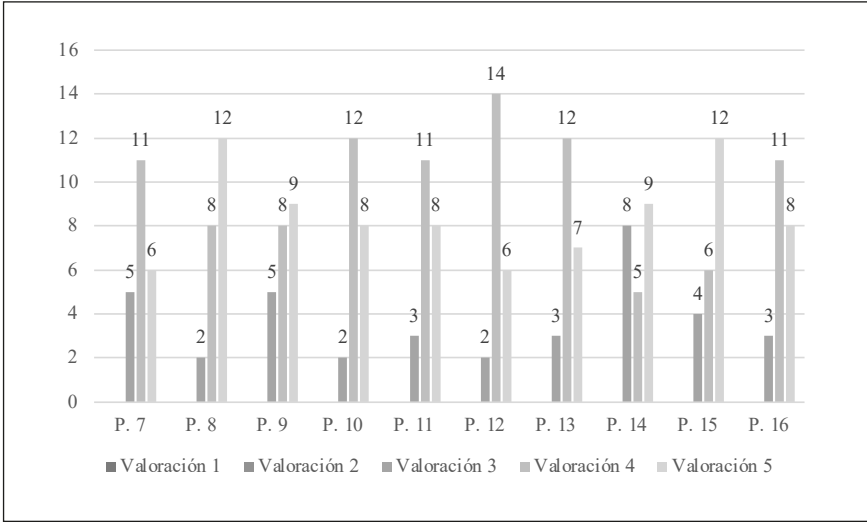
Tabla 103

Frecuencias de valoración de cada ítem en geriatras del HU Clínico San Carlos (N=20; 21.5%).

Pregunta (N.º)	Valoración 1 (%)	Valoración 2 (%)	Valoración 3 (%)	Valoración 4 (%)	Valoración 5 (%)	Valor p
P. 7	1 (5.00)	-	1 (5.00)	7 (35.00)	11 (55.00)	0.358
P. 8	1 (5.00)	-	2 (10.00)	5 (25.00)	12 (60.00)	0.284
P. 9	1 (5.00)	-	1 (5.00)	7 (35.00)	11 (55.00)	0.121
P. 10	1 (5.00)	-	1 (5.00)	9 (45.00)	9 (45.00)	0.100
P. 11	1 (5.00)	-	-	9 (45.00)	10 (50.00)	0.054
P. 12	1 (5.00)	-	1 (5.00)	6 (30.00)	12 (60.00)	0.679
P. 13	1 (5.00)	-	1 (5.00)	7 (35.00)	11 (55.00)	0.578
P. 14	1 (5.00)	1 (5.00)	2(10.00)	4 (20.00)	12 (60.00)	0.866
P. 15	1 (5.00)	-	3 (15.00)	3 (15.00)	13 (65.00)	0.552
P. 16	1 (5.00)	-	1 (5.00)	3 (15.00)	15 (75.00)	0.601
Total	10 (5.00)	1 (0.50)	13 (6.50)	60 (30.00)	116 (58.00)	

Figura 69

Frecuencias de valoración por ítem en geriatras del HU Clínico San Carlos (N=20; 21,5%).



En el Hospital Universitario La Paz (N=21; 22,6%) (**Tabla 104**) (**Figura 70**) las valoraciones 5 fueron predominantes con un porcentaje de 56.19%, de las 200 opiniones registradas, con el ítem P. 16 sobre la satisfacción general con un 71.42%. Se observó una mayor frecuencia en las valoraciones 3, alcanzando un 10.95% del total de las puntuaciones.

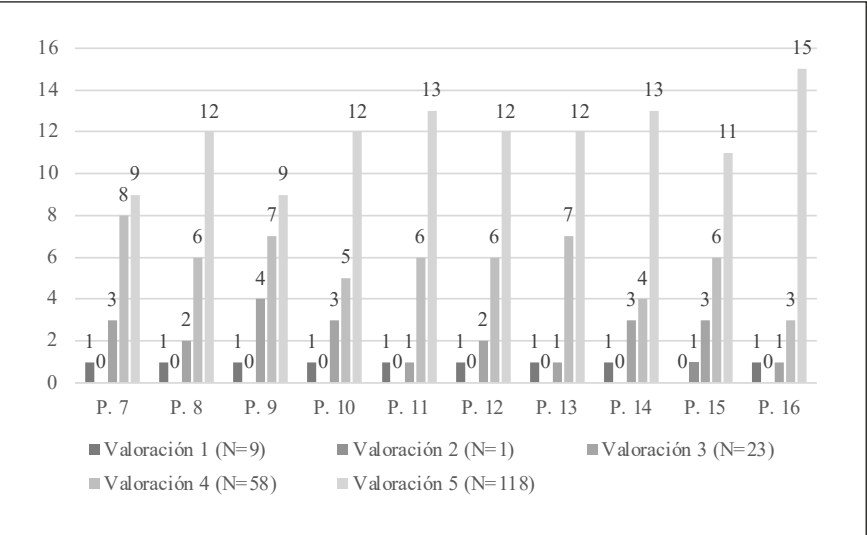
Tabla 104

Frecuencias de valoración de cada ítem en geriatras del HU La Paz (N=21; 22.6%).

Pregunta (N.º)	Valoración 1 (%)	Valoración 2 (%)	Valoración 3 (%)	Valoración 4 (%)	Valoración 5 (%)	Valor p
P. 7	1 (4.76)	-	3 (14.28)	8 (38.09)	9 (42.85)	0.358
P. 8	1 (4.76)	-	2 (9.52)	6 (28.57)	12 (57.14)	0.284
P. 9	1 (4.76)	-	4 (19.04)	7 (33.33)	9 (42.85)	0.121
P. 10	1 (4.76)	-	3 (14.28)	5 (23.80)	12 (57.14)	0.100
P. 11	1 (4.76)	-	1 (4.76)	6 (28.57)	13 (61.90)	0.054
P. 12	1 (4.76)	-	2 (9.52)	6 (28.57)	12 (57.14)	0.679
P. 13	1 (4.76)	-	1 (4.76)	7 (33.33)	12 (57.14)	0.578
P. 14	1 (4.76)	-	3 (14.28)	4 (19.04)	13 (61.90)	0.866
P. 15	-	1 (4.76)	3 (14.28)	6 (28.57)	11 (52.38)	0.552
P. 16	1 (4.76)	-	1 (4.76)	3 (14.28)	15 (71.42)	0.601
Total	9 (4.28)	1 (0.47)	23 (10.95)	58 (27.61)	118 (56.19)	

Figura 70

Frecuencias de valoración por ítem en geriatras del HU La Paz (N=21; 22,6%).



El Hospital Universitario Ramón y Cajal (N=16; 17,2%) (**Tabla 105**) (**Figura 71**) presentó una alta proporción de valoraciones 4 (45.62%) y 5 (46.87%) en todas las preguntas alcanzando de media entre las dos valoraciones un porcentaje del 92.49% de las 160 opiniones registradas, sin ninguna valoración 1 o 2. Los ítems P. 15 (Menos conocimientos) y P. 16 (Satisfacción) recibieron un 62.50% de valoraciones 5.

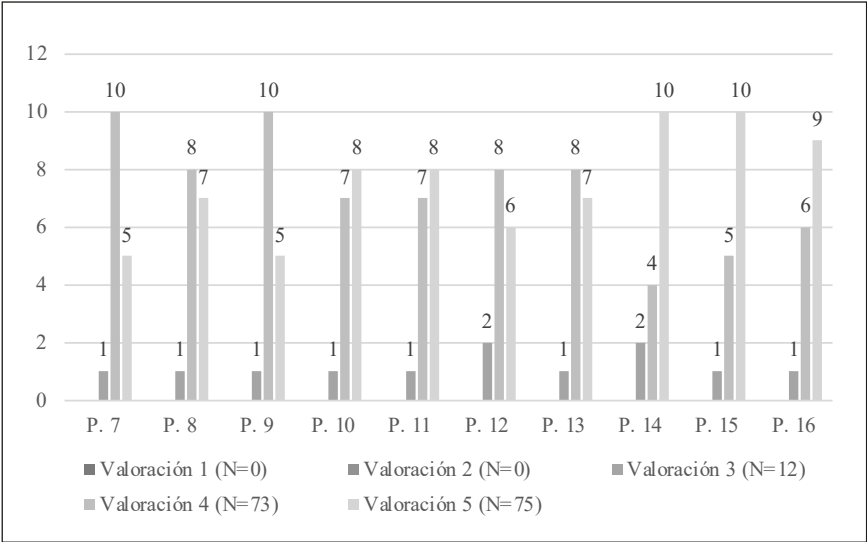
Tabla 105

Frecuencias de valoración de cada ítem en geriatras del HU Ramón y Cajal (N=16; 17.2%).

Pregunta (N.º)	Valoración 1 (%)	Valoración 2 (%)	Valoración 3 (%)	Valoración 4 (%)	Valoración 5 (%)	Valor p
P. 7	-	-	1 (6.25)	10 (62.50)	5 (31.25)	0.358
P. 8	-	-	1 (6.25)	8 (50.00)	7 (43.75)	0.284
P. 9	-	-	1 (6.25)	10 (62.50)	5 (31.25)	0.121
P. 10	-	-	1 (6.25)	7 (43.75)	8 (50.00)	0.100
P. 11	-	-	1 (6.25)	7 (43.75)	8 (50.00)	0.054
P. 12	-	-	2 (12.50)	8 (50.00)	6 (37.50)	0.679
P. 13	-	-	1 (6.25)	8 (50.00)	7 (43.75)	0.578
P. 14	-	-	2 (12.50)	4 (25.00)	10 (62.50)	0.866
P. 15	-	-	1 (6.25)	5 (31.25)	10 (62.50)	0.552
P. 16	-	-	1 (6.25)	6 (37.50)	9 (56.25)	0.601
Total	0 (0.00)	0 (0.00)	12 (7.50)	73 (45.62)	75 (46.87)	

Figura 71

Frecuencias de valoración por ítem en geriatras del HU Ramón y Cajal (N=16; 17,2%).



El Hospital Cruz Roja (N=29; 31,2%) (**Tabla 106**) (**Figura 72**) mostró una alta proporción de valoraciones 5, destacándose con el porcentaje más elevado entre hospitales con el 59.31% de las 290 opiniones registradas, especialmente en los ítems P. 10 (Acceso a datos) y la P. 16 (Satisfacción) con un 65.51%. Las valoraciones más bajas fueron insignificantes, con solo un 0.68% de valoraciones 2.

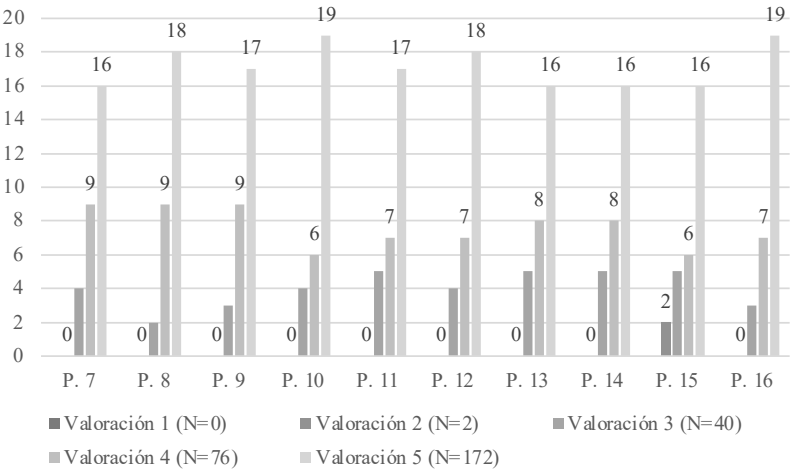
Tabla 106

Frecuencias de valoración de cada ítem en geriatras del HU Cruz Roja (N=29; 31.2%).

Pregunta (N.º)	Valoración 1 (%)	Valoración 2 (%)	Valoración 3 (%)	Valoración 4 (%)	Valoración 5 (%)	Valor p
P. 7	-	-	4 (13.79)	9 (31.03)	16 (55.17)	0.358
P. 8	-	-	2 (6.89)	9 (31.03)	18 (62.06)	0.284
P. 9	-	-	3 (10.34)	9 (31.03)	17 (58.62)	0.121
P. 10	-	-	4 (13.79)	6 (20.68)	19 (65.51)	0.100
P. 11	-	-	5 (17.24)	7 (24.13)	17 (58.62)	0.054
P. 12	-	-	4 (13.79)	7 (24.13)	18 (62.06)	0.679
P. 13	-	-	5 (17.24)	8 (27.58)	16 (55.17)	0.578
P. 14	-	-	5 (17.24)	8 (27.58)	16 (55.17)	0.866
P. 15	-	2 (6.89)	5 (17.24)	6 (20.68)	16 (55.17)	0.552
P. 16	-	-	3 (10.34)	7 (24.13)	19 (65.51)	0.601
Total	0 (0.00)	2 (0.68)	40 (13.79)	76 (26.20)	172 (59.31)	

Figura 72

Frecuencias de valoración por ítem en geriatras del HU Cruz Roja (N=29; 31,2%).



Se realizó un análisis estadístico realizado mediante la prueba de chi-cuadrada para determinar si existen diferencias significativas entre las respuestas de los geriatras de los cinco hospitales que participaron en la encuesta de utilidad de TU® sin encontrar diferencias estadísticamente significativas.



D. JOSÉ ANTONIO BLANCO SERRANO, Director de la Fundación Universidad Alfonso X el Sabio (FUAX)

HACE CONSTAR:

Que **D. Francisco Javier Gómez Pavón**, con **DNI 50158659Y**, profesor de la Universidad Alfonso X el Sabio, forma parte del equipo de investigadores que realiza el proyecto de investigación, autorizado de manera oficial el día 22 de diciembre de 2022 y seleccionado por una Comisión Evaluadora, con las siguientes características:

- **Título del Proyecto:** “Validación de la metodología de síntesis terciaria The Umbrella Collaboration Ö y de su aplicación informática y de inteligencia artificial The Umbrella Ö”
- **Código Proyecto:** 1.014.030
- **Centro:** Hospital Cruz Roja y Universidad Alfonso X El Sabio.
- **Carácter:** I+D+i. Multidisciplinar.
- **Entidad financiadora:** Fundación Universidad Alfonso X El Sabio-Banco Santander. XIV Convocatoria de Ayudas para Grupos de Investigación, Desarrollo e Innovación Uax-Santander 2022/2023.
- **Competitivo:** Si.
- **Financiación obtenida:** 6.000€
- **Duración:** 1 año. Evaluada Externamente por EQA y la organización.
- **Fecha de inicio:** 01/01/2023 **Fecha de término:** 31/12/2023
- **Director o investigador principal:** Francisco Javier Gómez Pavón (1/01/2023 a 20/04/2023) y Blanca Garmendia Prieto (21/04/2023 a 31/12/2023).
- **Grupo de investigación:** Blanca Garmendia Prieto, Javier Gómez Pavón, Pamela Lisette Carrillo, Beltrán Carrillo Manrique, Marta Rubinos Cuadrado, Jazmín Parellada Martín y Alejandra Palacios López.

Y para que así conste a efectos de acreditación del solicitante, se expide el presente documento, en Villanueva de la Cañada a 20 de septiembre de 2023.

FUNDACIÓN UNIVERSIDAD
ALFONSO X EL SABIO
C.I.F.: G-84760784
Avda. de la Universidad, nº 1
28691 Villanueva de la Cañada (Madrid)

Fdo. D. José Antonio Blanco Serrano
Director FUAX

Avda. de la Universidad, 1 • 28691 Villanueva de la Cañada (Madrid)
Tel.: 34 91 810 93 00 • E-mail: fundacion@uax.es • Internet: <http://fundacion.uax.es>

ANEXO 18

CERTIFICADO PRESENTACIÓN 64 CONGRESO DE LA SOCIEDAD ESPAÑOLA DE GERIATRÍA Y GERONTOLOGÍA, MALAGA 2024

DISEÑANDO UN NUEVO FUTURO,
CONSTRUYENDO REDES, TRABAJANDO JUNTOS

64 CONGRESO
S E G G
26 a 28 de Junio
44 CONGRESO
S A G G
2024 Málaga

CERTIFICADO DE PRESENTACIÓN

La Sociedad Española de Geriatria y Gerontología

Certifica que

Beltran Carrillo Manrique, Marta Rubinos Cuadrado, Jazmin Parellada Martin,
Alejandra Palacios López, Beltran Carrillo Rubinos, Juan Jose Baztan Cortés, Javier
Goméz Pavon

han presentado la Comunicación tipo **ORAL**, titulada:

**OC-052 - EFECTIVIDAD DE LA METODOLOGÍA “THE UMBRELLA
COLLABORATION®” INTEGRANDO INTELIGENCIA ARTIFICIAL EN LA
REALIZACIÓN DE REVISIONES UMBRELLA TRADICIONALES**

durante el **64 Congreso de la Sociedad Española de
Geriatria y Gerontología**

celebrado en Málaga del 26 al 28 de junio de 2024.

Y para que así conste a todos los efectos, firma la presente **CERTIFICACIÓN**,
en Málaga, a 28 de junio de 2024.

José Augusto García Navarro
Presidente de la Sociedad Española de Geriatria y Gerontología (SEGG)

DOCUMENTO FIRMADO DIGITALMENTE
<https://intranet.pacifico-meetings.com/Espresso/CertificadoFirmado.xhtml?ccf=AY917-C224006-PCA1030-PD11368-V1>

ANEXO 19

CERTIFICADO PRESENTACIÓN 20 CONGRESO DE LA EUROPEAN GERIATRIC MEDICINE SOCIETY, VALENCIA 2024



20th

VALENCIA, SEPT 18-20, 2024







POSTER CERTIFICATE

We herewith confirm the presentation of the Abstract

Title: UTILITY OF A NEW TERTIARY SYNTHESIS SYSTEM, “THE UMBRELLA COLLABORATON®”, IN GERIATRICS

Abstract Area: AI and Big Data

Authors: Beltran Carrillo (1), Marta Rubinos-Cuadrado (1), Jazmin Parellada-Martin (1), Alejandra Palacios-López (1), Beltrán Carrillo-Rubinos (1), Ana Carrillo-Rubinos (1), Fernando Canillas del Rey (2), Juan Jose Baztán-Cortes (2)

Affiliations: (1) The Umbrella Collaboration, (2) Hospital Universitario Cruz Roja, Madrid

Mirko Petrovic

EuGMS President



Nathalie van der Velde

EuGMS Academic Director



268

269

9

————— **PUBLICACIÓN** —————



The Leading eHealth Publisher

The Umbrella Collaboration®: An Innovative Tertiary Evidence Synthesis Methodology. Validation Study Protocol

Beltran Carrillo, Marta Rubinos-Cuadrado, Jazmin Parellada-Martin, Alejandra Palacios-López, Beltran Carrillo-Rubinos, Fernando Canillas-Del Rey, Juan Jose Baztán-Cortes, Javier Gómez-Pavon

Submitted to: JMIR Research Protocols
on: October 06, 2024

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on it's website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

<https://preprints.jmir.org/preprint/67248>[unpublished, non-peer-reviewed preprint]

JMIR Preprints

Carrillo et al

Table of Contents

Original Manuscript.....	5
Supplementary Files.....	25
.....	26
.....	27
Figures	28
Figure 1.....	29

<https://preprints.jmir.org/preprint/67248>[unpublished, non-peer-reviewed preprint]

The Umbrella Collaboration®: An Innovative Tertiary Evidence Synthesis Methodology. Validation Study Protocol

Beltran Carrillo^{1*} MD; Marta Rubinos-Cuadrado^{1*} RNC; Jazmin Parellada-Martin¹ BSCS; Alejandra Palacios-López¹ BME; Beltran Carrillo-Rubinos¹ BME; Fernando Canillas-Del Rey^{2,3} MD, PhD; Juan Jose Baztán-Cortes³ MD, PhD; Javier Gómez-Pavon^{2,3} MD, PhD

¹The Umbrella Collaboration Madrid ES
²Universidad Alfonso X el Sabio Villanueva de la Cañada, Madrid ES
³Hospital Universitario Cruz Roja Madrid ES
*these authors contributed equally

Corresponding Author:
Beltran Carrillo MD
The Umbrella Collaboration
C/ Ferraz, 49 - 1º izq
Madrid
ES

Abstract

Background: The synthesis of evidence in healthcare is essential for informed decision-making and policy development. This study aims to validate The Umbrella Collaboration® (TU®), an innovative, semi-automatic tertiary evidence synthesis methodology, by comparing it with Traditional Umbrella Reviews (TUR), which are currently the gold standard.

Objective: The primary objective of the present study is to assess whether a software-driven AI-assisted system of evidence synthesis, TU®, can match the effectiveness of traditional methods of tertiary synthesis, providing a potentially more timely, efficient, and comprehensive approach while remaining open to findings that could demonstrate superior performance. To support the primary objective of evaluating the effectiveness of TU® compared to traditional methodologies, this study also aims to assess the accessibility and comprehensibility of TU®'s outputs as a secondary objective

Methods: This protocol outlines a comparative study divided into two main parts. The first part involves a quantitative comparison of results obtained using TU® and TURs in geriatrics. We will evaluate the identification, size effect, direction, statistical significance, and certainty of outcomes, as well as the time and resources required for each methodology. Data for TURs will be sourced from Medline (via PubMed), while TU® will use AI-assisted informatics to replicate the research questions of the selected URTs. The second part of the study assesses the ease of use and comprehension of TU® through an online survey directed at health professionals, utilizing interactive features and detailed data access.

Results: Expected results include the assessment of concordance in identifying outcomes, the size effect, direction and significance of these outcomes, and the certainty of evidence. Additionally, we will measure the operational efficiency of each methodology by evaluating the time taken to complete projects. User perceptions of the ease of use and comprehension of TU® will be gathered through detailed surveys.

Conclusions: If TU® proves as effective as TURs but more time-efficient, accessible and easily updatable, it could significantly enhance the process of evidence synthesis, facilitating informed decision-making and improving healthcare. This study represents a step towards integrating innovative technologies into routine evidence synthesis practice, potentially transforming health research.

(JMIR Preprints 06/10/2024:67248)
DOI: <https://doi.org/10.2196/preprints.67248>

Preprint Settings

- 1) Would you like to publish your submitted manuscript as preprint?
✔ Please make my preprint PDF available to anyone at any time (recommended).

- Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.
2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?
✔ Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to all users.
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in [http://www.jmir.org](#), I will be able to remove the preprint from the public domain.

Original Manuscript

The Umbrella Collaboration®: An Innovative Tertiary Evidence Synthesis Methodology. Validation Study Protocol

Authors:

Beltrán Carrillo^{a,b}, Rubinos-Cuadrado, M^a, Parellada-Martin, J^a, Palacios-López, A^a, Carrillo-Rubinos, B^a, Canillas del Rey, F^{b,c}, Baztán-Cortés, JJ^b, Gómez-Pavón, J^{b,c}.

a: The Umbrella Collaboration®

b: Hospital Universitario Cruz Roja, (Madrid, Spain)

c: Universidad Alfonso X El Sabio, Villanueva de la Cañada (Madrid, Spain)

email autor: Beltran Carrillo. (bcm@theumbrellacollaboration.org)

ORCID:

Beltran Carrillo: 0000-0001-5748-9423

Rubinos-Cuadrado, M: 0000-0003-0108-2629

Parellada-Martin, J: 0000-0002-8528-4481

Palacios-López, A: 0000-0002-7263-9167

Carrillo-Rubinos, B: 0000-0002-0796-4414

Canillas del Rey, F:

Baztán-Cortes, JJ: 0009-00099792-0404

Gómez-Pavón, J: 0000-0002-4732-6265

ABSTRACT

BACKGROUND:

The synthesis of evidence in healthcare is essential for informed decision-making and policy development. This study aims to validate The Umbrella Collaboration® (TU®), an innovative, semi-automatic tertiary evidence synthesis methodology, by comparing it with Traditional Umbrella Reviews (TUR), which are currently the gold standard.

METHODS:

This protocol outlines a comparative study divided into two main parts. The first part involves a quantitative comparison of results obtained using TU® and TURs in geriatrics. We will evaluate the identification, size effect, direction, statistical significance, and certainty of outcomes, as well as the time and resources required for each methodology. Data for TURs will be sourced from Medline (via PubMed), while TU® will use AI-assisted informatics to replicate the research questions of the

selected URTs. The second part of the study assesses the ease of use and comprehension of TU[®] through an online survey directed at health professionals, utilizing interactive features and detailed data access.

RESULTS:

Expected results include the assessment of concordance in identifying outcomes, the size effect, direction and significance of these outcomes, and the certainty of evidence. Additionally, we will measure the operational efficiency of each methodology by evaluating the time taken to complete projects. User perceptions of the ease of use and comprehension of TU[®] will be gathered through detailed surveys.

DISCUSSION:

The implementation of new methodologies in evidence synthesis requires validation. This study will determine whether TU[®] can match the accuracy and comprehensiveness of TURs while offering benefits in terms of efficiency and user accessibility. The comparative study is designed to address the inherent challenges in validating a new methodology against established standards.

CONCLUSION:

If TU[®] proves as effective as TURs but more time-efficient, accessible and easily updatable, it could significantly enhance the process of evidence synthesis, facilitating informed decision-making and improving healthcare. This study represents a step towards integrating innovative technologies into routine evidence synthesis practice, potentially transforming health research.

FUNDING:

This project has received a research grant from the Fundación Alfonso X el Sabio.

CONFLICTS OF INTEREST:

The authors declare no conflicts of interest.

KEY WORDS: Tertiary Evidence Synthesis, The Umbrella Collaboration, Umbrella Reviews, Health Research Methodology, AI-assisted Synthesis, Evidence-based Decision Making

The Umbrella Collaboration®: An Innovative Tertiary Evidence Synthesis Methodology. Validation Study Protocol

INTRODUCTION

The synthesis of evidence in healthcare is a knowledge acquisition process designed to transform extensive volumes of data into manageable information to support decision-making based on the best possible evidence. It aims to integrate information from multiple sources on complex topics in a comprehensive, precise, transparent, and easily understandable manner. These principles ensure that synthesized evidence is accessible and useful to all stakeholders, including healthcare professionals, policymakers, and patients (1)(The Royal Society of Evidence Synthesis, 2018)¹.

Evidence synthesis plays a pivotal role in Knowledge Translation (KT) (2) (WHO, 2005)², serving as a bridge between research and healthcare. Consequently, evidence synthesis is essential for the development of health policies and informed decision-making (3) (Cottrell, 2014)³.

The role of all stakeholders in health decision-making underscores the importance of their ability to access, understand, and evaluate health information adequately (4) (Sorensen, 2012)⁴. A persistent challenge in KT is the low level of statistical and health literacy among the general population and health professionals, which significantly complicates effective health management (5,6) (Martinez-Garcia, 2022)⁵ (Baccolini, 2021)⁶. Therefore, it is crucial to democratize access to high-level health information and to promote active participation of all stakeholders in decisions affecting healthcare (7)(Baumann, 2022)⁷.

¹ Accessible: <https://royalsociety.org/-/media/policy/projects/evidence-synthesis/evidence-synthesis-statement-principles.pdf>

² World Health Organization. (2005). Bridging the “know-do” gap meeting on knowledge translation in global health”; 10-12 Oct 2005; Geneva, Switzerland. Geneva: WHO. (accessible: <https://www.measureevaluation.org/resources/training/capacity-building-resources/high-impact-research-training-curricula/bridging-the-know-do-gap.pdf>)

³ Cottrell E, Whitlock E, Kato E, Uhl S, Belinson S, Chang C, Hoomans T, Meltzer D, Noorani H, Robinson K, Schoelles K, Motu'apuaka M, Anderson J, Paynter R, Guise JM. Defining the Benefits of Stakeholder Engagement in Systematic Reviews [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); 2014 Mar. Report No.: 14-EHC006-EF. PMID: 24783309. <https://www.ncbi.nlm.nih.gov/books/NBK196180/>

⁴ Sorensen K, Van den Broucke S, Fullam J, Doyle G, Pelikan J, Slonska Z, Brand H; (HLS-EU) Consortium Health Literacy Project European. Health literacy and public health: a systematic review and integration of definitions and models. BMC Public Health. 2012 Jan 25;12:80. doi: 10.1186/1471-2458-12-80. PMID: 22276600; PMCID: PMC3292515. <https://bmcpublihealth.biomedcentral.com/articles/10.1186/1471-2458-12-80>

⁵ Martínez-García, J. J., Canizalez-Román, A., Velázquez-Román, J. A., Flores-Villaseñor, H. M., & León-Sicaños, N. M. (2022). Evaluación del conocimiento de métodos básicos de epidemiología e investigación en médicos residentes. *Revista Médica de la Universidad Autónoma de Sinaloa REVMEUAS*, 11(2), 115-123. <https://www.medigraphic.com/cgi-bin/new/resumen.cgi?IDARTICULO=106198>

⁶ Baccolini V, Rosso A, Di Paolo C, Isonne C, Salerno C, Migliara G, Prencipe GP, Massimi A, Marzuillo C, De Vito C, Villari P, Romano F. What is the Prevalence of Low Health Literacy in European Union Member States? A Systematic Review and Meta-analysis. J Gen Intern Med. 2021 Mar;36(3):753-761. doi: 10.1007/s11606-020-06407-8. Epub 2021 Jan 5. PMID: 33403622; PMCID: PMC7947142. <https://link.springer.com/article/10.1007/s11606-020-06407-8>

⁷ Baumann LA, Reinhold AK, Brütt AL. Public and patient involvement in health policy decision-making on the health system level - A scoping review. Health Policy. 2022 Oct;126(10):1023-1038. doi: 10.1016/j.healthpol.2022.07.007. Epub 2022 Jul 20. PMID: 35918211. <https://www.sciencedirect.com/science/article/pii/S0168851022001919?via%3Dihub>

The evolution of evidence synthesis methodologies has led to the development of tertiary synthesis, designed to condense knowledge from multiple systematic reviews with or without meta-analyses (SR/MA). This synthesis, often referred to as Umbrella Reviews now referred to as Traditional Umbrella Reviews (TUR) to distinguish them from the experimental methodology under study, The Umbrella Collaboration® (TU®), builds upon the concept of primary studies (individual studies with participant samples) and secondary studies (systematic reviews and analysis of those primary studies). Tertiary synthesis represents a third level, named with other terms such as overviews, meta-epidemiological studies, meta-meta-analyses, meta-synthesis, and meta-reviews also describing this approach (8,9) (Giuseppe Biondi-Zoccai, 2016)⁸ (Choi, 2023)⁹. Tertiary synthesis has gained prominence in contexts where broad research questions are posed, rapid results are needed, and resources for extensive systematic reviews are limited. TURs follow a structured methodological process that involves several clearly defined stages (9–14) (Fusar-Poli, 2018)¹⁰ (Aromataris, 2020)¹¹ (Cant, 2022)¹² (Belbasis, 2022)¹³ (Pollock, 2023)¹⁴ (Choi, 2023)¹⁵ (Aromataris, 2024)¹⁶. Although organizations like Cochrane Collaboration (14) (Pollock, 2023)¹⁷ and the Joanna Briggs Institute (JBI) (11) (Aromataris, 2024)¹⁸ have developed and continually updated detailed methodologies for these reviews, there remains considerable divergence in how TUR authors implement these steps in practice. This methodological freedom leads to significant variations among different TURs in terms of rigor and approach.

The urgency for high-quality, timely information during crises like the SARS-CoV-2 pandemic has

⁸ Biondi-Zoccai, G. (2016). Umbrella reviews. *Evidence synthesis with overviews of reviews and meta-epidemiologic studies*. Cham, Switzerland: Springer International.

⁹ Choi GJ, Kang H. Introduction to Umbrella Reviews as a Useful Evidence-Based Practice. J Lipid Atheroscler. 2023 Jan;12(1):3-11. doi: 10.12997/jla.2023.12.1.3. Epub 2022 Oct 21. PMID: 36761061; PMCID: PMC9884555. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9884555/>

¹⁰ Fusar-Poli P, Radau J. Ten simple rules for conducting umbrella reviews. Evid Based Ment Health. 2018 Aug;21(3):95-100. doi: 10.1136/ebmental-2018-300014. Epub 2018 Jul 13. PMID: 30006442; PMCID: PMC10270421. <https://mentalhealth.bmj.com/content/21/3/95.long>

¹¹ Aromataris E, Fernandez R, Godfrey C, Holly C, Khalil H, Tungpunkom P. Chapter 10: Umbrella Reviews. Aromataris E, Munn Z, editors. *JBI Manual for Evidence Synthesis*. JBI; 2020. Available from <https://synthesismanual.jbi.global>. <https://doi.org/10.46658/JBIMES-20-11>

¹² Cant, R., Ryan, C., & Kelly, M. A. (2022). A nine-step pathway to conduct an umbrella review of literature. <https://onlinelibrary.wiley.com/doi/full/10.1111/nae2.12039>

¹³ Belbasis L, Bellou V, Ioannidis JPA. Conducting umbrella reviews. BMJ Med. 2022 Nov 22;1(1):e000071. doi: 10.1136/bmjmed-2021-000071. PMID: 36936579; PMCID: PMC9951359.

¹⁴ Pollock M, Fernandes RM, Becker LA, Pieper D, Hartling L. Chapter V: Overviews of Reviews. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.4 (updated August 2023). Cochrane, 2023. Available from www.training.cochrane.org/handbook.

¹⁵ Choi GJ, Kang H. Introduction to Umbrella Reviews as a Useful Evidence-Based Practice. J Lipid Atheroscler. 2023 Jan;12(1):3-11. doi: 10.12997/jla.2023.12.1.3. Epub 2022 Oct 21. PMID: 36761061; PMCID: PMC9884555.

¹⁶ Aromataris E, Lockwood C, Porritt K, Pilla B, Jordan Z, editors. *JBI Manual for Evidence Synthesis*. JBI; 2024. Available from: <https://synthesismanual.jbi.global>. <https://doi.org/10.46658/JBIMES-24-01>

¹⁷ Pollock M, Fernandes RM, Becker LA, Pieper D, Hartling L. Chapter V: Overviews of Reviews. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.4 (updated August 2023). Cochrane, 2023. Available from www.training.cochrane.org/handbook.

¹⁸ Aromataris E, Lockwood C, Porritt K, Pilla B, Jordan Z, editors. *JBI Manual for Evidence Synthesis*. JBI; 2024. Available from: <https://synthesismanual.jbi.global>. <https://doi.org/10.46658/JBIMES-24-01>

highlighted the critical need for faster evidence synthesis methods, even if it means accepting certain limitations in comprehensiveness, detail, and precision (15) (Tricco, 2022)¹⁹. This demand has driven the development of innovative approaches such as The Umbrella Collaboration® (TU®), which leverages AI-assisted software to facilitate tertiary evidence synthesis under human oversight. While AI tools such as PICO Portal, DistillerSR, Covidence, and Rayyan have already improved secondary evidence synthesis by streamlining data management and analysis (16) (Qureshi, 2023)²⁰, technologies like large language models (LLMs), including ChatGPT, are beginning to demonstrate potential for conducting systematic reviews autonomously, though human supervision remains essential to mitigate risks such as errors and hallucinations (17) (Teperikidis, 2024)²¹. Despite these developments in secondary synthesis, the application of AI and software engineering in tertiary synthesis is still in its early stages, with no dedicated software currently available. TU® is at the forefront of this field, pioneering the integration of AI and software engineering with human oversight to ensure accuracy and minimize technology-induced errors. As AI continues to evolve, it is likely that fully automated processes for both secondary and tertiary synthesis will emerge, potentially revolutionizing clinical research and practice. However, building confidence in these technologies will require ongoing development and rigorous validation.

This article was previously presented as a meeting abstract at the 64^o Spanish Society of Geriatrics and Gerontology (Sociedad Española de Geriatria y Gerontologia) Congress on June 26 - 28, 2024.

The Umbrella Collaboration® (patent pending)

TU® is primarily a software-driven system engineered to streamline tertiary evidence synthesis, relying on programmed algorithms to automate the majority of its functions. The core of the system is built on a software infrastructure that processes and synthesizes data from SR/MA abstracts stored in Medline. While AI plays a crucial role, particularly through the use of LLMs and machine learning (ML), it is utilized selectively within the broader software framework to enhance specific tasks.

LLMs are used in generating related search terms, expanding upon human-generated queries to enhance the comprehensiveness of literature searches. Any LLM can be adapted to TU® software, up

¹⁹ Tricco AC, Straus SE, Ghaffar A, Langlois EV. Rapid reviews for health policy and systems decision-making: more important than ever before. *Syst Rev*. 2022 Jul 30;11(1):153. doi: 10.1186/s13643-022-01887-7. PMID: 35906637; PMCID: PMC9338614.
²⁰ Qureshi R, Shaughnessy D, Gill KAR, Robinson KA, Li T, Agai E. Are ChatGPT and large language models "the answer" to bringing us closer to systematic review automation? *Syst Rev*. 2023 Apr 29;12(1):72. doi: 10.1186/s13643-023-02243-z. PMID: 37120563; PMCID: PMC10148473.
²¹ Teperikidis L, Boulmpou A, Papadopoulos C, Biondi-Zoccai G. Using ChatGPT to perform a systematic review: a tutorial. *Minerva Cardiol Angiol*. 2024 Jul 26. doi: 10.23736/S2724-5683.24.06568-2. Epub ahead of print. PMID: 39056432.

to date we have used ChatGPT 4 (18) (OpenAI)²². This function is crucial in broadening the scope of the searches while ensuring that the results remain relevant and precise. The AI component is designed to support, not replace, human oversight, ensuring that the final selection of literature is both accurate and comprehensive (19) (Gwon, 2024)²³.

As the TU® database grows, ML will be integrated to further refine and optimize the software's performance. Training the system on an expanding dataset is expected to enhance its ability to select, categorize, and analyze relevant research, thereby improving both efficiency and accuracy over time. This approach allows TU® to evolve, continuously improving its utility in evidence synthesis through the iterative learning process (20) (Kothamali, 2024)²⁴.

Overall, TU® represents a hybrid model where traditional software engineering and targeted AI applications work in tandem. This balance ensures that while the software performs most functions automatically, AI enhances specific tasks, such as search term generation and future predictive analysis, under strict human supervision. This strategic integration of AI elements within a primarily software-driven system ensures the reliability and precision of the evidence synthesis process.

The outcomes of TU® are made available through an interactive public web application specifically designed to be accessible and comprehensible to all stakeholders, regardless of their statistical expertise. The combination of intuitive graphical formats and plain language ensures that the findings are easily interpreted by a wide audience.

Additionally, the platform supports continuous updates, automatically integrating new SR/MA data every 24 hours, thereby ensuring the most current and reliable evidence synthesis. This approach aligns with the concept of Living Systematic Reviews (LSRs) (21) (Elliot, 2017)²⁵, which advocate for frequent updates to maintain relevance in rapidly evolving fields. While Cochrane Collaboration recommends updating LSRs monthly (22) (Cochrane, 2019)²⁶, TU® surpasses this standard by ensuring updates are incorporated daily, providing near real-time evidence synthesis.

²² OpenAI. 2024. *ChatGPT* (versión GPT-4). <https://chat.openai.com/chat>.
²³ Gwon YN, Kim JH, Chung HS, Jung EJ, Chun J, Lee S, Shim SR. The Use of Generative AI for Scientific Literature Searches for Systematic Reviews: ChatGPT and Microsoft Bing AI Performance Evaluation. *JMIR Med Inform*. 2024 May 14;12:e51187. doi: 10.2196/51187. PMID: 38771247; PMCID: PMC11107769.
²⁴ Kothamali, P. R., Karne, V. K., & Dandiyala, S. S. M. (2024, July). Integrating AI and Machine Learning in Quality Assurance for Automation Engineering. In *International Journal for Research Publication and Seminar* (Vol. 15, No. 3, pp. 93-102).
²⁵ Elliott, J. H., Synnot, A., Turner, T., Simmonds, M., Akl, E. A., McDonald, S., ... & Pearson, L. (2017). Living systematic review: 1. Introduction—the why, what, when, and how. *Journal of clinical epidemiology*, 91, 23-30.
²⁶ Cochrane: Guidance for the production and publication of Cochrane living systematic reviews: Cochrane Reviews in living mode. 2019. https://community.cochrane.org/sites/default/files/uploads/inline-files/Transform/201912_LSR_Revised_Guidance.pdf

The daily updates provided by TU® are further complemented by its remarkable efficiency. Pilot tests have demonstrated TU®'s capability to complete tertiary evidence synthesis projects within hours, a significant reduction in time compared to traditional methods. If validated by the upcoming research, this advancement could demonstrate TU®'s potential to streamline the synthesis process, delivering rapid yet reliable results while upholding the highest standards of accuracy. Should these findings be confirmed, TU® may emerge as an invaluable tool for accelerating the pace of evidence-based research.

A schematic diagram is provided to clearly illustrate TU®'s workflow, showing the process from data acquisition to the generation of synthesized results. Abstracts of SR/MA, retrieved from the Medline database via PubMed, form the foundation of the analysis. The data is processed through a range of techniques, including Natural Language Processing (NLP), sentiment analysis (SA), web scraping (WS) and ML. The expected results comprise synthesized evidence on intervention effectiveness and risk exposures, presented in a graphical and visual format. These results are conveyed in plain language, making them easily understandable by all stakeholders, regardless of their statistical literacy.

The project is continuously updated through automated and on-demand searches, with data from new studies seamlessly integrated into the existing body of evidence. Each inclusion restarts the synthesis process, creating a dynamic, cyclical workflow that ensures the results of the project remain up-to-date (Figure 1).

Figure 1. The Umbrella Collaboration® Workflow.

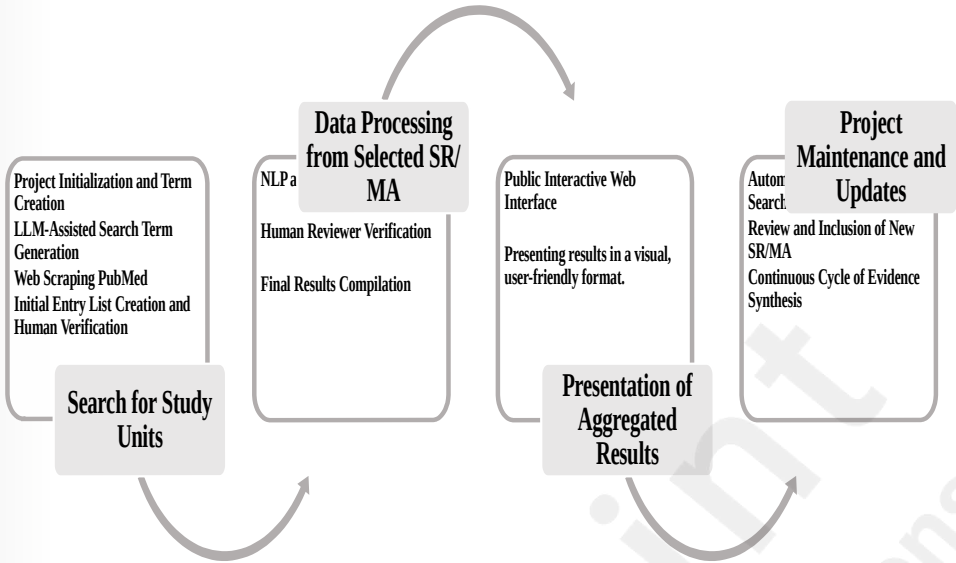


Figure 1 The Umbrella Collaboration® Workflow

The implementation of new methodologies in the scientific field requires a comparative validation process with established methods to ensure their reliability and effectiveness. TU®, being an innovative methodology still in its theoretical-conceptual stage, must be evaluated against established methodologies. Therefore, the aim of this study is to validate TU® by comparing its performance and outcomes with the gold standard, TURs, to establish its credibility and potential superiority.

Objectives

The primary objective of the present study is to assess whether a software-driven AI-assisted system of evidence synthesis, TU®, can match the effectiveness of traditional methods of tertiary synthesis, providing a potentially more timely, efficient, and comprehensive approach while remaining open to findings that could demonstrate superior performance. To support the primary objective of evaluating the effectiveness of TU® compared to traditional methodologies, this study also aims to assess the accessibility and comprehensibility of TU®'s outputs as a secondary objective.

METHODS AND MATERIALS

Study Design

Part 1: Quantitative Comparison of Methodologies

The first part of this study focuses on a quantitative comparison between the two tertiary synthesis methodologies. To facilitate this comparison, a targeted search in PubMed will identify relevant TURs in geriatrics, focusing on representative reviews rather than an exhaustive search. Our approach will involve a focused search in PubMed, using specific terms relevant to geriatrics, to find suitable TURs that serve as a benchmark for this comparative analysis. This targeted search is sufficient for our methodological comparison and does not require the comprehensive search strategy typical of systematic reviews, as our goal is not to cover the entire scope of available literature but to enable a parallel evaluation of synthesis methodologies. Therefore, while the search strategy may appear basic, it is intentionally designed to fulfill the specific needs of our project without aiming for exhaustive literature retrieval, which is beyond the scope of this project.

Study Variables

In the quantitative comparison, several critical variables will be analyzed. Key among these is the identification and evaluation of Outcomes of Interest (outcomes). This includes assessing the degree of concordance between the methodologies in identifying outcomes, using a concordance matrix to document and compare the outcomes identified by TU® and TURs. Additionally, we will analyze the total number of outcomes identified by each methodology, providing a descriptive and statistical comparison. It is essential to define the concept of an "outcome of interest" within the context of tertiary evidence synthesis. An outcome of interest refers to specific aspects identified and evaluated by systematic reviews with meta-analyses (SR/MA) that examine the same research question. These outcomes are critical for understanding the overall impact of various interventions on health conditions or the effects of exposures to risks.

A crucial aspect of this analysis involves the comparison of effect sizes for the identified outcomes. Given the diverse metrics used in TURs—such as Standardized Mean Difference (SMD), Mean Difference (MD), Relative Risk (RR), Odds Ratio (OR), and Hazard Ratio (HR)—and the unique metric employed by TU®, we will standardize all effect sizes to Cohen's d. This standardization facilitates a direct comparison, ensuring consistency in the interpretation of results.

We will also examine the direction of the effects for each outcome, categorizing them as favorable, unfavorable, or unknown for interventions, and as increasing, decreasing, or unknown for exposures

to risks. The statistical significance of the outcomes will be compared by analyzing p-values and confidence intervals, assessing whether the results are statistically significant across both methodologies.

Furthermore, the certainty of the evidence associated with each outcome will be evaluated. For TURs, this will be done using the GRADE system (Grading of Recommendations Assessment, Development and Evaluation) (23) (Guyatt, 2008)²⁷, which categorizes evidence into very low, low, moderate, and high levels. TU® applies a SA-based scoring system on a scale from -1 to +1. Both scales will be normalized to a similar quantitative range (0-1) to facilitate comparison.

Finally, the execution time of each methodology will be assessed, with TU® providing exact time measurements and TURs relying on an estimated timeframe of 6 to 12 months based on existing literature.

Data Collection and Research Question Replication

Data collection will begin with a targeted search in PubMed to identify TURs in geriatrics, using the search terms "umbrella" AND "geriatric". The identified TURs will serve as benchmarks for our comparative analysis. The research questions from these selected umbrella reviews will be directly replicated in TU® without modification, ensuring a precise comparison of outcomes generated by each methodology. TU® will be configured to replicate these questions, employing automated searches and synthesis through NLP, WS, SA and ML, with human reviewer verifying and extracting data as necessary. This approach allows us to assess the comparative effectiveness and efficiency of TU® relative to traditional methods, particularly in identifying and analyzing outcomes critical to evaluating health interventions and exposure risks. Data from both TURs and TU® will be systematically collected and recorded in a database to facilitate precise comparisons of outcomes, effect sizes, and other critical variables, ensuring a thorough evaluation of both methodologies.

Part 2: Evaluation of Ease of Use and Comprehension

The second part of the study focuses on evaluating and comparing the ease of use and comprehension of the results generated by TU® with those from TURs. This evaluation will be conducted through an anonymous and voluntary online survey directed at health professionals, designed to assess their experience with both methodologies.

²⁷ Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, Schünemann HJ; GRADE Working Group. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*. 2008 Apr 26;336(7650):924-6. doi: 10.1136/bmj.39489.470347.AD. PMID: 18436948; PMCID: PMC2335261. <https://www.bmj.com/content/336/7650/924.long>

The survey, developed using Google Survey for ease of access and data analysis, comprises 16 items. The initial 6 items gather demographic information about the survey respondents, while the subsequent 10 questions directly compare the usefulness and clarity of the results produced by TU® and TURs, using a Likert scale ranging from 1 to 5. This scale will measure respondent’s perceptions of the clarity, comprehensibility, and ease of use of the results provided by both methodologies. (Table 1).

To ensure thorough evaluation, informational sessions will be held in geriatric departments of university hospitals in Madrid (Spain). During these sessions, the concept of tertiary evidence synthesis and the functionality of TU® will be introduced. Health professionals, including those from geriatrics and other rotating specialties, will be given access to the TU® platform and guided through its use by an expert. This hands-on experience will be complemented by providing the participants with TUR result tables to facilitate a direct comparison.

Participants will be able to access the survey via a QR code (Figure 2), provided during the sessions, allowing them to complete it either immediately or at their convenience. The survey's responses will be analyzed descriptively, focusing on the overall user experience with TU® and its potential advantages in terms of ease of interpretation and presentation compared to traditional methods.

Table 1. Survey Questions on the Utility of The Umbrella Collaboration®

	Question	Answer Likert Scale
1 - 6	Respondent Affiliation Data	
7	Do you consider that the interactive interface of the "The Umbrella" (TU®) methodology facilitates the understanding of results compared to the static tables of Traditional Umbrella Reviews (URT)?	1: Strongly disagree 2: 3: 4: 5: Strongly agree
8	¿Do you believe that the visualisation of results using bubble plots in "The Umbrella" (TU®) helps to quickly identify the most relevant outcomes?	1: Strongly disagree 2: 3: 4: 5: Strongly agree
9	Does the graphical representation of the outcomes (illustrated by figures, colours, and sizes) in "The Umbrella" (TU®) enhance your ability to assess the clinical relevance of the results?	1: Strongly disagree 2: 3: 4: 5: Strongly agree
10	Do you find that access to detailed data by clicking on the figures of the outcomes in "The Umbrella" (TU®) interface enhances your evidence analysis experience?	1: Strongly disagree 2: 3:

		4: 5: Strongly agree
11	Do you think that "The Umbrella" (TU®) methodology allows for a quicker interpretation of data compared to Traditional Umbrella Reviews (URT)?	1: Strongly disagree 2: 3: 4: 5: Strongly agree
12	Is the information provided by "The Umbrella" (TU®) useful in your field of work (clinical, research, or educational) for evidence-based decision making?	1: Strongly disagree 2: 3: 4: 5: Strongly agree
13	Does the ease of use of "The Umbrella" (TU®) interface facilitate greater data exploration compared to traditional methods?	1: Strongly disagree 2: 3: 4: 5: Strongly agree
14	Do you consider the historical evolution of evidence provided by "The Umbrella" (TU®) methodology to be useful?	1: Strongly disagree 2: 3: 4: 5: Strongly agree
15	Does "The Umbrella" (TU®) methodology require less statistical knowledge to interpret the results compared to Traditional Umbrella Reviews (URT)?	1: Strongly disagree 2: 3: 4: 5: Strongly agree
16	Overall, are you satisfied with "The Umbrella" (TU®) methodology as a tool for tertiary evidence synthesis?	1: Strongly disagree 2: 3: 4: 5: Strongly agree

Figure 2. QR Survey Accessing Code on the Ease of Use and Comprehension of TU®



Statistical Analysis: Data Analysis and Statistical Methods

The quantitative analysis will compare the results obtained by both methodologies. Contingency tables will be constructed to contrast the identification of outcomes, the size effect, the direction of the effect, and the statistical significance, ensuring statistical congruence between the methods. For the evaluation of certainty levels, TU[®] scores will be normalized to a scale comparable to the GRADE levels, which will also be transformed into a numerical scale between 0-1.

The chi-squared test will be used to determine significant differences between the methodologies. Additionally, Pearson and Spearman correlation analyses will be conducted to quantify the relationship between TU[®] certainty levels and GRADE certainty levels of the TURs. Pearson correlation is useful for quantifying the strength and direction of a linear relationship between two continuous variables. Moreover, the analysis will be complemented with Spearman correlation due to potential violations of normality assumptions in the data.

Statistical analysis will be performed using IBM SPSS Statistics 26, with an alpha level of 0.05 to determine statistical significance.

RESULTS

In the results section of this study, we will focus on evaluating the effectiveness of the experimental methodology in accurately identifying and analyzing relevant outcomes. Among the results we will include detailed assessments of effect sizes, the direction of effects, statistical significance, and the certainty of evidence for each outcome. We will compare these findings with those derived from TUR to determine if the experimental approach yields results that are at least equivalent in quality and comprehensiveness.

Finally, the efficiency of the experimental methodology will be evaluated by recording the time taken to complete the synthesis process. While the time required for TUR is estimated to range between 6 to 12 months based on existing literature, we will document the actual time taken by the experimental approach. This will provide a practical measure of the potential time savings offered by the software driven AI-assisted method, highlighting its feasibility and effectiveness in a real-world context.

To evaluate the ease of use and comprehension of TU[®] within its environment, a detailed survey will be specifically designed. Survey respondents will have access to the TU[®] platform, where they can interact with various interactive screens displaying the results of the synthesis process across different projects completed to date.

ETHICAL CONSIDERATIONS

This study does not involve human participants, personal data, or animals. All data will be sourced from published materials, and the analyses will be conducted in accordance with established ethical standards for secondary data analysis. Additionally, the survey component will be conducted anonymously and on a voluntary basis. Given these considerations, we do not deem it necessary to seek approval from an ethics committee, as the study adheres to standard ethical practices for research of this nature.

DISCUSSION

The primary objective of this study is to evaluate whether The Umbrella Collaboration[®] (TU[®]), a software-driven system designed to facilitate tertiary evidence synthesis with AI-assisted methodologies, can match the effectiveness of traditional umbrella reviews (TURs). TU[®] integrates within its software development advanced technologies such as NLP, SA, WS and ML to enhance the efficiency of evidence synthesis.

The semi-automated processes implemented by TU[®] could signify a significant advancement in making evidence synthesis more accessible and timely, with the potential for continuous updates as new data becomes available. The study's findings could pave the way for broader adoption of AI-driven methodologies in evidence synthesis, potentially reducing the time and resources needed for comprehensive reviews. TU[®]'s integration of software engineering project and AI with traditional methods could streamline the review process, enabling faster aggregation and interpretation of data across various research domains.

By offering a platform that supports real-time updates and provides accessible synthesis outputs, TU[®] has the potential to enhance the utility of tertiary synthesis for a wide range of stakeholders, including those with limited statistical expertise. However, a key limitation of TU[®] is its reliance on a single database (Medline via PubMed), which might not capture all relevant studies. This study will

examine whether this limitation can be mitigated by the system’s other capabilities. Ultimately, the timely and comprehensible evidence synthesis provided by TU® could facilitate more informed decision-making in clinical settings, particularly in rapidly evolving areas of medical research.

LIMITATIONS

The decision to utilize only one database, Medline via PubMed, in TU® is both a recognized limitation and a deliberate choice shaped by resource constraints and technical considerations. While systematic reviews typically require searching multiple databases to capture all relevant literature (24) (Gusenbauer, 2020)²⁸, our approach focuses on testing whether TU® can achieve outcomes comparable to TURs despite these limitations. PubMed's strong coverage and the use of LLMs to expand search terms have proven effective, with a pilot study showing that 86% of references from TURs were identified using only PubMed, and of the 511 references included, only 11 were not indexed in PubMed. This highlights the potential of our approach to match traditional methods, although the limitation of using a single database is acknowledged.

Furthermore, while searching multiple databases is often recommended to avoid language and indexing biases, especially those related to non-English literature (25,26) (Jia, 2020)²⁹ (Mao, 2020)³⁰, TU® mitigates some of these biases by focusing on abstracts in English, as all PubMed abstracts are provided in this language regardless of the original publication's language. However, the absence of Chinese databases in our approach is a notable limitation, given that only a small proportion of Chinese journals are indexed in Medline.

CONCLUSION

This study aims to validate TU® as a tool for tertiary evidence synthesis in health. If this methodology proves to be as effective as TURs, but more efficient in terms of project execution time and more accessible in terms of ease of use and comprehension, it could significantly enhance the way evidence synthesis is conducted, facilitating informed decision-making, and improving health outcomes. The results of this study may represent a step towards the integration of innovative technologies into the routine practice of evidence synthesis, with the potential to transform the field

²⁸ Gusenbauer M, Haddaway NR. Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources. *Res Synth Methods*. 2020 Mar;11(2):181-217. doi: 10.1002/jrsm.1378. Epub 2020 Jan 28. PMID: 31614060; PMCID: PMC7079055.

²⁹ Jia Y, Huang D, Wen J, Wang Y, Rosman L, Chen Q, Robinson KA, Gagnier JJ, Ehrhardt S, Celentano DD. Assessment of Language and Indexing Biases Among Chinese-Sponsored Randomized Clinical Trials. *JAMA Netw Open*. 2020 May 1;3(5):e205894. doi: 10.1001/jamanetworkopen.2020.5894. PMID: 32463469; PMCID: PMC7256669.

³⁰ Mao C, Li M. Language Bias Among Chinese-Sponsored Randomized Clinical Trials in Systematic Reviews and Meta-analyses—Can Anything Be Done? *JAMA Netw Open*. 2020;3(5):e206370. doi:10.1001/jamanetworkopen.2020.6370

of health research.

FUNDING

This project has received a research grant from the *Fundación Alfonso X el Sabio*.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

REFERENCES:

1. Evidence synthesis for policy A STATEMENT OF PRINCIPLES [Internet]. The Royal Society of Evidence Synthesis; 2018. Available from: <https://royalsociety.org/-/media/policy/projects/evidence-synthesis/evidence-synthesis-statement-principles.pdf>

2. World Health Organization, 2005. bridging-the-know-do-gap.pdf [Internet]. [cited 2024 Jun 2]. Available from: <https://www.measureevaluation.org/resources/training/capacity-building-resources/high-impact-research-training-curricula/bridging-the-know-do-gap.pdf>

3. Cottrell E, Whitlock E, Kato E, Uhl S, Belinson S, Chang C, et al. Defining the Benefits of Stakeholder Engagement in Systematic Reviews [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); 2014 [cited 2024 May 26]. (AHRQ Methods for Effective Health Care). Available from: <http://www.ncbi.nlm.nih.gov/books/NBK196180/>

4. Sørensen K, Van den Broucke S, Fullam J, Doyle G, Pelikan J, Slonska Z, et al. Health literacy and public health: A systematic review and integration of definitions and models. BMC Public Health [Internet]. 2012 Jan 25 [cited 2024 May 26];12(1):80. Available from: <https://doi.org/10.1186/1471-2458-12-80>

5. Martínez-García JJ, Canizalez-Román A, Velázquez-Román JA, Flores-Villaseñor HM, León-Sicaíros NM. Evaluación del conocimiento de métodos básicos de epidemiología e investigación en médicos residentes. Rev Médica Univ Autónoma Sinaloa REVMEDUAS [Internet]. 2022 Jul 13 [cited 2024 May 26];11(2):115–23. Available from: <https://www.medigraphic.com/cgi-bin/new/resumen.cgi?IDARTICULO=106198>

6. Baccolini V, Rosso A, Di Paolo C, Isonne C, Salerno C, Migliara G, et al. What is the Prevalence of Low Health Literacy in European Union Member States? A Systematic Review and Meta-analysis. J Gen Intern Med [Internet]. 2021 Mar 1 [cited 2024 May 26];36(3):753–61. Available from: <https://doi.org/10.1007/s11606-020-06407-8>

7. Baumann LA, Reinhold AK, Brütt AL. Public and patient involvement in health policy decision-making on the health system level – A scoping review. Health Policy [Internet]. 2022 Oct 1 [cited 2024 May 26];126(10):1023–38. Available from: <https://www.sciencedirect.com/science/article/pii/S0168851022001919>

8. Biondi-Zoccai G, editor. Umbrella Reviews: Evidence Synthesis with Overviews of Reviews and Meta-Epidemiologic Studies [Internet]. Cham: Springer International Publishing; 2016 [cited 2024 May 26]. Available from: <http://link.springer.com/10.1007/978-3-319-25655-9>

9. Choi GJ, Kang H. Introduction to Umbrella Reviews as a Useful Evidence-Based Practice. J Lipid Atheroscler [Internet]. 2023 Jan [cited 2024 May 26];12(1):3–11. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9884555/>

10. Fusar-Poli P, Radua J. Ten simple rules for conducting umbrella reviews. BMJ Ment Health [Internet]. 2018 Aug 1 [cited 2024 May 26];21(3):95–100. Available from: <https://mentalhealth.bmj.com/content/21/3/95>

11. Aromataris E, Lockwood C, Porritt K, Pilla B. JBI Manual for Evidence Synthesis - JBI Global Wiki [Internet]. [cited 2024 Jun 5]. Available from: <https://jbi-global-wiki.refined.site/space/MANUAL>

12. Cant R, Ryan C, Kelly MA. A nine-step pathway to conduct an umbrella review of literature. Nurse Author Ed [Internet]. 2022 [cited 2024 May 26];32(2):31–4. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/nae2.12039>

13. Belbasis L, Brooker RD, Zavalis E, Pezzullo AM, Axfors C, Ioannidis JP. Mapping and systematic appraisal of umbrella reviews in epidemiological research: a protocol for a meta-epidemiological study. Syst Rev [Internet]. 2023 Jul 14 [cited 2024 May 26];12(1):123. Available from: <https://doi.org/10.1186/s13643-023-02265-7>

14. Pollock M. Chapter V: Overviews of Reviews [Internet]. 2023 [cited 2024 May 26]. Available

from: <https://training.cochrane.org/handbook/current/chapter-v>

15. Tricco AC, Straus SE, Ghaffar A, Langlois EV. Rapid reviews for health policy and systems decision-making: more important than ever before. Syst Rev [Internet]. 2022 Jul 30 [cited 2024 Jun 5];11(1):153. Available from: <https://doi.org/10.1186/s13643-022-01887-7>

16. Qureshi R, Shaughnessy D, Gill KAR, Robinson KA, Li T, Agai E. Are ChatGPT and large language models “the answer” to bringing us closer to systematic review automation? Syst Rev [Internet]. 2023 Apr 29 [cited 2024 Aug 17];12(1):72. Available from: <https://doi.org/10.1186/s13643-023-02243-z>

17. Teperikidis L, Boulmpou A, Papadopoulos C, Biondi-Zoccai G. Using ChatGPT to perform a systematic review: a tutorial. Minerva Cardiol Angiol. 2024 Jul 26;

18. ChatGPT [Internet]. [cited 2024 Sep 7]. Available from: <https://openai.com/chatgpt/>

19. Gwon YN, Kim JH, Chung HS, Jung EJ, Chun J, Lee S, et al. The Use of Generative AI for Scientific Literature Searches for Systematic Reviews: ChatGPT and Microsoft Bing AI Performance Evaluation. JMIR Med Inform [Internet]. 2024 May 14 [cited 2024 Aug 15];12:51187. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11107769/>

20. Kothamali PR, Karne VK, Dandiyala SSM. Integrating AI and Machine Learning in Quality Assurance for Automation Engineering. Int J Res Publ Semin [Internet]. 2024 Jul 25 [cited 2024 Aug 15];15(3):93–102. Available from: <https://jrps.shodhsagar.com/index.php/j/article/view/1445>

21. Elliott JH, Synnot A, Turner T, Simmonds M, Akl EA, McDonald S, et al. Living systematic review: 1. Introduction-the why, what, when, and how. J Clin Epidemiol. 2017 Nov;91:23–30.

22. 201912_LSR_Revised_Guidance.pdf [Internet]. [cited 2024 Aug 15]. Available from: https://community.cochrane.org/sites/default/files/uploads/inline-files/Transform/201912_LSR_Revised_Guidance.pdf

23. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ [Internet]. 2008 Apr 24 [cited 2024 Jun 6];336(7650):924–6. Available from: <https://www.bmj.com/content/336/7650/924>

24. Gusenbauer M, Haddaway NR. Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources. Res Synth Methods [Internet]. 2020 Mar [cited 2024 Aug 16];11(2):181–217. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7079055/>

25. Jia Y, Huang D, Wen J, Wang Y, Rosman L, Chen Q, et al. Assessment of Language and Indexing Biases Among Chinese-Sponsored Randomized Clinical Trials. JAMA Netw Open [Internet]. 2020 May 28 [cited 2024 Aug 16];3(5):e205894. Available from: <https://doi.org/10.1001/jamanetworkopen.2020.5894>

26. Mao C, Li M. Language Bias Among Chinese-Sponsored Randomized Clinical Trials in Systematic Reviews and Meta-analyses—Can Anything Be Done? JAMA Netw Open [Internet]. 2020 May 28 [cited 2024 Aug 16];3(5):e206370. Available from: <https://doi.org/10.1001/jamanetworkopen.2020.6370>

Supplementary Files

Untitled.

Table 1. Survey Questions on the Utility of The Umbrella Collaboration*

Question	Answer Likert Scale
1-6 Respondent Affiliation Data	
7 Do you consider that the interactive interface of the "The Umbrella" (TU) methodology facilitates the understanding of results compared to the static tables of Traditional Umbrella Reviews (URT)?	1: Strongly disagree 2: 3: 4: 5: Strongly agree
8 (Do you believe that the visualisation of results using bubble plots in "The Umbrella" (TU) helps to quickly identify the most relevant outcomes?	1: Strongly disagree 2: 3: 4: 5: Strongly agree
9 Does the graphical representation of the outcomes (illustrated by figures, colours, and sizes) in "The Umbrella" (TU) enhance your ability to assess the clinical relevance of the results?	1: Strongly disagree 2: 3: 4: 5: Strongly agree
10 Do you find that access to detailed data by clicking on the figures of the outcomes in "The Umbrella" (TU) interface enhances your evidence analysis experience?	1: Strongly disagree 2: 3: 4: 5: Strongly agree
11 Do you think that "The Umbrella" (TU) methodology allows for a quicker interpretation of data compared to Traditional Umbrella Reviews (URT)?	1: Strongly disagree 2: 3: 4: 5: Strongly agree
12 Is the information provided by "The Umbrella" (TU) useful in your field of work (clinical, research, or educational) for evidence-based decision making?	1: Strongly disagree 2: 3: 4: 5: Strongly agree
13 Does the ease of use of "The Umbrella" (TU) interface facilitate greater data exploration compared to traditional methods?	1: Strongly disagree 2: 3: 4: 5: Strongly agree
14 Do you consider the historical evolution of evidence provided by "The Umbrella" (TU) methodology to be useful?	1: Strongly disagree 2: 3: 4: 5: Strongly agree
15 Does "The Umbrella" (TU) methodology require less statistical knowledge to interpret the results compared to Traditional Umbrella Reviews (URT)?	1: Strongly disagree 2: 3:

Untitled.

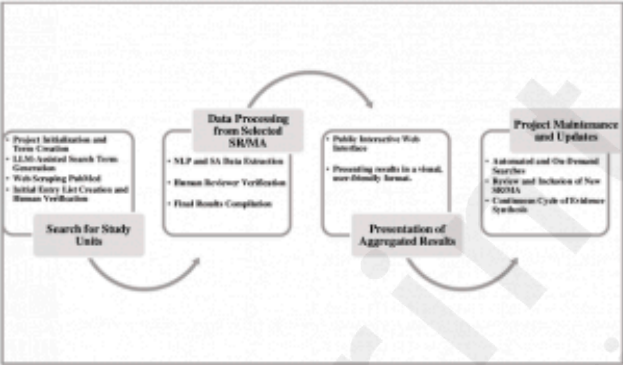
Table 1. Survey Questions on the Utility of The Umbrella Collaboration*

Question	Answer Likert Scale
1-6 Respondent Affiliation Data	
7 Do you consider that the interactive interface of the "The Umbrella" (TU) methodology facilitates the understanding of results compared to the static tables of Traditional Umbrella Reviews (URT)?	1: Strongly disagree 2: 3: 4: 5: Strongly agree
8 (Do you believe that the visualisation of results using bubble plots in "The Umbrella" (TU) helps to quickly identify the most relevant outcomes?	1: Strongly disagree 2: 3: 4: 5: Strongly agree
9 Does the graphical representation of the outcomes (illustrated by figures, colours, and sizes) in "The Umbrella" (TU) enhance your ability to assess the clinical relevance of the results?	1: Strongly disagree 2: 3: 4: 5: Strongly agree
10 Do you find that access to detailed data by clicking on the figures of the outcomes in "The Umbrella" (TU) interface enhances your evidence analysis experience?	1: Strongly disagree 2: 3: 4: 5: Strongly agree
11 Do you think that "The Umbrella" (TU) methodology allows for a quicker interpretation of data compared to Traditional Umbrella Reviews (URT)?	1: Strongly disagree 2: 3: 4: 5: Strongly agree
12 Is the information provided by "The Umbrella" (TU) useful in your field of work (clinical, research, or educational) for evidence-based decision making?	1: Strongly disagree 2: 3: 4: 5: Strongly agree
13 Does the ease of use of "The Umbrella" (TU) interface facilitate greater data exploration compared to traditional methods?	1: Strongly disagree 2: 3: 4: 5: Strongly agree
14 Do you consider the historical evolution of evidence provided by "The Umbrella" (TU) methodology to be useful?	1: Strongly disagree 2: 3: 4: 5: Strongly agree
15 Does "The Umbrella" (TU) methodology require less statistical knowledge to interpret the results compared to Traditional Umbrella Reviews (URT)?	1: Strongly disagree 2: 3:

Figures

The Umbrella Collaboration Workflow.

Figure 1. The Umbrella Collaboration® Workflow.



10

BIBLIOGRAFÍA

1. PLoS Medicine Editors. Focusing the Spotlight on Lack of Access to Health Information. *PLoS Med* [Internet]. 2013 Apr 30 [cited 2024 May 26];10(4):e1001438. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3640082/>
2. Portal [Internet]. [cited 2024 Oct 4]. The Council of Europe: guardian of Human Rights, Democracy and the Rule of Law for 700 million citizens - Portal - www.coe.int. Available from: <https://www.coe.int/es/web/portal>
3. Land MK, Pakenham-Walsh N. Access to Health Information Under International Human Rights Law [Internet]. Rochester, NY; 2012 [cited 2024 May 26]. Available from: <https://papers.ssrn.com/abstract=3670747>
4. Evidence-informed Policy Network (EVIPNet) Europe [Internet]. [cited 2024 May 26]. Available from: [https://www.who.int/europe/initiatives/evidence-informed-policy-network-\(evipnet\)-europe](https://www.who.int/europe/initiatives/evidence-informed-policy-network-(evipnet)-europe)
5. World Health Organization, 2005. bridging-the-know-do-gap.pdf [Internet]. [cited 2024 Jun 2]. Available from: <https://www.measureevaluation.org/resources/training/capacity-building-resources/high-impact-research-training-curricula/bridging-the-know-do-gap.pdf>
6. Straus SE, Tetroe J, Graham I. Defining knowledge translation. *CMAJ Can Med Assoc J* [Internet]. 2009 Aug 4 [cited 2024 Nov 23];181(3–4):165–8. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2717660/>
7. Hu JCH. Access to health knowledge for health equality: a multi-phase review focused on disability-health. *Int J Equity Health* [Internet]. 2024 Jan 10 [cited 2024 May 26];23:6. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10777648/>
8. Kitson A, Straus SE. The knowledge-to-action cycle: identifying the gaps. *CMAJ Can Med Assoc J* [Internet]. 2010 Feb 9 [cited 2024 May 26];182(2):E73–7. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2817340/>
9. Cottrell E, Whitlock E, Kato E, Uhl S, Belinson S, Chang C, et al. Defining the Benefits of Stakeholder Engagement in Systematic Reviews [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); 2014 [cited 2024 May 26]. (AHRQ Methods for Effective Health Care). Available from: <http://www.ncbi.nlm.nih.gov/books/NBK196180/>
10. Government of Canada CI of HR. Knowledge User Engagement - CIHR [Internet]. 2012 [cited 2024 May 26]. Available from: <https://www.cihr-irsc.gc.ca/e/49505.html>
11. Yu X, Wu S, Sun Y, Wang P, Wang L, Su R, et al. Exploring the diverse definitions of ‘evidence’: a scoping review. *BMJ Evid-Based Med* [Internet]. 2024 Feb 1 [cited 2024 May 26];29(1):37–43. Available from: <https://ebm.bmj.com/content/29/1/37>
12. Blunt CJ. The origins of evidence pyramids [Internet]. CJ Blunt. 2022 [cited 2024 Oct 6]. Available from: <http://cjblunt.com/the-origins-of-evidence-pyramids/>
13. Murad MH, Asi N, Alsawas M, Alahdab F. New evidence pyramid. *BMJ Evid-Based Med* [Internet]. 2016 Aug 1 [cited 2024 Nov 23];21(4):125–7. Available from: <https://ebm.bmj.com/content/21/4/125>
14. Gugnani N. Real world evidence: will the “pyramid” of evidence need some redefining...? *Evid Based Dent* [Internet]. 2024 Sep [cited 2024 Oct 6];25(3):119–20. Available from: <https://www.nature.com/articles/s41432-024-01035-1>
15. evidence-synthesis-statement-principles.pdf [Internet]. [cited 2024 Sep 6]. Available from: <https://royalsociety.org/-/media/policy/projects/evidence-synthesis/evidence-synthesis-statement-principles.pdf>
16. Fusar-Poli P, Radua J. Ten simple rules for conducting umbrella reviews. *BMJ Ment Health* [Internet]. 2018 Aug 1 [cited 2024 May 26];21(3):95–100. Available from: <https://mentalhealth.bmj.com/content/21/3/95>

17. Choi GJ, Kang H. Introduction to Umbrella Reviews as a Useful Evidence-Based Practice. *J Lipid Atheroscler* [Internet]. 2023 Jan [cited 2024 May 26];12(1):3–11. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9884555/>
18. Gough D, Davies P, Jantvedt G, Langlois E, Littell J, Lotfi T, et al. Evidence Synthesis International (ESI): Position Statement. *Syst Rev* [Internet]. 2020 Jul 10 [cited 2024 May 26];9(1):155. Available from: <https://doi.org/10.1186/s13643-020-01415-5>
19. Cochrane Handbook for Systematic Reviews of Interventions [Internet]. [cited 2024 Oct 4]. Available from: <https://training.cochrane.org/handbook/current>
20. Page MJ, Shamseer L, Altman DG, Tetzlaff J, Sampson M, Tricco AC, et al. Epidemiology and Reporting Characteristics of Systematic Reviews of Biomedical Research: A Cross-Sectional Study. *PLOS Med* [Internet]. 2016 May 24 [cited 2024 Nov 23];13(5):e1002028. Available from: <https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1002028>
21. Alabousi M, Alabousi A, McGrath TA, Cobey KD, Budhram B, Frank RA, et al. Epidemiology of systematic reviews in imaging journals: evaluation of publication trends and sustainability? *Eur Radiol* [Internet]. 2019 Feb 1 [cited 2024 Nov 23];29(2):517–26. Available from: <https://doi.org/10.1007/s00330-018-5567-z>
22. Ioannidis JP a. The Mass Production of Redundant, Misleading, and Conflicted Systematic Reviews and Meta-analyses. *Milbank Q* [Internet]. 2016 [cited 2024 May 26];94(3):485–514. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/1468-0009.12210>
23. Ahn E, Kang H. Introduction to systematic review and meta-analysis. *Korean J Anesthesiol* [Internet]. 2018 Apr [cited 2024 Nov 23];71(2):103–12. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5903119/>
24. Vetter TR. Systematic Review and Meta-analysis: Sometimes Bigger Is Indeed Better. *Anesth Analg* [Internet]. 2019 Mar [cited 2024 Nov 23];128(3):575. Available from: https://journals.lww.com/anesthesia-analgesia/fulltext/2019/03000/systematic_review_and_meta_analysis_sometimes.25.aspx
25. Phillips V, Barker E. Systematic reviews: Structure, form and content. *J Perioper Pract* [Internet]. 2021 Sep 1 [cited 2024 Nov 23];31(9):349–53. Available from: <https://doi.org/10.1177/1750458921994693>
26. Fogarty AE, Conger A, Burnham T. Systematic reviews: Not always a pain. *Interv Pain Med* [Internet]. 2022 Jan 1 [cited 2024 Nov 23];1:100128. Available from: <https://www.sciencedirect.com/science/article/pii/S2772594422001248>
27. Chapter I: Introduction [Internet]. [cited 2024 Nov 23]. Available from: <https://training.cochrane.org/handbook/current/chapter-i>
28. Home Page | JBI [Internet]. [cited 2024 Nov 23]. Available from: <https://jbi.global/>
29. Campbell Collaboration [Internet]. [cited 2024 Nov 23]. Available from: <https://www.campbellcollaboration.org/>
30. Borah R, Brown AW, Capers PL, Kaiser KA. Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. *BMJ Open* [Internet]. 2017 Feb 1 [cited 2024 May 26];7(2):e012545. Available from: <https://bmjopen.bmj.com/content/7/2/e012545>
31. Garritty C, Gartlehner G, Nussbaumer-Streit B, King VJ, Hamel C, Kamel C, et al. Cochrane Rapid Reviews Methods Group offers evidence-informed guidance to conduct rapid reviews. *J Clin Epidemiol* [Internet]. 2021 Feb 1 [cited 2024 May 26];130:13–22. Available from: [https://www.jclinepi.com/article/S0895-4356\(20\)31146-X/fulltext](https://www.jclinepi.com/article/S0895-4356(20)31146-X/fulltext)
32. Boisen ML, Hartnett JN, Goba A, Vandi MA, Grant DS, Schieffelin JS, et al. Epidemiology and Management of the 2013–16 West African Ebola Outbreak. *Annu Rev Virol* [Internet].

- 2016 Sep 29 [cited 2024 May 26];3(Volume 3, 2016):147–71. Available from: <https://www.annualreviews.org/content/journals/10.1146/annurev-virology-110615-040056>
33. Sampson M, Shojania KG, Garritty C, Horsley T, Ocampo M, Moher D. Systematic reviews can be produced and published faster. *J Clin Epidemiol* [Internet]. 2008 Jun 1 [cited 2024 Nov 23];61(6):531–6. Available from: [https://www.jclinepi.com/article/S0895-4356\(08\)00053-X/abstract](https://www.jclinepi.com/article/S0895-4356(08)00053-X/abstract)
 34. Clark J, Glasziou P, Mar CD, Bannach-Brown A, Stehlik P, Scott AM. A full systematic review was completed in 2 weeks using automation tools: a case study. *J Clin Epidemiol* [Internet]. 2020 May 1 [cited 2024 Nov 23];121:81–90. Available from: [https://www.jclinepi.com/article/S0895-4356\(19\)30719-X/fulltext](https://www.jclinepi.com/article/S0895-4356(19)30719-X/fulltext)
 35. Scott AM, Glasziou P, Clark J. We extended the 2-week systematic review (2weekSR) methodology to larger, more complex systematic reviews: A case series. *J Clin Epidemiol* [Internet]. 2023 May 1 [cited 2024 Nov 23];157:112–9. Available from: [https://www.jclinepi.com/article/S0895-4356\(23\)00050-1/fulltext](https://www.jclinepi.com/article/S0895-4356(23)00050-1/fulltext)
 36. Tufanaru C, Surian D, Scott AM, Glasziou P, Coiera E. The 2-week systematic review (2weekSR) method was successfully blind-replicated by another team: a case study. *J Clin Epidemiol* [Internet]. 2024 Jan 1 [cited 2024 Nov 23];165. Available from: [https://www.jclinepi.com/article/S0895-4356\(23\)00273-1/fulltext](https://www.jclinepi.com/article/S0895-4356(23)00273-1/fulltext)
 37. Uttley L, Quintana DS, Montgomery P, Carroll C, Page MJ, Falzon L, et al. The problems with systematic reviews: a living systematic review. *J Clin Epidemiol* [Internet]. 2023 Apr 1 [cited 2024 May 27];156:30–41. Available from: <https://www.sciencedirect.com/science/article/pii/S0895435623000112>
 38. Garner P, Hopewell S, Chandler J, MacLehose H, Akl EA, Beyene J, et al. When and how to update systematic reviews: consensus and checklist. *BMJ* [Internet]. 2016 Jul 20 [cited 2024 May 26];354:i3507. Available from: <https://www.bmj.com/content/354/bmj.i3507>
 39. Hoffmeyer B, Fonnes S, Andresen K, Rosenberg J. Use of inactive Cochrane reviews in academia: A citation analysis. *Scientometrics* [Internet]. 2023 May 1 [cited 2024 May 26];128(5):2923–34. Available from: <https://doi.org/10.1007/s11192-023-04691-9>
 40. Beller EM, Chen JKH, Wang ULH, Glasziou PP. Are systematic reviews up-to-date at the time of publication? *Syst Rev* [Internet]. 2013 May 28 [cited 2024 May 26];2(1):36. Available from: <https://doi.org/10.1186/2046-4053-2-36>
 41. Bashir R, Surian D, Dunn AG. Time-to-update of systematic reviews relative to the availability of new evidence. *Syst Rev* [Internet]. 2018 Nov 17 [cited 2024 May 26];7(1):195. Available from: <https://doi.org/10.1186/s13643-018-0856-9>
 42. Shojania KG, Sampson M, Ansari MT, Ji J, Doucette S, Moher D. How Quickly Do Systematic Reviews Go Out of Date? A Survival Analysis. *Ann Intern Med* [Internet]. 2007 Aug 21 [cited 2024 Nov 23];147(4):224–33. Available from: <https://www.acpjournals.org/doi/10.7326/0003-4819-147-4-200708210-00179>
 43. Sørensen K, Van den Broucke S, Fullam J, Doyle G, Pelikan J, Slonska Z, et al. Health literacy and public health: A systematic review and integration of definitions and models. *BMC Public Health* [Internet]. 2012 Jan 25 [cited 2024 May 26];12(1):80. Available from: <https://doi.org/10.1186/1471-2458-12-80>
 44. Windish DM, Huot SJ, Green ML. Medicine Residents' Understanding of the Biostatistics and Results in the Medical Literature. *JAMA* [Internet]. 2007 Sep 5 [cited 2024 May 26];298(9):1010–22. Available from: <https://doi.org/10.1001/jama.298.9.1010>
 45. Best AM, Laskin DM. Oral and maxillofacial surgery residents have poor understanding of biostatistics. *J Oral Maxillofac Surg Off J Am Assoc Oral Maxillofac Surg*. 2013 Jan;71(1):227–34.
 46. Bookstaver PB, Miller AD, Felder TM, Tice DL, Norris LB, Sutton SS. Assessing pharmacy residents' knowledge of biostatistics and research study design. *Ann Pharmacother*. 2012;46(7–8):991–9.
 47. Bougie O, Posner G, Black AY. Critical Appraisal Skills Among Canadian Obstetrics and Gynaecology Residents: How Do They Fare? *J Obstet Gynaecol Can JOGC J Obstet Gynecol Can JOGC*. 2015 Jul;37(7):639–47.
 48. Martínez-García JJ, Canizalez-Román A, Velázquez-Román JA, Flores-Villaseñor HM, León-Sicaire NM. Evaluación del conocimiento de métodos básicos de epidemiología e investigación en médicos residentes. *Rev Médica Univ Autónoma Sinaloa REVMEDUAS* [Internet]. 2022 Jul 13 [cited 2024 May 26];11(2):115–23. Available from: <https://www.medigraphic.com/cgi-bin/new/resumen.cgi?IDARTICULO=106198>
 49. Hoffrage U, Gigerenzer G. Using natural frequencies to improve diagnostic inferences. *Acad Med J Assoc Am Med Coll*. 1998 May;73(5):538–40.
 50. Labarge AS, McCaffrey RJ, Brown TA. Neuropsychologists' abilities to determine the predictive value of diagnostic tests. *Arch Clin Neuropsychol Off J Natl Acad Neuropsychol*. 2003 Mar;18(2):165–75.
 51. Bramwell R, West H, Salmon P. Health professionals' and service users' interpretation of screening test results: experimental study. *BMJ* [Internet]. 2006 Aug 5 [cited 2024 Nov 23];333(7562):284. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1526944/>
 52. Wegwarth O, Schwartz LM, Woloshin S, Gaissmaier W, Gigerenzer G. Do Physicians Understand Cancer Screening Statistics? A National Survey of Primary Care Physicians in the United States. *Ann Intern Med* [Internet]. 2012 Mar 6 [cited 2024 Nov 23];156(5):340–9. Available from: <https://www.acpjournals.org/doi/10.7326/0003-4819-156-5-201203060-00005>
 53. Lakhli C, Lejeune FX, Rouault M, Khamassi M, Rohaut B. Illusion of knowledge in statistics among clinicians: evaluating the alignment between objective accuracy and subjective confidence, an online survey. *Cogn Res Princ Implic* [Internet]. 2023 Apr 20 [cited 2024 Nov 23];8:23. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10118231/>
 54. Lopez C, Kim B, Sacks K. Health Literacy in the United States: Enhancing Assessments and Reducing Disparities. *SSRN Electron J* [Internet]. 2022 [cited 2024 May 26]; Available from: <https://www.ssrn.com/abstract=4182046>
 55. Baccolini V, Rosso A, Di Paolo C, Isonne C, Salerno C, Migliara G, et al. What is the Prevalence of Low Health Literacy in European Union Member States? A Systematic Review and Meta-analysis. *J Gen Intern Med* [Internet]. 2021 Mar 1 [cited 2024 May 26];36(3):753–61. Available from: <https://doi.org/10.1007/s11606-020-06407-8>
 56. ahdat S, Hamzehgardeshi L, Hessam S, Hamzehgardeshi Z. Patient Involvement in Health Care Decision Making: A Review. *Iran Red Crescent Med J* [Internet]. 2014 Jan [cited 2024 May 26];16(1):e12454. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3964421/>
 57. Krist AH, Tong ST, Aycock RA, Longo DR. Engaging patients in decision-making and behavior change to promote prevention. *Stud Health Technol Inform* [Internet]. 2017 [cited 2024 May 26];240:284–302. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6996004/>
 58. Baumann LA, Reinhold AK, Brütt AL. Public and patient involvement in health policy decision-making on the health system level – A scoping review. *Health Policy* [Internet]. 2022 Oct 1 [cited 2024 May 26];126(10):1023–38. Available from: <https://www.sciencedirect.com/science/article/pii/S0168851022001919>
 59. Siontis KC, Hernandez-Boussard T, Ioannidis JPA. Overlapping meta-analyses on the same topic: survey of published studies. *The BMJ* [Internet]. 2013 Jul 19 [cited 2024 May 26];347:f4501. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3716360/>

60. Shrier I, Boivin JF, Platt RW, Steele RJ, Brophy JM, Carnevale F, et al. The interpretation of systematic reviews with meta-analyses: an objective or subjective process? *BMC Med Inform Decis Mak* [Internet]. 2008 May 21 [cited 2024 Nov 23];8:19. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2408567/>
61. Valentine JC, Cooper H, Patall EA, Tyson D, Robinson JC. A method for evaluating research syntheses: The quality, conclusions, and consensus of 12 syntheses of the effects of after-school programs. *Res Synth Methods* [Internet]. 2010 [cited 2024 May 26];1(1):20–38. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/jrsm.3>
62. Mueller M, D'Addario M, Egger M, Cevallos M, Dekkers O, Mugglin C, et al. Methods to systematically review and meta-analyse observational studies: a systematic scoping review of recommendations. *BMC Med Res Methodol* [Internet]. 2018 May 21 [cited 2024 May 26];18(1):44. Available from: <https://doi.org/10.1186/s12874-018-0495-9>
63. Hamel C, Michaud A, Thuku M, Skidmore B, Stevens A, Nussbaumer-Streit B, et al. Defining Rapid Reviews: a systematic scoping review and thematic analysis of definitions and defining characteristics of rapid reviews. *J Clin Epidemiol* [Internet]. 2021 Jan 1 [cited 2024 Nov 23];129:74–85. Available from: [https://www.jclinepi.com/article/S0895-4356\(20\)31127-6/abstract](https://www.jclinepi.com/article/S0895-4356(20)31127-6/abstract)
64. Rapid reviews to strengthen health policy and systems: a practical guide [Internet]. [cited 2024 Jun 5]. Available from: <https://ahpsr.who.int/publications/i/item/2017-08-10-rapid-reviews-to-strengthen-health-policy-and-systems-a-practical-guide>
65. Tricco AC, Straus SE, Ghaffar A, Langlois EV. Rapid reviews for health policy and systems decision-making: more important than ever before. *Syst Rev* [Internet]. 2022 Jul 30 [cited 2024 Jun 5];11(1):153. Available from: <https://doi.org/10.1186/s13643-022-01887-7>
66. Booth A, Sommer I, Noyes J, Houghton C, Campbell F. Rapid reviews methods series: guidance on rapid qualitative evidence synthesis. *BMJ Evid-Based Med* [Internet]. 2024 Jun 1 [cited 2024 Nov 23];29(3):194–200. Available from: <https://ebm.bmj.com/content/29/3/194>
67. Klerings I, Robalino S, Booth A, Escobar-Liquitay CM, Sommer I, Gartlehner G, et al. Rapid reviews methods series: Guidance on literature search. *BMJ Evid-Based Med* [Internet]. 2023 Dec 1 [cited 2024 Nov 23];28(6):412–7. Available from: <https://ebm.bmj.com/content/28/6/412>
68. Nussbaumer-Streit B, Sommer I, Hamel C, Devane D, Noel-Storr A, Puljak L, et al. Rapid reviews methods series: Guidance on team considerations, study selection, data extraction and risk of bias assessment. *BMJ Evid-Based Med* [Internet]. 2023 Dec 1 [cited 2024 Nov 23];28(6):418–23. Available from: <https://ebm.bmj.com/content/28/6/418>
69. Garritty C, Tricco AC, Smith M, Pollock D, Kamel C, King VJ. Rapid Reviews Methods Series: Involving patient and public partners, healthcare providers and policymakers as knowledge users. *BMJ Evid-Based Med* [Internet]. 2024 Feb 1 [cited 2024 Nov 23];29(1):55–61. Available from: <https://ebm.bmj.com/content/29/1/55>
70. Gartlehner G, Kahwati L, Hilscher R, Thomas I, Kugley S, Crotty K, et al. Data extraction for evidence synthesis using a large language model: A proof-of-concept study. *Res Synth Methods* [Internet]. 2024 [cited 2024 Oct 4];15(4):576–89. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/jrsm.1710>
71. King VJ, Stevens A, Nussbaumer-Streit B, Kamel C, Garritty C. Paper 2: Performing rapid reviews. *Syst Rev* [Internet]. 2022 Jul 30 [cited 2024 Nov 23];11:151. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9338520/>
72. Marshall IJ, Marshall R, Wallace BC, Brassey J, Thomas J. Rapid reviews may produce different results to systematic reviews: a meta-epidemiological study. *J Clin Epidemiol* [Internet]. 2019 May 1 [cited 2024 Nov 23];109:30–41. Available from: [https://www.jclinepi.com/article/S0895-4356\(18\)30893-X/fulltext](https://www.jclinepi.com/article/S0895-4356(18)30893-X/fulltext)

73. Wagner G, Nussbaumer-Streit B, Greimel J, Ciapponi A, Gartlehner G. Trading certainty for speed - how much uncertainty are decisionmakers and guideline developers willing to accept when using rapid reviews: an international survey. *BMC Med Res Methodol* [Internet]. 2017 Aug 14 [cited 2024 Nov 23];17(1):121. Available from: <https://doi.org/10.1186/s12874-017-0406-5>
74. 201912_LSR_Revised_Guidance.pdf [Internet]. [cited 2024 Aug 15]. Available from: https://community.cochrane.org/sites/default/files/uploads/inline-files/Transform/201912_LSR_Revised_Guidance.pdf
75. Lansky A, Wethington HR. Living Systematic Reviews and Other Approaches for Updating Evidence. *Am J Public Health* [Internet]. 2020 Nov [cited 2024 Oct 5];110(11):1687–8. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7542271/>
76. Simmonds M, Elliott JH, Synnot A, Turner T. Living Systematic ReviewsSystematic reviews. In: Evangelou E, Veroniki AA, editors. *Meta-Research: Methods and Protocols* [Internet]. New York, NY: Springer US; 2022 [cited 2024 Nov 23]. p. 121–34. Available from: https://doi.org/10.1007/978-1-0716-1566-9_7
77. Millard T, Synnot A, Elliott J, Green S, McDonald S, Turner T. Feasibility and acceptability of living systematic reviews: results from a mixed-methods evaluation. *Syst Rev* [Internet]. 2019 Dec 14 [cited 2024 Nov 23];8(1):325. Available from: <https://doi.org/10.1186/s13643-019-1248-5>
78. Tsafnat G, Glasziou P, Choong MK, Dunn A, Galgani F, Coiera E. Systematic review automation technologies. *Syst Rev* [Internet]. 2014 Jul 9 [cited 2024 Nov 23];3(1):74. Available from: <https://doi.org/10.1186/2046-4053-3-74>
79. Scott AM, Forbes C, Clark J, Carter M, Glasziou P, Munn Z. Systematic review automation tools improve efficiency but lack of knowledge impedes their adoption: a survey. *J Clin Epidemiol* [Internet]. 2021 Oct 1 [cited 2024 Nov 23];138:80–94. Available from: [https://www.jclinepi.com/article/S0895-4356\(21\)00208-0/abstract](https://www.jclinepi.com/article/S0895-4356(21)00208-0/abstract)
80. Affengruber L, Nussbaumer-Streit B, Hamel C, Maten MV der, Thomas J, Mavergames C, et al. Rapid review methods series: Guidance on the use of supportive software. *BMJ Evid-Based Med* [Internet]. 2024 Aug 1 [cited 2024 Nov 23];29(4):264–71. Available from: <https://ebm.bmj.com/content/29/4/264>
81. Gwon YN, Kim JH, Chung HS, Jung EJ, Chun J, Lee S, et al. The Use of Generative AI for Scientific Literature Searches for Systematic Reviews: ChatGPT and Microsoft Bing AI Performance Evaluation. *JMIR Med Inform* [Internet]. 2024 May 14 [cited 2024 Aug 15];12:51187. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11107769/>
82. Haddaway NR, Callaghan MW, Collins AM, Lamb WF, Minx JC, Thomas J, et al. On the use of computer-assistance to facilitate systematic mapping. *Campbell Syst Rev* [Internet]. 2020 [cited 2024 Nov 23];16(4):e1129. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/cl2.1129>
83. Clark J, McFarlane C, Cleo G, Ramos CI, Marshall S. The Impact of Systematic Review Automation Tools on Methodological Quality and Time Taken to Complete Systematic Review Tasks: Case Study. *JMIR Med Educ* [Internet]. 2021 May 31 [cited 2024 Nov 23];7(2):e24418. Available from: <https://mededu.jmir.org/2021/2/e24418>
84. Khalil H, Ameen D, Zarnegar A. Tools to support the automation of systematic reviews: a scoping review. *J Clin Epidemiol* [Internet]. 2022 Apr 1 [cited 2024 Aug 15];144:22–42. Available from: <https://www.sciencedirect.com/science/article/pii/S0895435621004029>
85. Sallam M. ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. *Healthc Basel Switz*. 2023 Mar 19;11(6):887.

86. Issaiy M, Ghanaati H, Kolahi S, Shakiba M, Jalali AH, Zarei D, et al. Methodological insights into ChatGPT's screening performance in systematic reviews. *BMC Med Res Methodol* [Internet]. 2024 Mar 27 [cited 2024 Nov 23];24(1):78. Available from: <https://doi.org/10.1186/s12874-024-02203-8>
87. Konet A, Thomas I, Gartlehner G, Kahwati L, Hilscher R, Kugley S, et al. Performance of two large language models for data extraction in evidence synthesis. *Res Synth Methods* [Internet]. [cited 2024 Aug 15];n/a(n/a). Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/jrsm.1732>
88. Guimarães NS, Joviano-Santos JV, Reis MG, Chaves RRM, Observatory of Epidemiology N Health Research (OPENS). Development of search strategies for systematic reviews in health using ChatGPT: a critical analysis. *J Transl Med* [Internet]. 2024 Jan 2 [cited 2024 Sep 25];22(1):1. Available from: <https://doi.org/10.1186/s12967-023-04371-5>
89. O'Connor AM, Tsafnat G, Thomas J, Glasziou P, Gilbert SB, Hutton B. A question of trust: can we build an evidence base to gain trust in systematic review automation technologies? *Syst Rev* [Internet]. 2019 Jun 18 [cited 2024 Nov 23];8(1):143. Available from: <https://doi.org/10.1186/s13643-019-1062-0>
90. van Altena A J., Spijker R, Olabarriaga S d. Usage of automation tools in systematic reviews. *Res Synth Methods* [Internet]. 2019 [cited 2024 Nov 23];10(1):72–82. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/jrsm.1335>
91. Johnson EE, O'Keefe H, Sutton A, Marshall C. The Systematic Review Toolbox: keeping up to date with tools to support evidence synthesis. *Syst Rev* [Internet]. 2022 Dec 1 [cited 2024 Nov 23];11(1):258. Available from: <https://doi.org/10.1186/s13643-022-02122-z>
92. Golan R, Reddy R, Muthigi A, Ramasamy R. Artificial intelligence in academic writing: a paradigm-shifting technological advance. *Nat Rev Urol* [Internet]. 2023 Jun [cited 2024 Nov 23];20(6):327–8. Available from: <https://www.nature.com/articles/s41585-023-00746-x>
93. Fabiano N, Gupta A, Bhambra N, Luu B, Wong S, Maaz M, et al. How to optimize the systematic review process using AI tools. *JCPP Adv* [Internet]. [cited 2024 May 26];n/a(n/a):e12234. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/jcv2.12234>
94. Cowie K, Rahmatullah A, Hardy N, Holub K, Kallmes K. Web-Based Software Tools for Systematic Literature Review in Medicine: Systematic Search and Feature Analysis. *JMIR Med Inform* [Internet]. 2022 May 2 [cited 2024 Jul 31];10(5):e33219. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9112080/>
95. Department-of-Defense-Responsible-Artificial-Intelligence-Strategy-and-Implementation-Pathway.pdf [Internet]. [cited 2024 May 26]. Available from: <https://media.defense.gov/2022/Jun/22/2003022604/-1/-1/0/Department-of-Defense-Responsible-Artificial-Intelligence-Strategy-and-Implementation-Pathway.PDF>
96. Ge L, Agrawal R, Singer M, Kannapiran P, De Castro Molina JA, Teow KL, et al. Leveraging artificial intelligence to enhance systematic reviews in health research: advanced tools and challenges. *Syst Rev* [Internet]. 2024 Oct 25 [cited 2024 Oct 28];13(1):269. Available from: <https://doi.org/10.1186/s13643-024-02682-2>
97. Gartlehner G, Wagner G, Lux L, Affengruber L, Dobrescu A, Kaminski-Hartenthaler A, et al. Assessing the accuracy of machine-assisted abstract screening with DistillerAI: a user study. *Syst Rev* [Internet]. 2019 Nov 15 [cited 2024 Nov 23];8(1):277. Available from: <https://doi.org/10.1186/s13643-019-1221-3>
98. Gates A, Guitard S, Pillay J, Elliott SA, Dyson MP, Newton AS, et al. Performance and usability of machine learning for screening in systematic reviews: a comparative evaluation of three tools. *Syst Rev* [Internet]. 2019 Nov 15 [cited 2024 Nov 23];8(1):278. Available from: <https://doi.org/10.1186/s13643-019-1222-2>

99. Hamel C, Kelly SE, Thavorn K, Rice DB, Wells GA, Hutton B. An evaluation of DistillerSR's machine learning-based prioritization tool for title/abstract screening – impact on reviewer-relevant outcomes. *BMC Med Res Methodol* [Internet]. 2020 Oct 15 [cited 2024 Nov 23];20(1):256. Available from: <https://doi.org/10.1186/s12874-020-01129-1>
100. Tsou AY, Treadwell JR, Erinoff E, Schoelles K. Machine learning for screening prioritization in systematic reviews: comparative performance of Abstrackr and EPPI-Reviewer. *Syst Rev* [Internet]. 2020 Apr 2 [cited 2024 Nov 23];9(1):73. Available from: <https://doi.org/10.1186/s13643-020-01324-7>
101. O'Neill G, Geisler P. Health Literacy as a Pathway to Wellbeing: Del J Public Health [Internet]. 2023 Dec 31 [cited 2024 Nov 23];9(5):12. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10810290/>
102. Stormacq C, Wosinski J, Boillat E, Van den Broucke S. Effects of health literacy interventions on health-related outcomes in socioeconomically disadvantaged adults living in the community: a systematic review. *JBI Evid Synth* [Internet]. 2020 Jul [cited 2024 May 26];18(7):1389. Available from: https://journals.lww.com/jbisir/fulltext/2020/07000/effects_of_health_literacy_interventions_on.3.aspx
103. Sorensen L, Gyrd-Hansen D, Kristiansen IS, Nexøe J, Nielsen JB. Laypersons' understanding of relative risk reductions: Randomised cross-sectional study. *BMC Med Inform Decis Mak* [Internet]. 2008 Jul 17 [cited 2024 May 26];8(1):31. Available from: <https://doi.org/10.1186/1472-6947-8-31>
104. Hopkins WG. New View of Statistics: Effect Magnitudes [Internet]. 2002 [cited 2024 May 26]. Available from: <http://www.sportsci.org/resource/stats/effectmag.html>
105. National Institutes of Health (NIH) [Internet]. [cited 2024 May 26]. Plain Language. Available from: <https://www.nih.gov/institutes-nih/nih-office-director/office-communications-public-liaison/clear-communication/plain-language>
106. Plain English summaries | NIHR [Internet]. [cited 2024 Nov 23]. Available from: <https://www.nihr.ac.uk/plain-english-summaries>
107. Evidence-based Practice Center (EPC) Reports [Internet]. [cited 2024 May 26]. Available from: <https://www.ahrq.gov/research/findings/evidence-based-reports/index.html>
108. Higgins J. Cochrane Handbook for Systematic Reviews of Interventions [Internet]. 2023 [cited 2024 May 26]. Available from: <https://training.cochrane.org/handbook>
109. Lima JP, Tangamornsuksan W, Guyatt GH. Trustworthy evidence-based versus untrustworthy guidelines: detecting the difference. *Fam Med Community Health* [Internet]. 2023 Oct 6 [cited 2024 Aug 23];11(4). Available from: <https://fmch.bmj.com/content/11/4/e002437>
110. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ* [Internet]. 2017 Sep 21 [cited 2024 Jun 3];358:j4008. Available from: <https://www.bmj.com/content/358/bmj.j4008>
111. Ioannidis JPA. Integration of evidence from multiple meta-analyses: a primer on umbrella reviews, treatment networks and multiple treatments meta-analyses. *CMAJ* [Internet]. 2009 Oct 13 [cited 2024 Nov 23];181(8):488–93. Available from: <https://www.cmaj.ca/content/181/8/488>
112. Belbasis L, Brooker RD, Zavalis E, Pezzullo AM, Axfors C, Ioannidis JP. Mapping and systematic appraisal of umbrella reviews in epidemiological research: a protocol for a meta-epidemiological study. *Syst Rev* [Internet]. 2023 Jul 14 [cited 2024 May 26];12(1):123. Available from: <https://doi.org/10.1186/s13643-023-02265-7>
113. Fontelo P, Liu F. A review of recent publication trends from top publishing countries. *Syst Rev* [Internet]. 2018 Sep 27 [cited 2024 May 26];7(1):147. Available from: <https://doi.org/10.1186/s13643-018-0819-1>

114. Cant R, Ryan C, Kelly MA. A nine-step pathway to conduct an umbrella review of literature. *Nurse Author Ed* [Internet]. 2022 [cited 2024 May 26];32(2):31–4. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/nae2.12039>
115. Papatheodorou S. Umbrella reviews: what they are and why we need them. *Eur J Epidemiol*. 2019 Jun;34(6):543–6.
116. Smith V, Devane D, Begley CM, Clarke M. Methodology in conducting a systematic review of systematic reviews of healthcare interventions. *BMC Med Res Methodol* [Internet]. 2011 Feb 3 [cited 2024 Nov 23];11(1):15. Available from: <https://doi.org/10.1186/1471-2288-11-15>
117. Aromataris E, Fernandez R, Godfrey CM, Holly C, Khalil H, Tungpunkom P. Summarizing systematic reviews: methodological development, conduct and reporting of an umbrella review approach. *JB I Evid Implement* [Internet]. 2015 Sep [cited 2024 May 26];13(3):132. Available from: https://journals.lww.com/ijebh/fulltext/2015/09000/summarizing_systematic_reviews__methodological.4.aspx
118. Pollock M, Fernandes RM, Becker LA, Featherstone R, Hartling L. What guidance is available for researchers conducting overviews of reviews of healthcare interventions? A scoping review and qualitative metasummary. *Syst Rev* [Internet]. 2016 Nov 14 [cited 2024 May 26];5(1):190. Available from: <https://doi.org/10.1186/s13643-016-0367-5>
119. Biondi-Zoccai G, editor. *Umbrella Reviews: Evidence Synthesis with Overviews of Reviews and Meta-Epidemiologic Studies* [Internet]. Cham: Springer International Publishing; 2016 [cited 2024 May 26]. Available from: <http://link.springer.com/10.1007/978-3-319-25655-9>
120. Hunt H, Pollock A, Campbell P, Estcourt L, Brunton G. An introduction to overviews of reviews: planning a relevant research question and objective for an overview. *Syst Rev* [Internet]. 2018 Mar 1 [cited 2024 May 26];7(1):39. Available from: <https://doi.org/10.1186/s13643-018-0695-8>
121. Hartling L, Chisholm A, Thomson D, Dryden DM. A Descriptive Analysis of Overviews of Reviews Published between 2000 and 2011. *PLOS ONE* [Internet]. 2012 Nov 15 [cited 2024 May 26];7(11):e49667. Available from: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0049667>
122. Choi GJ, Kang H. The umbrella review: a useful strategy in the rain of evidence. *Korean J Pain* [Internet]. 2022 Apr 1 [cited 2024 Nov 3];35(2):127. Available from: <https://pmc.ncbi.nlm.nih.gov/articles/PMC8977198/>
123. Belbasis L, Bellou V, Ioannidis JPA. Conducting umbrella reviews. *BMJ Med* [Internet]. 2022 Nov 1 [cited 2024 Jun 5];1(1). Available from: <https://bmjmedicine.bmj.com/content/1/1/e000071>
124. Bolland MJ, Grey A, Reid IR. Differences in Overlapping Meta-Analyses of Vitamin D Supplements and Falls. *J Clin Endocrinol Metab* [Internet]. 2014 Nov 1 [cited 2024 May 26];99(11):4265–72. Available from: <https://doi.org/10.1210/jc.2014-2562>
125. Cooper H, Koenka AC. The overview of reviews: Unique challenges and opportunities when research syntheses are the principal elements of new integrative scholarship. *Am Psychol*. 2012;67(6):446–62.
126. Pollock M. Chapter V: Overviews of Reviews [Internet]. 2023 [cited 2024 May 26]. Available from: <https://training.cochrane.org/handbook/current/chapter-v>
127. Niforatos JD, Weaver M, Johansen ME. Assessment of Publication Trends of Systematic Reviews and Randomized Clinical Trials, 1995 to 2017. *JAMA Intern Med* [Internet]. 2019 Nov 1 [cited 2024 May 26];179(11):1593–4. Available from: <https://doi.org/10.1001/jamainternmed.2019.3013>
128. Lunny C, Neelakant T, Chen A, Shinger G, Stevens A, Tasnim S, et al. Bibliometric study of 'overviews of systematic reviews' of health interventions: Evaluation of prevalence, citation

- and journal impact factor. *Res Synth Methods* [Internet]. 2022 [cited 2024 Nov 23];13(1):109–20. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/jrsm.1530>
129. Thomson D, Russell K, Becker L, Klassen T, Hartling L. The evolution of a new publication type: Steps and challenges of producing overviews of reviews. *Res Synth Methods* [Internet]. 2010 [cited 2024 May 26];1(3–4):198–211. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/jrsm.30>
130. Kazrin A, Durac J, Agteros T. Meta-meta analysis: a new method for evaluating therapy outcome. *Behav Res Ther*. 1979;17(4):397–9.
131. Pieper D, Antoine SL, Morfeld JC, Mathes T, Eikermann M. Methodological approaches in conducting overviews: current state in HTA agencies. *Res Synth Methods* [Internet]. 2014 [cited 2024 Nov 23];5(3):187–99. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/jrsm.1107>
132. Lunny C, McKenzie JE, McDonald S. Retrieval of overviews of systematic reviews in MEDLINE was improved by the development of an objectively derived and validated search strategy. *J Clin Epidemiol* [Internet]. 2016 Jun 1 [cited 2024 Nov 23];74:107–18. Available from: [https://www.jclinepi.com/article/S0895-4356\(15\)00578-8/abstract](https://www.jclinepi.com/article/S0895-4356(15)00578-8/abstract)
133. McKenzie JE, Brennan SE. Overviews of systematic reviews: great promise, greater challenge. *Syst Rev* [Internet]. 2017 Sep 8 [cited 2024 Jun 5];6(1):185. Available from: <https://doi.org/10.1186/s13643-017-0582-8>
134. Pieper D, Buechter R, Jerinic P, Eikermann M. Overviews of reviews often have limited rigor: a systematic review. *J Clin Epidemiol*. 2012 Dec;65(12):1267–73.
135. Zhao L, Shen C, Liu M, Zhang J, Cheng L, Li Y, et al. Comparison of Reporting and Transparency in Published Protocols and Publications in Umbrella Reviews: Scoping Review. *J Med Internet Res* [Internet]. 2023 Aug 2 [cited 2024 Nov 23];25(1):e43299. Available from: <https://www.jmir.org/2023/1/e43299>
136. Pollock M, Fernandes RM, Becker LA, Featherstone R, Hartling L. What guidance is available for researchers conducting overviews of reviews of healthcare interventions? A scoping review and qualitative metasummary. *Syst Rev* [Internet]. 2016 Nov 14 [cited 2024 Nov 23];5(1):190. Available from: <https://doi.org/10.1186/s13643-016-0367-5>
137. Lunny C, Brennan SE, McDonald S, McKenzie JE. Toward a comprehensive evidence map of overview of systematic review methods: paper 1—purpose, eligibility, search and data extraction. *Syst Rev* [Internet]. 2017 Nov 21 [cited 2024 Jun 5];6(1):231. Available from: <https://doi.org/10.1186/s13643-017-0617-1>
138. Lunny C, Brennan SE, McDonald S, McKenzie JE. Toward a comprehensive evidence map of overview of systematic review methods: paper 2—risk of bias assessment; synthesis, presentation and summary of the findings; and assessment of the certainty of the evidence. *Syst Rev* [Internet]. 2018 Oct 12 [cited 2024 Nov 23];7(1):159. Available from: <https://doi.org/10.1186/s13643-018-0784-8>
139. Pieper D, Antoine SL, Morfeld JC, Mathes T, Eikermann M. Methodological approaches in conducting overviews: current state in HTA agencies. *Res Synth Methods* [Internet]. 2014 [cited 2024 Nov 23];5(3):187–99. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/jrsm.1107>
140. Becker LA, Oxman AD. Overviews of Reviews. In: *Cochrane Handbook for Systematic Reviews of Interventions* [Internet]. John Wiley & Sons, Ltd; 2008 [cited 2024 Nov 23]. p. 607–31. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470712184.ch22>
141. Aromataris E, Lockwood C, Porritt K, Pilla B. *JB I Manual for Evidence Synthesis - JB I Global Wiki* [Internet]. [cited 2024 Jun 5]. Available from: <https://jbi-global-wiki.refined.site/space/MANUAL>

142. Lunny C, Brennan SE, McDonald S, McKenzie JE. Toward a comprehensive evidence map of overview of systematic review methods: paper 2—risk of bias assessment; synthesis, presentation and summary of the findings; and assessment of the certainty of the evidence. *Syst Rev* [Internet]. 2018 Oct 12 [cited 2024 Jun 5];7(1):159. Available from: <https://doi.org/10.1186/s13643-018-0784-8>
143. Gianfredi V, Nucci D, Amerio A, Signorelli C, Odone A, Dinu M. What Can We Expect from an Umbrella Review? *Adv Nutr* [Internet]. 2022 Apr 1 [cited 2024 Oct 4];13(2):684–5. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8970837/>
144. Shi X, Wallach JD. Umbrella reviews: a useful study design in need of standardisation. *BMJ* [Internet]. 2022 Aug 5 [cited 2024 Oct 4];o1740. Available from: <https://www.bmj.com/lookup/doi/10.1136/bmj.o1740>
145. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. New York: Routledge; 1988. 567 p.
146. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol* [Internet]. 2011 Apr 1 [cited 2024 Jul 31];64(4):383–94. Available from: [https://www.jclinepi.com/article/S0895-4356\(10\)00330-6/fulltext](https://www.jclinepi.com/article/S0895-4356(10)00330-6/fulltext)
147. Stone JC, Barker TH, Aromataris E, Ritskes-Hoitinga M, Sears K, Klugar M, et al. From critical appraisal to risk of bias assessment: clarifying the terminology for study evaluation in JBI systematic reviews. *JBI Evid Synth* [Internet]. 2023 Mar [cited 2024 Jul 2];21(3):472. Available from: https://journals.lww.com/jbisir/fulltext/2023/03000/from_critical_appraisal_to_risk_of_bias.3.aspx
148. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* [Internet]. 2008 Apr 24 [cited 2024 Jun 6];336(7650):924–6. Available from: <https://www.bmj.com/content/336/7650/924>
149. Hultcrantz M, Rind D, Akl EA, Treweek S, Mustafa RA, Iorio A, et al. The GRADE Working Group clarifies the construct of certainty of evidence. *J Clin Epidemiol* [Internet]. 2017 May 18 [cited 2024 Oct 20];87:4. Available from: <https://pmc.ncbi.nlm.nih.gov/articles/PMC6542664/>
150. Clinical practice guidelines: Concepts, limitations and challenges [Internet]. 2020 [cited 2024 Nov 23]. Available from: <https://www.medwave.cl/revisiones/metodoinvestreport/7887.html>
151. Gunasekaran KP. Exploring Sentiment Analysis Techniques in Natural Language Processing: A Comprehensive Review [Internet]. 2023 [cited 2024 Jul 29]. Available from: <http://arxiv.org/abs/2305.14842>
152. Zhang W, Deng Y, Liu B, Pan SJ, Bing L. Sentiment Analysis in the Era of Large Language Models: A Reality Check [Internet]. *arXiv*; 2023 [cited 2024 Jul 29]. Available from: <http://arxiv.org/abs/2305.15005>
153. Parasuraman A, Zeithaml V, Berry L. Servqual: A Multiple-Item Scale For Measuring Consumer Perc - ProQuest [Internet]. [cited 2024 Jul 3]. Available from: <https://www.proquest.com/openview/7d007e04d78261295e5524f15bef6837/1?pq-origsite=gscholar&cbl=41988>
154. CIS [Internet]. [cited 2024 Nov 24]. ¿Cómo se hacen las encuestas? Available from: <https://www.cis.es/estudios/preguntas-frecuentes/como-se-hacen-las-encuestas>
155. Núñez González E, Estévez Guerra GJ, Hernández Marrero P, Marrero Medina CD. Una propuesta destinada a complementar el cuestionario Font-Roja de satisfacción laboral. *Gac Sanit* [Internet]. 2007 Apr [cited 2024 Nov 24];21(2):136–41. Available from: https://scielo.isciii.es/scielo.php?script=sci_abstract&pid=S0213-91112007000200008&lng=es&nrm=iso&tlng=es

156. OpenAI [Internet]. [cited 2024 Nov 24]. Available from: <https://openai.com/>
157. uems.eu [Internet]. [cited 2024 Nov 24]. Home. Available from: <https://www.uems.eu/>
158. Hinkle DE, Wiersma W, Jurs SG. *Applied Statistics for the Behavioral Sciences*. Houghton Mifflin; 2003. 790 p.
159. Hopkins WG, Marshall SW, Batterham AM, Hanin J. *Progressive Statistics for Studies in Sports Medicine and Exercise Science*. *Med Sci Sports Exerc* [Internet]. 2009 Jan [cited 2024 Sep 13];41(1):3. Available from: https://journals.lww.com/acsm-msse/fulltext/2009/01000/progressive_statistics_for_studies_in_sports.2.aspx
160. Rana R, Singhal R. Chi-square Test and its Application in Hypothesis Testing. *J Pract Cardiovasc Sci* [Internet]. 2015 Apr [cited 2024 Nov 20];1(1):69. Available from: https://journals.lww.com/jpcs/fulltext/2015/01010/chi_square_test_and_its_application_in_hypothesis.17.aspx
161. Conneely M, Leahy S, Dore L, Trépel D, Robinson K, Jordan F, et al. The effectiveness of interventions to reduce adverse outcomes among older adults following Emergency Department discharge: umbrella review. *BMC Geriatr* [Internet]. 2022 May 28 [cited 2024 Oct 20];22(1):462. Available from: <https://doi.org/10.1186/s12877-022-03007-5>
162. Visser D, Wattel EM, Gerrits KHL, Wouden JC van der, Meiland FJM, Groot AJ de, et al. Effectiveness and characteristics of physical fitness training on aerobic fitness in vulnerable older adults: an umbrella review of systematic reviews. *BMJ Open* [Internet]. 2022 May 1 [cited 2024 Nov 6];12(5):e058056. Available from: <https://bmjopen.bmj.com/content/12/5/e058056>
163. Yuan Y, Lin S, Lin W, Huang F, Zhu P. Modifiable predictive factors and all-cause mortality in the non-hospitalized elderly population: An umbrella review of meta-analyses. *Exp Gerontol* [Internet]. 2022 Jun 15 [cited 2024 Nov 6];163:111792. Available from: <https://www.sciencedirect.com/science/article/pii/S0531556522001000>
164. Travica N, Lotfaliany M, Marriott A, Safavynia SA, Lane MM, Gray L, et al. Peri-Operative Risk Factors Associated with Post-Operative Cognitive Dysfunction (POCD): An Umbrella Review of Meta-Analyses of Observational Studies. *J Clin Med* [Internet]. 2023 Jan [cited 2024 Nov 6];12(4):1610. Available from: <https://www.mdpi.com/2077-0383/12/4/1610>
165. Veronese N, Honvo G, Bruyère O, Rizzoli R, Barbagallo M, Maggi S, et al. Knee osteoarthritis and adverse health outcomes: an umbrella review of meta-analyses of observational studies. *Aging Clin Exp Res* [Internet]. 2023 Feb 1 [cited 2024 Nov 6];35(2):245–52. Available from: <https://doi.org/10.1007/s40520-022-02289-4>
166. Li J, Cao D, Huang Y, Chen B, Chen Z, Wang R, et al. Zinc Intakes and Health Outcomes: An Umbrella Review. *Front Nutr* [Internet]. 2022 Feb 8 [cited 2024 Nov 6];9. Available from: <https://www.frontiersin.org/journals/nutrition/articles/10.3389/fnut.2022.798078/full>
167. Shen Y, Liu D, Li S, He Y, Tan F, Sun X, et al. Effects of Exercise on Patients Important Outcomes in Older People With Sarcopenia: An Umbrella Review of Meta-Analyses of Randomized Controlled Trials. *Front Med* [Internet]. 2022 Feb 3 [cited 2024 Oct 20];9. Available from: <https://www.frontiersin.org/journals/medicine/articles/10.3389/fmed.2022.811746/full>
168. Beck Jepsen D, Robinson K, Ogliari G, Montero-Odasso M, Kamkar N, Ryg J, et al. Predicting falls in older adults: an umbrella review of instruments assessing gait, balance, and functional mobility. *BMC Geriatr* [Internet]. 2022 Jul 25 [cited 2024 Nov 6];22(1):615. Available from: <https://doi.org/10.1186/s12877-022-03271-5>
169. Gazzaniga G, Menichelli D, Scaglione F, Farcomeni A, Pani A, Pastori D. Effect of digoxin on all-cause and cardiovascular mortality in patients with atrial fibrillation with and without heart failure: an umbrella review of systematic reviews and 12 meta-analyses. *Eur J Clin Pharmacol* [Internet]. 2023 Mar 6 [cited 2024 Oct 20];79(4):473. Available from: <https://pmc.ncbi.nlm.nih.gov/articles/PMC10039090/>

170. Musazadeh V, Kavyani Z, Mirhosseini N, Dehghan P, Vajdi M. Effect of vitamin D supplementation on type 2 diabetes biomarkers: an umbrella of interventional meta-analyses. *Diabetol Metab Syndr* [Internet]. 2023 Apr 19 [cited 2024 Nov 6];15(1):76. Available from: <https://doi.org/10.1186/s13098-023-01010-3>
171. Veronese N, Galvano D, D'Antiga F, Vecchiato C, Furegon E, Allocco R, et al. Interventions for reducing loneliness: An umbrella review of intervention studies. *Health Soc Care Community* [Internet]. 2021 [cited 2024 Oct 20];29(5):e89–96. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/hsc.13248>
172. Veronese N, Smith L, Bolzetta F, Cester A, Demurtas J, Punzi L. Efficacy of conservative treatments for hand osteoarthritis. *Wien Klin Wochenschr* [Internet]. 2021 Mar 1 [cited 2024 Oct 20];133(5):234–40. Available from: <https://doi.org/10.1007/s00508-020-01702-0>
173. Veronese N, Demurtas J, Thompson T, Solmi M, Pesolillo G, Celotto S, et al. Effect of low-dose aspirin on health outcomes: An umbrella review of systematic reviews and meta-analyses. *Br J Clin Pharmacol* [Internet]. 2020 [cited 2024 Nov 6];86(8):1465–75. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/bcp.14310>
174. Soysal P, Veronese N, Ippoliti S, Pizzol D, Carrie AM, Stefanescu S, et al. The impact of urinary incontinence on multiple health outcomes: an umbrella review of meta-analysis of observational studies. *Aging Clin Exp Res* [Internet]. 2023 Jan 13 [cited 2024 Nov 24];35(3):479–95. Available from: <http://www.scopus.com/inward/record.url?scp=85146238196&partnerID=8YFLogXK>
175. Zhang GQ, Chen JL, Luo Y, Mathur MB, Anagnostis P, Nurmatov U, et al. Menopausal hormone therapy and women's health: An umbrella review. *PLOS Med* [Internet]. 2021 Aug 2 [cited 2024 Nov 6];18(8):e1003731. Available from: <https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1003731>
176. Marx W, Veronese N, Kelly JT, Smith L, Hockey M, Collins S, et al. The Dietary Inflammatory Index and Human Health: An Umbrella Review of Meta-Analyses of Observational Studies. *Adv Nutr* [Internet]. 2021 Sep 1 [cited 2024 Nov 6];12(5):1681–90. Available from: <https://www.sciencedirect.com/science/article/pii/S216183132200463X>
177. Trott M, Smith L, Veronese N, Pizzol D, Barnett Y, Gorely T, et al. Eye disease and mortality, cognition, disease, and modifiable risk factors: an umbrella review of meta-analyses of observational studies. *Eye* [Internet]. 2022 Feb [cited 2024 Nov 6];36(2):369–78. Available from: <https://www.nature.com/articles/s41433-021-01684-x>
178. Celotto S, Veronese N, Barbagallo M, Ometto F, Smith L, Pardhan S, et al. An umbrella review of systematic reviews with meta-analyses evaluating positive and negative outcomes of Hydroxychloroquine and chloroquine therapy. *Int J Infect Dis* [Internet]. 2021 Feb 1 [cited 2024 Nov 6];103:599–606. Available from: [https://www.ijidonline.com/article/S1201-9712\(20\)32541-8/fulltext](https://www.ijidonline.com/article/S1201-9712(20)32541-8/fulltext)
179. Demurtas J, Fanelli GN, Romano SL, Solari M, Yang L, Soysal P, et al. Stem cells for treatment of cardiovascular diseases: An umbrella review of randomized controlled trials. *Ageing Res Rev* [Internet]. 2021 May 1 [cited 2024 Nov 24];67:101257. Available from: <https://www.sciencedirect.com/science/article/pii/S1568163721000040>
180. Demurtas J, Schoene D, Torbahn G, Marengoni A, Grande G, Zou L, et al. Physical Activity and Exercise in Mild Cognitive Impairment and Dementia: An Umbrella Review of Intervention and Observational Studies. *J Am Med Dir Assoc* [Internet]. 2020 Oct 1 [cited 2024 Nov 24];21(10):1415–1422.e6. Available from: <https://www.sciencedirect.com/science/article/pii/S1525861020307374>
181. Obaid M, Douiri A, Flach C, Prasad V, Marshall I. Can we prevent poststroke cognitive impairment? An umbrella review of risk factors and treatments. *BMJ Open* [Internet]. 2020 Sep 1 [cited 2024 Nov 6];10(9):e037982. Available from: <https://bmjopen.bmj.com/content/10/9/e037982>

182. Veronese N, Demurtas J, Smith L, Reginster JY, Bruyère O, Beaudart C, et al. Glucosamine sulphate: an umbrella review of health outcomes. *Ther Adv Musculoskelet Dis* [Internet]. 2020 Jan 1 [cited 2024 Nov 6];12:1759720X20975927. Available from: <https://doi.org/10.1177/1759720X20975927>
183. Trott M, Smith L, Xiao T, Veronese N, Koyanagi A, Jacob L, et al. Hearing impairment and diverse health outcomes. *Wien Klin Wochenschr* [Internet]. 2021 Oct 1 [cited 2024 Nov 6];133(19):1028–41. Available from: <https://doi.org/10.1007/s00508-021-01893-0>
184. Neil-Sztramko SE, Teggart K, Moore C, Sherifali D, Fitzpatrick-Lewis D, Coletta G, et al. Community-based group physical activity and/or nutrition interventions to promote mobility in older adults: an umbrella review. *BMC Geriatr* [Internet]. 2022 Jun 29 [cited 2024 Nov 6];22(1):539. Available from: <https://doi.org/10.1186/s12877-022-03170-9>
185. Antonazzo B, Marano G, Romagnoli E, Ronzoni S, Frati G, Sani G, et al. Impact of arterial hypertension and its management strategies on cognitive function and dementia: a comprehensive umbrella review. *Minerva Cardiol Angiol*. 2022 Jun;70(3):285–97.
186. Demurtas J, Celotto S, Beaudart C, Sanchez-Rodriguez D, Balci C, Soysal P, et al. The efficacy and safety of influenza vaccination in older people: An umbrella review of evidence from meta-analyses of both observational and randomized controlled studies. *Ageing Res Rev* [Internet]. 2020 Sep 1 [cited 2024 Nov 6];62:101118. Available from: <https://www.sciencedirect.com/science/article/pii/S1568163720302531>
187. Conroy T, Heuzenroeder L, Feo R. In-hospital interventions for reducing readmissions to acute care for adults aged 65 and over: An umbrella review. *Int J Qual Health Care* [Internet]. 2020 Aug 1 [cited 2024 Nov 6];32(7):414–30. Available from: <https://doi.org/10.1093/intqhc/mzaa064>
188. Gielen E, Beckwée D, Delaere A, De Breucker S, Vandewoude M, Bautmans I, et al. Nutritional interventions to improve muscle mass, muscle strength, and physical performance in older people: an umbrella review of systematic reviews and meta-analyses. *Nutr Rev* [Internet]. 2021 Jan 9 [cited 2024 Nov 6];79(2):121–47. Available from: <https://doi.org/10.1093/nutrit/nuaa011>
189. Biondi-Zoccai G, Antonazzo B, Giordano A, Versaci F, Frati G, Ronzoni S, et al. Oral antiplatelet therapy in the elderly undergoing percutaneous coronary intervention: an umbrella review. *J Thorac Dis* [Internet]. 2020 Apr [cited 2024 Nov 6];12(4). Available from: <https://jtd.amegroups.org/article/view/35967>
190. Ofori-Asenso R, Chin KL, Sahle BW, Mazidi M, Zullo AR, Liew D. Frailty Confers High Mortality Risk across Different Populations: Evidence from an Overview of Systematic Reviews and Meta-Analyses. *Geriatrics* [Internet]. 2020 Mar [cited 2024 Nov 6];5(1):17. Available from: <https://www.mdpi.com/2308-3417/5/1/17>
191. Veronese N, Demurtas J, Soysal P, Smith L, Torbahn G, Schoene D, et al. Sarcopenia and health-related outcomes: an umbrella review of observational studies. *Eur Geriatr Med* [Internet]. 2019 Dec 1 [cited 2024 Nov 6];10(6):853–62. Available from: <https://doi.org/10.1007/s41999-019-00233-w>
192. Beckwée D, Delaere A, Aelbrecht S, Baert V, Beaudart C, Bruyere O, et al. Exercise Interventions for the Prevention and Treatment of Sarcopenia. A Systematic Umbrella Review. *J Nutr Health Aging* [Internet]. 2019 Jun 1 [cited 2024 Nov 6];23(6):494–502. Available from: <https://www.sciencedirect.com/science/article/pii/S1279770723011958>
193. De Spiegeleer A, Beckwée D, Bautmans I, Petrovic M, Bautmans I, Beaudart C, et al. Pharmacological Interventions to Improve Muscle Mass, Muscle Strength and Physical Performance in Older People: An Umbrella Review of Systematic Reviews and Meta-analyses. *Drugs Aging* [Internet]. 2018 Aug 1 [cited 2024 Nov 6];35(8):719–34. Available from: <https://doi.org/10.1007/s40266-018-0566-y>

194. Jadcak AD, Makwana N, Luscombe-Marsh N, Visvanathan R, Schultz TJ. Effectiveness of exercise interventions on physical function in community-dwelling frail older people: an umbrella review of systematic reviews. *JBI Evid Synth* [Internet]. 2018 Mar [cited 2024 Nov 6];16(3):752. Available from: https://journals.lww.com/jbisrir/abstract/2018/03000/effectiveness_of_exercise_interventions_on.17.aspx
195. 1Poscia A, Milovanovic S, La Milia DI, Duplaga M, Grysztar M, Landi F, et al. Effectiveness of nutritional interventions addressed to elderly persons: umbrella systematic review with meta-analysis. *Eur J Public Health* [Internet]. 2018 Apr 1 [cited 2024 Nov 6];28(2):275–83. Available from: <https://doi.org/10.1093/eurpub/ckx199>
196. Stubbs B, Denking MD, Brefka S, Dallmeier D. What works to prevent falls in older adults dwelling in long term care facilities and hospitals? An umbrella review of meta-analyses of randomised controlled trials. *Maturitas*. 2015 Jul;81(3):335–42.
197. Thukral S, Kovac S, Paturu M. Chapter 30 - Chi square. In: Eltorai AEM, Liu T, Chand R, Kalva SP, editors. *Translational Interventional Radiology* [Internet]. Academic Press; 2023 [cited 2024 Nov 3]. p. 145–8. (Handbook for Designing and Conducting Clinical). Available from: <https://www.sciencedirect.com/science/article/pii/B9780128230268000286>
198. Lee DK. Alternatives to P value: confidence interval and effect size. *Korean J Anesthesiol* [Internet]. 2016 Oct 25 [cited 2024 Nov 3];69(6):555. Available from: <https://pmc.ncbi.nlm.nih.gov/articles/PMC5133225/>
199. Luchini C, Veronese N, Nottegar A, Shin JI, Gentile G, Granzio U, et al. Assessing the quality of studies in meta-research: Review/guidelines on the most important quality assessment tools. *Pharm Stat* [Internet]. 2021 [cited 2024 Nov 13];20(1):185–95. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/pst.2068>
200. Seehra J, Pandis N, Koletsi D, Fleming PS. Use of quality assessment tools in systematic reviews was varied and inconsistent. *J Clin Epidemiol* [Internet]. 2016 Jan 1 [cited 2024 Nov 13];69:179–184.e5. Available from: [https://www.jclinepi.com/article/S0895-4356\(15\)00336-4/abstract](https://www.jclinepi.com/article/S0895-4356(15)00336-4/abstract)
201. Bigna JJR, Um LN, Nansseu JRN. A comparison of quality of abstracts of systematic reviews including meta-analysis of randomized controlled trials in high-impact general medicine journals before and after the publication of PRISMA extension for abstracts: a systematic review and meta-analysis. *Syst Rev*. 2016 Oct 13;5(1):174.
202. Li T, Hua F, Dan S, Zhong Y, Levey C, Song Y. Reporting quality of systematic review abstracts in operative dentistry: An assessment using the PRISMA for Abstracts guidelines. *J Dent* [Internet]. 2020 Nov 1 [cited 2024 Nov 16];102:103471. Available from: <https://www.sciencedirect.com/science/article/pii/S0300571220302177>
203. El Ansari W, AlRumaihi K, El-Ansari K, Arafa M, Elbardisi H, Majzoub A, et al. Reporting quality of abstracts of systematic reviews/meta-analyses: An appraisal of Arab Journal of Urology across 12 years: the PRISMA-Abstracts checklist. *Arab J Urol* [Internet]. 2023 Jan 2 [cited 2024 Nov 16];21(1):52–65. Available from: <https://doi.org/10.1080/2090598X.2022.2113127>
204. Zhong Y, Wang Y, Dan S, Zhao T, Li T, Qin D, et al. THE REPORTING QUALITY OF SYSTEMATIC REVIEW ABSTRACTS IN LEADING GENERAL DENTAL JOURNALS: A METHODOLOGICAL STUDY. *J Evid-Based Dent Pract* [Internet]. 2023 Mar 1 [cited 2024 Nov 16];23(1):101831. Available from: <https://www.sciencedirect.com/science/article/pii/S1532338222001592>
205. Berthet V. The Impact of Cognitive Biases on Professionals' Decision-Making: A Review of Four Occupational Areas. *Front Psychol* [Internet]. 2022 Jan 4 [cited 2024 Nov 16];12. Available from: <https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2021.802439/full>
206. Tricco AC, Langlois EtienneV, Straus SE, Alliance for Health Policy and Systems Research, World Health Organization. Rapid reviews to strengthen health policy and systems: a practical guide [Internet]. Geneva: World Health Organization; 2017 [cited 2024 Nov 23]. Available from: <https://iris.who.int/handle/10665/258698>
207. Heinen L, Goossen K, Lunny C, Hirt J, Puljak L, Pieper D. The optimal approach for retrieving systematic reviews was achieved when searching MEDLINE and Epistemonikos in addition to reference checking: a methodological validation study. *BMC Med Res Methodol* [Internet]. 2024 Nov 9 [cited 2024 Nov 15];24(1):271. Available from: <https://doi.org/10.1186/s12874-024-02384-2>
208. Bramer WM, Rethlefsen ML, Kleijnen J, Franco OH. Optimal database combinations for literature searches in systematic reviews: a prospective exploratory study. *Syst Rev* [Internet]. 2017 Dec 6 [cited 2024 Aug 16];6(1):245. Available from: <https://doi.org/10.1186/s13643-017-0644-y>
209. Aromataris E, Riitano D. Systematic Reviews: Constructing a Search Strategy and Searching for Evidence. *AJN Am J Nurs* [Internet]. 2014 May [cited 2024 Nov 11];114(5):49. Available from: https://journals.lww.com/ajnonline/FullText/2014/05000/Systematic_Reviews___Constructing_a_Search.27.aspx
210. Altman DG, Bland JM. How to obtain the confidence interval from a P value. *BMJ* [Internet]. 2011 Aug 8 [cited 2024 Jul 31];343:d2090. Available from: <https://www.bmj.com/content/343/bmj.d2090>
211. Andrade C. Mean Difference, Standardized Mean Difference (SMD), and Their Use in Meta-Analysis: As Simple as It Gets. *J Clin Psychiatry* [Internet]. 2020 Sep 22 [cited 2024 Nov 4];81(5):11349. Available from: <https://www.psychiatrist.com/jcp/mean-difference-standardized-mean-difference-smd-and-their-use-in-meta-analysis/>
212. Xu C, Yu T, Furuya-Kanamori L, Lin L, Zorzela L, Zhou X, et al. Validity of data extraction in evidence synthesis practice of adverse events: reproducibility study. *The BMJ* [Internet]. 2022 May 10 [cited 2024 Nov 24];377:e069155. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9086856/>
213. Mathes T, Klaffen P, Pieper D. Frequency of data extraction errors and methods to increase data extraction quality: a methodological review. *BMC Med Res Methodol* [Internet]. 2017 Nov 28 [cited 2024 Nov 24];17(1):152. Available from: <https://doi.org/10.1186/s12874-017-0431-4>
214. Kovacs M, Hoekstra R, Aczel B. The Role of Human Fallibility in Psychological Research: A Survey of Mistakes in Data Management. *Adv Methods Pract Psychol Sci* [Internet]. 2021 Oct 1 [cited 2024 Nov 24];4(4):25152459211045930. Available from: <https://doi.org/10.1177/25152459211045930>
215. Belouafa S, Habti F, Benhar S, Belafkih B, Tayane S, Hamdouch S, et al. Statistical tools and approaches to validate analytical methods: methodology and practical examples. *Int J Metrol Qual Eng* [Internet]. 2017 [cited 2024 Nov 4];8:9. Available from: <https://www.metrology-journal.org/articles/ijmqe/abs/2017/01/ijmqe160046/ijmqe160046.html>
216. Sullivan GM, Feinn R. Using Effect Size—or Why the P Value Is Not Enough. *J Grad Med Educ* [Internet]. 2012 Sep [cited 2024 Nov 24];4(3):279–82. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3444174/>
217. Demidenko E. The p-Value You Can't Buy. *Am Stat* [Internet]. 2016 Jan 2 [cited 2024 Nov 24];70(1):33–8. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4867863/>
218. Greenland S. Valid P-Values Behave Exactly as They Should: Some Misleading Criticisms of P-Values and Their Resolution With S-Values. *Am Stat* [Internet]. 2019 Mar 29 [cited 2024 Nov 24];73(sup1):106–14. Available from: <https://doi.org/10.1080/00031305.2018.1529625>
219. LU Y, BELITSKAYA-LEVY I. The debate about p-values. *Shanghai Arch Psychiatry* [Internet]. 2015 Dec 25 [cited 2024 Nov 24];27(6):381–5. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4858512/>

220. Brzezińska J. The Problem of Zero Cells in the Analysis of Contingency Tables. *Zesz Nauk Univ Ekon W Krakowie*. 2015 Nov 16;49–61.
221. Sapra RL, Saluja S. Understanding statistical association and correlation. *Curr Med Res Pract* [Internet]. 2021 Feb [cited 2024 Nov 24];11(1):31. Available from: https://journals.lww.com/cmre/fulltext/2021/11010/understanding_statistical_association_and.7.aspx
222. Clinical practice guidelines: Concepts, limitations and challenges [Internet]. 2020 [cited 2024 Oct 20]. Available from: <https://www.medwave.cl/revisiones/metodoinvestreport/7887.html>
223. Schlesinger S, Schwingshackl L, Neuenschwander M, Barbaresco J. A critical reflection on the grading of the certainty of evidence in umbrella reviews. *Eur J Epidemiol* [Internet]. 2019 Sep 1 [cited 2024 Jun 15];34(9):889–90. Available from: <https://doi.org/10.1007/s10654-019-00531-4>
224. Norman G, Wilson P, Dumville J, Bower P, Cullum N. Rapid evidence synthesis to enable innovation and adoption in health and social care. *Syst Rev* [Internet]. 2022 Nov 23 [cited 2024 Nov 24];11(1):250. Available from: <https://doi.org/10.1186/s13643-022-02106-z>
225. Skarlinski MD, Cox S, Laurent JM, Braza JD, Hinks M, Hammerling MJ, et al. Language agents achieve superhuman synthesis of scientific knowledge [Internet]. *arXiv*; 2024 [cited 2024 Nov 23]. Available from: <http://arxiv.org/abs/2409.13740>
226. Neto LSS, Rosa T dos S, Freire MD, Correa H de L, Pedreira RC, Dias FCF, et al. Geriatric and Gerontology Research: A Scientometric Investigation of Open Access Journal Articles Indexed in the Scopus Database. *Ann Geriatr Med Res* [Internet]. 2023 Jul 28 [cited 2024 Oct 27];27(3):183. Available from: <https://pmc.ncbi.nlm.nih.gov/articles/PMC10556720/>
227. McNabney MK, Green AR, Burke M, Le ST, Butler D, Chun AK, et al. Complexities of care: Common components of models of care in geriatrics. *J Am Geriatr Soc* [Internet]. 2022 [cited 2024 Oct 27];70(7):1960–72. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/jgs.17811>
228. Bonczar M, Ostrowski P, D'Antoni AV, Tubbs RS, Iwanaga J, Ghosh SK, et al. How to write an umbrella review? A step-by-step tutorial with tips and tricks. *Folia Morphol* [Internet]. 2023 [cited 2024 Oct 26];82(1):1–6. Available from: https://journals.viamedica.pl/fovia_morphologica/article/view/FM.a2022.0104
229. Sharp MK, Baki DABA, Quigley J, Tyner B, Devane D, Mahtani KR, et al. The effectiveness and acceptability of evidence synthesis summary formats for clinical guideline development groups: a mixed-methods systematic review. *Implement Sci IS* [Internet]. 2022 Oct 27 [cited 2024 Jul 31];17:74. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9615384/>
230. Yaguana JAM, Machuca MEM, Vences FVV. Tratamiento y representación de datos provenientes de escalas tipo Likert. *Cienc Lat Rev Científica Multidiscip* [Internet]. 2023 Jul 19 [cited 2024 Oct 27];7(4):736–47. Available from: <https://ciencialatina.org/index.php/cienciala/article/view/6905>
231. Benrimoh D, Tanguay-Sela M, Perlman K, Israel S, Mehlretter J, Armstrong C, et al. Using a simulation centre to evaluate preliminary acceptability and impact of an artificial intelligence-powered clinical decision support system for depression treatment on the physician–patient interaction. *BJPsych Open* [Internet]. 2021 Jan 6 [cited 2024 Nov 1];7(1):e22. Available from: <https://pmc.ncbi.nlm.nih.gov/articles/PMC8058891/>
232. Tanguay-Sela M, Benrimoh D, Popescu C, Perez T, Rollins C, Snook E, et al. Evaluating the perceived utility of an artificial intelligence-powered clinical decision support system for depression treatment using a simulation center. *Psychiatry Res*. 2022 Feb;308:114336.
233. Helenason J, Ekström C, Falk M, Papachristou P. Exploring the feasibility of an artificial intelligence based clinical decision support system for cutaneous melanoma detection in primary care – a mixed method study. *Scand J Prim Health Care* [Internet]. [cited 2024 Nov 24];42(1):51–60. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10851794/>

234. Brooke John. SUS: A 'Quick and Dirty' Usability Scale. In: *Usability Evaluation In Industry*. CRC Press; 1996.
235. Panicker RO, George AE. Adoption of Automated Clinical Decision Support System: A Recent Literature Review and a Case Study. *Arch Med Health Sci* [Internet]. 2023 Jun [cited 2024 Nov 1];11(1):86. Available from: https://journals.lww.com/armh/fulltext/2023/11010/adoption_of_automated_clinical_decision_support.16.aspx
236. Van Biesen W, Van Cauwenberge D, Decruyenaere J, Leune T, Sterckx S. An exploration of expectations and perceptions of practicing physicians on the implementation of computerized clinical decision support systems using a Qsort approach. *BMC Med Inform Decis Mak* [Internet]. 2022 Jul 16 [cited 2024 Nov 1];22(1):185. Available from: <https://doi.org/10.1186/s12911-022-01933-3>
237. Safi S, Thiessen T, Schmailzl KJ. Acceptance and Resistance of New Digital Technologies in Medicine: Qualitative Study. *JMIR Res Protoc* [Internet]. 2018 Dec 4 [cited 2024 Nov 24];7(12):e11072. Available from: <https://www.researchprotocols.org/2018/12/e11072>
238. Ayorinde A, Mensah DO, Walsh J, Ghosh I, Ibrahim SA, Hogg J, et al. Health Care Professionals' Experience of Using AI: Systematic Review With Narrative Synthesis. *J Med Internet Res* [Internet]. 2024 Oct 30 [cited 2024 Nov 24];26(1):e55766. Available from: <https://www.jmir.org/2024/1/e55766>
239. Poll DrJVD, Yong Y, Weijer NVD. Validity of AI-generated One-off Questionnaires for Organizational Transformation. *Int J Econ Bus Manag Res* [Internet]. 2022 [cited 2024 Nov 18];06(02):122–30. Available from: https://ijebmr.com/uploads/pdf/archivepdf/2022/IJEBMR_908.pdf
240. Bogner K, Landrock U. Response Biases in Standardised Surveys (GESIS Survey Guidelines) Response Biases in Standardised Surveys (GESIS Survey Guidelines) [Internet]. GESIS - Leibniz Institute for the Social Sciences; 2016 [cited 2024 Nov 18]. Available from: https://www.gesis.org/fileadmin/admin/Dateikatalog/pdf/guidelines/response_biases_standardized_surveys_bogner_landrock_2016.pdf
241. Pearson H. Can AI review the scientific literature — and figure out what it all means? *Nature* [Internet]. 2024 Nov 13 [cited 2024 Nov 23];635(8038):276–8. Available from: <https://www.nature.com/articles/d41586-024-03676-9>
242. Ross-White A, Godfrey C. Is there an optimum number needed to retrieve to justify inclusion of a database in a systematic review search? *Health Inf Libr J* [Internet]. 2017 [cited 2024 Nov 24];34(3):217–24. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/hir.12185>
243. Booth A. How much searching is enough? Comprehensive versus optimal retrieval for technology assessments. *Int J Technol Assess Health Care* [Internet]. 2010 Oct [cited 2024 Nov 24];26(4):431–5. Available from: <https://www.cambridge.org/core/journals/international-journal-of-technology-assessment-in-health-care/article/abs/how-much-searching-is-enough-comprehensive-versus-optimal-retrieval-for-technology-assessments/420FE198E07703FB1AAFE762958CFE4A>
244. Ballard M, Montgomery P. Risk of bias in overviews of reviews: a scoping review of methodological guidance and four-item checklist. *Res Synth Methods* [Internet]. 2017 [cited 2024 Nov 24];8(1):92–108. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/jrsm.1229>
245. Mohseni M, Ameri H, Arab-Zozani M. Potential limitations in systematic review studies assessing the effect of the main intervention for treatment/therapy of COVID-19 patients: An overview. *Front Med* [Internet]. 2022 Sep 15 [cited 2024 Nov 24];9. Available from: <https://www.frontiersin.org/journals/medicine/articles/10.3389/fmed.2022.966632/full>
246. Borenstein M, editor. Introduction to meta-analysis. Nachdr. Chichester: Wiley; 2013. 421 p.

247. Zhang J, Yu KF. What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes. *JAMA*. 1998 Nov 18;280(19):1690–1.
248. Ellis PD. *The Essential Guide to Effect Sizes: Statistical Power, Meta-Analysis, and the Interpretation of Research Results*. Cambridge University Press; 2010. 193 p.
249. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. *Introduction to Meta-Analysis*. John Wiley & Sons; 2021. 547 p.
250. Lu Y, Cen W, Wang W, Huang Y, Chen H. How big is a big hazard ratio in clinical trials? *Int J Clin Trials* [Internet]. 2023 Jul 26 [cited 2024 Nov 24];10(3):195–200. Available from: <https://www.ijclinicaltrials.com/index.php/ijct/article/view/592>
251. Stites SD, Cao H, James R, Harkins K, Coykendall C, Flatt JD. A systematic review of measures of gender and biological sex: Exploring candidates for Alzheimer's disease and related dementias (AD/ADRD) research. *Alzheimers Dement Diagn Assess Dis Monit* [Internet]. 2023 [cited 2024 Nov 6];15(1):e12359. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/dad2.12359>
252. Sousa-Pinto B, Anto A, Berger M, Dramburg S, Pfaar O, Klimek L, et al. Real-world data using mHealth apps in rhinitis, rhinosinusitis and their multimorbidities. *Clin Transl Allergy*. 2022 Nov;12(11):e12208.
253. Wang N, Chen J, Chen W, Shi Z, Yang H, Liu P, et al. The effectiveness of case management for cancer patients: an umbrella review. *BMC Health Serv Res* [Internet]. 2022 Oct 14 [cited 2024 Nov 6];22(1):1247. Available from: <https://doi.org/10.1186/s12913-022-08610-1>
254. Hoffmann I, Shojaa M, Kohl M, von Stengel S, Becker C, Gosch M, et al. Exercise Reduces the Number of Overall and Major Osteoporotic Fractures in Adults. Does Supervision Make a Difference? Systematic Review and Meta-Analysis. *J Bone Miner Res* [Internet]. 2022 Nov 1 [cited 2024 Nov 6];37(11):2132–48. Available from: <https://doi.org/10.1002/jbmr.4683>
255. Lei XG, Huang Z, Rashid T, Yang X. Effectiveness of mHealth/eHealth interventions on obesity treatment: a protocol for umbrella review of meta-analyses. *BMJ Open* [Internet]. 2022 Jun 1 [cited 2024 Nov 6];12(6):e052443. Available from: <https://bmjopen.bmj.com/content/12/6/e052443>
256. Kast S, Shojaa M, Kohl M, von Stengel S, Gosch M, Jakob F, et al. Effects of different exercise intensity on bone mineral density in adults: a comparative systematic review and meta-analysis. *Osteoporos Int* [Internet]. 2022 Aug 1 [cited 2024 Nov 6];33(8):1643–57. Available from: <https://doi.org/10.1007/s00198-022-06329-7>
257. Chen Z, Huang Y, Cao D, Qiu S, Chen B, Li J, et al. Vitamin C Intake and Cancers: An Umbrella Review. *Front Nutr* [Internet]. 2022 Jan 20 [cited 2024 Nov 6];8. Available from: <https://www.frontiersin.org/journals/nutrition/articles/10.3389/fnut.2021.812394/full>
258. Tannou T, Lihoreau T, Gagnon-Roy M, Grondin M, Bier N. Effectiveness of smart living environments to support older adults to age in place in their community: an umbrella review protocol. *BMJ Open*. 2022 Jan 25;12(1):e054235.
259. Martínez-García M, Castrejón-Pérez RC, Rodríguez-Hernández AP, Sandoval-Motta S, Vallejo M, Borges-Yáñez SA, et al. Incidence of Arterial Hypertension in People With Periodontitis and Characterization of the Oral and Subgingival Microbiome: A Study Protocol. *Front Cardiovasc Med* [Internet]. 2022 Jan 7 [cited 2024 Nov 6];8. Available from: <https://www.frontiersin.org/journals/cardiovascular-medicine/articles/10.3389/fcvm.2021.763293/full>
260. Chauhan S, Kumar S, Bharti R, Patel R. Prevalence and determinants of activity of daily living and instrumental activity of daily living among elderly in India. *BMC Geriatr* [Internet]. 2022 Jan 19 [cited 2024 Nov 6];22(1):64. Available from: <https://doi.org/10.1186/s12877-021-02659-z>
261. Kroke A, Schmidt A, Amini AM, Kalotai N, Lehmann A, Haardt J, et al. Dietary protein intake

and health-related outcomes: a methodological protocol for the evidence evaluation and the outline of an evidence to decision framework underlying the evidence-based guideline of the German Nutrition Society. *Eur J Nutr* [Internet]. 2022 Jun 1 [cited 2024 Nov 6];61(4):2091–101. Available from: <https://doi.org/10.1007/s00394-021-02789-5>

262. Mages M, Shojaa M, Kohl M, von Stengel S, Becker C, Gosch M, et al. Exercise Effects on Bone Mineral Density in Men. *Nutrients* [Internet]. 2021 Dec [cited 2024 Nov 6];13(12):4244. Available from: <https://www.mdpi.com/2072-6643/13/12/4244>
263. Moretto L, Herzig R, Beckmann K, Wolfer N, Dennler M, Glaus TM. Reversible facial nerve paralysis in a cat suspected to be associated with systemic hypertension. *JFMS Open Rep*. 2021;7(2):20551169211063454.
264. Zitzmann AL, Shojaa M, Kast S, Kohl M, von Stengel S, Borucki D, et al. The effect of different training frequency on bone mineral density in older adults. A comparative systematic review and meta-analysis. *Bone* [Internet]. 2022 Jan 1 [cited 2024 Nov 6];154:116230. Available from: <https://www.sciencedirect.com/science/article/pii/S8756328221003963>
265. Sawyer ADM, van Lenthe F, Kamphuis CBM, Terragni L, Roos G, Poelman MP, et al. Dynamics of the complex food environment underlying dietary intake in low-income groups: a systems map of associations extracted from a systematic umbrella literature review. *Int J Behav Nutr Phys Act* [Internet]. 2021 Jul 13 [cited 2024 Nov 6];18(1):96. Available from: <https://doi.org/10.1186/s12966-021-01164-1>
266. Michalski E, Ostrowska M, Gietka P, Sudol-Szopińska I. Magnetic resonance imaging of the knee joint in juvenile idiopathic arthritis. *Reumatologia* [Internet]. 2020 Dec 23 [cited 2024 Nov 6];58(6):416–23. Available from: <https://reu.termedia.pl/Magnetic-resonance-imaging-of-the-knee-joint-in-juvenile-idiopathic-arthritis,131007,0,2.html>
267. Lauretani F, Longobucco Y, Ferrari Pellegrini F, De Iorio AM, Fazio C, Federici R, et al. Comprehensive Model for Physical and Cognitive Frailty: Current Organization and Unmet Needs. *Front Psychol* [Internet]. 2020 Nov 26 [cited 2024 Nov 6];11. Available from: <https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2020.569629/full>
268. Toward a theory-based specification of non-pharmacological treatments in aging and dementia: Focused reviews and methodological recommendations. *Alzheimers Dement* [Internet]. 2021 [cited 2024 Nov 6];17(2):255–70. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/alz.12188>
269. Florisson S, Aagesen EK, Bertelsen AS, Nielsen LP, Rosholm JU. Are older adults insufficiently included in clinical trials?—An umbrella review. *Basic Clin Pharmacol Toxicol* [Internet]. 2021 [cited 2024 Nov 6];128(2):213–23. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/bcpt.13536>
270. Falandry C, Stefani L, Andre L, Granger M, Barbavara C, Habchi H, et al. Interventions to improve physical performances of older people with cancer before complex medico-surgical procedures: Protocol for an umbrella review of systematic reviews and meta-analyses. *Medicine (Baltimore)* [Internet]. 2020 Sep 25 [cited 2024 Nov 6];99(39):e21780. Available from: https://journals.lww.com/md-journal/fulltext/2020/09250/interventions_to_improve_physical_performances_of.2.aspx
271. Gremese E, Ferraccioli ES, Alivernini S, Tolusso B, Ferraccioli G. Basic immunology may lead to translational therapeutic rationale: SARS-CoV-2 and rheumatic diseases. *Eur J Clin Invest* [Internet]. 2020 [cited 2024 Nov 6];50(9):e13342. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/eci.13342>
272. Soysal P, Hurst C, Demurtas J, Firth J, Howden R, Yang L, et al. Handgrip strength and health outcomes: Umbrella review of systematic reviews with meta-analyses of observational studies. *J Sport Health Sci* [Internet]. 2021 May 1 [cited 2024 Nov 6];10(3):290–5. Available from: <https://www.sciencedirect.com/science/article/pii/S2095254620300752>

273. Solmi M, Veronese N, Galvano D, Favaro A, Ostinelli EG, Noventa V, et al. Factors Associated With Loneliness: An Umbrella Review Of Observational Studies. *J Affect Disord* [Internet]. 2020 Jun 15 [cited 2024 Nov 6];271:131–8. Available from: <https://www.sciencedirect.com/science/article/pii/S0165032719335803>
274. Maretzke F, Bechthold A, Egert S, Ernst JB, Melo van Lent D, Pilz S, et al. Role of Vitamin D in Preventing and Treating Selected Extraskelatal Diseases—An Umbrella Review. *Nutrients* [Internet]. 2020 Apr [cited 2024 Nov 6];12(4):969. Available from: <https://www.mdpi.com/2072-6643/12/4/969>
275. McGrath R, Johnson N, Klawitter L, Mahoney S, Trautman K, Carlson C, et al. What are the association patterns between handgrip strength and adverse health conditions? A topical review. *SAGE Open Med*. 2020;8:2050312120910358.
276. Biondi-Zoccai G, Garmendia CM, Abbate A, Giordano A, Frati G, Sciarretta S, et al. Atherothrombosis Prevention and Treatment with Anti-interleukin-1 Agents. *Curr Atheroscler Rep* [Internet]. 2020 Jan 13 [cited 2024 Nov 6];22(1):4. Available from: <https://doi.org/10.1007/s11883-020-0819-1>
277. Wangmo T, Lipps M, Kressig RW, Ienca M. Ethical concerns with the use of intelligent assistive technology: findings from a qualitative study with professional stakeholders. *BMC Med Ethics* [Internet]. 2019 Dec 19 [cited 2024 Nov 6];20(1):98. Available from: <https://doi.org/10.1186/s12910-019-0437-z>
278. Roller-Wirnsberger R, Zitta S, Herzog C, Dornan H, Lindner S, Rehatschek H, et al. Massive open online courses (MOOCs) for long-distance education in geriatric medicine across Europe. *Eur Geriatr Med* [Internet]. 2019 Dec 1 [cited 2024 Nov 6];10(6):989–94. Available from: <https://doi.org/10.1007/s41999-019-00252-7>
279. Soysal P, Veronese N, Smith L, Torbahn G, Jackson SE, Yang L, et al. Orthostatic hypotension and health outcomes: an umbrella review of observational studies. *Eur Geriatr Med* [Internet]. 2019 Dec 1 [cited 2024 Nov 6];10(6):863–70. Available from: <https://doi.org/10.1007/s41999-019-00239-4>
280. MacEntee MI. A typology of systematic reviews for synthesising evidence on health care. *Gerodontology* [Internet]. 2019 [cited 2024 Nov 6];36(4):303–12. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/ger.12439>
281. Kivnick HQ, Driessen MC, Santavasy C, Wardwell C, Duncan Davis L. ‘Who’s Been Putting Socks in My Drawer?’ Narrative Case Study of an Elder Role Model. *The Gerontologist*. 2020 Jul 15;60(5):831–40.
282. Grabovac I, Veronese N, Stefanac S, Haider S, Jackson SE, Koyanagi A, et al. Human Immunodeficiency Virus Infection and Diverse Physical Health Outcomes: An Umbrella Review of Meta-analyses of Observational Studies. *Clin Infect Dis* [Internet]. 2020 Apr 15 [cited 2024 Nov 6];70(9):1809–15. Available from: <https://doi.org/10.1093/cid/ciz539>
283. Pescatello LS, Buchner DM, Jakicic JM, Powell KE, Kraus WE, Bloodgood B, et al. Physical Activity to Prevent and Treat Hypertension: A Systematic Review. *Med Sci Sports Exerc* [Internet]. 2019 Jun [cited 2024 Nov 6];51(6):1314. Available from: https://journals.lww.com/acsm-msse/fulltext/2019/06000/physical_activity_to_prevent_and_treat.26.aspx
284. Chen L, Ye M, Kahana E. A Self-Reliant Umbrella: Defining Successful Aging Among the Old-Old (80+) in Shanghai. *J Appl Gerontol* [Internet]. 2020 Mar 1 [cited 2024 Nov 6];39(3):242–9. Available from: <https://doi.org/10.1177/0733464819842500>
285. Sharka R, Abed H, Hector M. Oral health-related quality of life and satisfaction of edentulous patients using conventional complete dentures and implant-retained overdentures: An umbrella systematic review. *Gerodontology* [Internet]. 2019 [cited 2024 Nov 6];36(3):195–204. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/ger.12399>

286. Bangsbo J, Blackwell J, Boraxbekk CJ, Caserotti P, Dela F, Evans AB, et al. Copenhagen Consensus statement 2019: physical activity and ageing. *Br J Sports Med* [Internet]. 2019 Jul 1 [cited 2024 Nov 6];53(14):856–8. Available from: <https://bjsm.bmj.com/content/53/14/856>
287. Smith L, Luchini C, Demurtas J, Soysal P, Stubbs B, Hamer M, et al. Telomere length and health outcomes: An umbrella review of systematic reviews and meta-analyses of observational studies. *Ageing Res Rev* [Internet]. 2019 May 1 [cited 2024 Nov 6];51:1–10. Available from: <https://www.sciencedirect.com/science/article/pii/S1568163718303234>
288. Smith AT, Han JH. Altered Mental Status in the Emergency Department. *Semin Neurol* [Internet]. 2019 Feb [cited 2024 Nov 6];39(01):005–19. Available from: <http://www.thieme-connect.de/DOI/DOI?10.1055/s-0038-1677035>
289. Zou R, Kuang G, Ågren H, Nordberg A, Långström B, Tu Y. Free Energy Profile for Penetration of Pittsburgh Compound-B into the Amyloid β Fibril. *ACS Chem Neurosci* [Internet]. 2019 Mar 20 [cited 2024 Nov 6];10(3):1783–90. Available from: <https://doi.org/10.1021/acscchemneuro.8b00662>
290. Veronese N, Demurtas J, Pesolillo G, Celotto S, Barnini T, Calusi G, et al. Magnesium and health outcomes: an umbrella review of systematic reviews and meta-analyses of observational and intervention studies. *Eur J Nutr* [Internet]. 2020 Feb 1 [cited 2024 Nov 6];59(1):263–72. Available from: <https://doi.org/10.1007/s00394-019-01905-w>
291. Vink A, Hanser S. Music-Based Therapeutic Interventions for People with Dementia: A Mini-Review. *Medicines* [Internet]. 2018 Dec [cited 2024 Nov 6];5(4):109. Available from: <https://www.mdpi.com/2305-6320/5/4/109>
292. Veronese N, Demurtas J, Celotto S, Caruso MG, Maggi S, Bolzetta F, et al. Is chocolate consumption associated with health outcomes? An umbrella review of systematic reviews and meta-analyses. *Clin Nutr* [Internet]. 2019 Jun 1 [cited 2024 Nov 6];38(3):1101–8. Available from: [https://www.clinicalnutritionjournal.com/article/S0261-5614\(18\)30204-8/abstract](https://www.clinicalnutritionjournal.com/article/S0261-5614(18)30204-8/abstract)
293. Veronese N, Solmi M, Caruso MG, Giannelli G, Osella AR, Evangelou E, et al. Dietary fiber and health outcomes: an umbrella review of systematic reviews and meta-analyses. *Am J Clin Nutr* [Internet]. 2018 Mar 1 [cited 2024 Nov 6];107(3):436–44. Available from: <https://www.sciencedirect.com/science/article/pii/S0002916522028131>
294. Ying J, Yap P, Gandhi M, Liew TM. Iterating a framework for the prevention of caregiver depression in dementia: a multi-method approach. *Int Psychogeriatr* [Internet]. 2018 Aug [cited 2024 Nov 6];30(8):1119–30. Available from: <https://www.cambridge.org/core/journals/international-psychogeriatrics/article/iterating-a-framework-for-the-prevention-of-caregiver-depression-in-dementia-a-multimethod-approach/FECD9DEBE7D3B4BC8CAD45DAE0465314>
295. Parker SG, McCue P, Phelps K, McCleod A, Arora S, Nockels K, et al. What is Comprehensive Geriatric Assessment (CGA)? An umbrella review. *Age Ageing*. 2018 Jan 1;47(1):149–55.
296. Finnigan-Fox G, Matlock DD, Tate CE, Knoepke CE, Allen LA. Hospice, She Yelped: Examining the Quantity and Quality of Decision Support Available to Patient and Families Considering Hospice. *J Pain Symptom Manage*. 2017 Dec;54(6):916-921.e1.
297. Catalá-López F, Hutton B, Driver JA, Ridao M, Valderas JM, Gènova-Maleras R, et al. Anorexia nervosa and cancer: a protocol for a systematic review and meta-analysis of observational studies. *Syst Rev* [Internet]. 2017 Jul 11 [cited 2024 Nov 6];6(1):137. Available from: <https://doi.org/10.1186/s13643-017-0540-5>
298. Shenkin SD, Harrison JK, Wilkinson T, Dodds RM, Ioannidis JPA. Systematic reviews: guidance relevant for studies of older people. *Age Ageing*. 2017 Sep 1;46(5):722–8.
299. Bortolato B, Köhler CA, Evangelou E, León-Caballero J, Solmi M, Stubbs B, et al. Systematic assessment of environmental risk factors for bipolar disorder: an umbrella review of systematic reviews and meta-analyses. *Bipolar Disord* [Internet]. 2017 [cited 2024 Nov 6];19(2):84–96. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/bdi.12490>

300. Apóstolo J, Cooke R, Bobrowicz-Campos E, Santana S, Marcucci M, Cano A, et al. Predicting risk and outcomes for frail older adults: a protocol for an umbrella review of available frailty screening tools. *JBIS Evid Synth* [Internet]. 2015 Dec [cited 2024 Nov 6];13(12):14. Available from: https://journals.lww.com/jbisir/fulltext/2015/13120/predicting_risk_and_outcomes_for_frail_older.3.aspx
301. Catalá-López F, Hutton B, Driver JA, Page MJ, Ridao M, Valderas JM, et al. Cancer and central nervous system disorders: protocol for an umbrella review of systematic reviews and updated meta-analyses of observational studies. *Syst Rev* [Internet]. 2017 Apr 4 [cited 2024 Nov 6];6(1):69. Available from: <https://doi.org/10.1186/s13643-017-0466-y>
302. Evenepoel P, Cunningham J, Ferrari S, Haarhaus M, Javaid MK, Lafage-Proust MH, et al. European Consensus Statement on the diagnosis and management of osteoporosis in chronic kidney disease stages G4-G5D. *Nephrol Dial Transplant Off Publ Eur Dial Transpl Assoc - Eur Ren Assoc*. 2021 Jan 1;36(1):42–59.
303. Bakker TJEM. Er is disciplinebreed sprake van relatieve onbevoegdheid en onbekwaamheid op het gebied van kwetsbare ouderen. *Tijdschr Gerontol Geriatr* [Internet]. 2016 Oct 14 [cited 2024 Nov 6];(5). Available from: <https://tvvgg.nl/article/view/16396>
304. Hategan A, Bourgeois JA, Goldberg J. Aging with autism spectrum disorder: an emerging public health problem. *Int Psychogeriatr* [Internet]. 2017 Apr [cited 2024 Nov 6];29(4):695–7. Available from: <https://www.cambridge.org/core/journals/international-psychogeriatrics/article/aging-with-autism-spectrum-disorder-an-emerging-public-health-problem/EFD780ABDF27D8BBD8E4A814CD995968>
305. Hermans K, Spruytte N, Cohen J, Van Audenhove C, Declercq A. Usefulness, feasibility and face validity of the interRAI Palliative Care instrument according to care professionals in nursing homes: A qualitative study. *Int J Nurs Stud* [Internet]. 2016 Oct 1 [cited 2024 Nov 6];62:90–9. Available from: <https://www.sciencedirect.com/science/article/pii/S0020748916300980>
306. Tamaki K, Ishigaki S, Ogawa T, Oguchi H, Kato T, Suganuma T, et al. Japan Prosthodontic Society position paper on “occlusal discomfort syndrome”. *J Prosthodont Res* [Internet]. 2016 Jul 1 [cited 2024 Nov 6];60(3):156–66. Available from: <https://www.sciencedirect.com/science/article/pii/S1883195815001048>
307. Apostolo J, Cooke R, Bobrowicz-Campos E, Santana S, Marcucci M, Cano A, et al. Predicting risk and outcomes for frail older adults: a protocol for an umbrella review of available frailty screening tools. *JBIS Database Syst Rev Implement Rep* [Internet]. 2015 [cited 2024 Nov 24];13(12):14–24. Available from: <http://joannabriggslibrary.org/index.php/jbisir/article/view/2468>
308. Wan W, Cao L, Kalionis B, Xia S, Tai X. Applications of Induced Pluripotent Stem Cells in Studying the Neurodegenerative Diseases. *Stem Cells Int* [Internet]. 2015 [cited 2024 Nov 6];2015(1):382530. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1155/2015/382530>
309. Morris JN, Howard EP, Steel K, Schreiber R, Fries BE, Lipsitz LA, et al. Predicting risk of hospital and emergency department use for home care elderly persons through a secondary analysis of cross-national data. *BMC Health Serv Res* [Internet]. 2014 Nov 14 [cited 2024 Nov 6];14(1):519. Available from: <https://doi.org/10.1186/s12913-014-0519-z>
310. Dybedal GS, Tanum L, Sundet K, Gaarden TL, Bjølseth TM. Neuropsychological functioning in late-life depression. *Front Psychol* [Internet]. 2013 Jun 27 [cited 2024 Nov 6];4. Available from: <https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2013.00381/full>
311. Cherry KE, Walker EJ, Brown JS, Volaufova J, LaMotte LR, Welsh DA, et al. Social engagement and health in younger, older, and oldest-old adults in the Louisiana Healthy Aging Study. *J Appl Gerontol Off J South Gerontol Soc*. 2013 Feb;32(1):51–75.
312. Lee UJ, Scott VCS, Rashid R, Behniwal A, Wein AJ, Maliski SL, et al. Defining and Managing

- Overactive Bladder: Disagreement Among the Experts. *Urology* [Internet]. 2013 Feb 1 [cited 2024 Nov 6];81(2):257–62. Available from: [https://www.goldjournal.net/article/S0090-4295\(12\)01178-8/abstract](https://www.goldjournal.net/article/S0090-4295(12)01178-8/abstract)
313. Sandoval C, True C. Equine Wellness Care in Ambulatory Practice. *Vet Clin North Am Equine Pract* [Internet]. 2012 Apr 1 [cited 2024 Nov 6];28(1):189–205. Available from: <https://www.sciencedirect.com/science/article/pii/S0749073912000090>
314. Porter EJ, Lasiter S. Reducing My Risk of Intrusion: An Intention of Old Homebound Women Who Live Alone. *Nurs Res* [Internet]. 2008 Oct [cited 2024 Nov 6];57(5):351. Available from: https://journals.lww.com/nursingresearchonline/abstract/2008/09000/reducing_my_risk_of_intrusion__an_intention_of_old.7.aspx
315. Collaboration Among Counties, Consumers, Professionals and Organizations. *J Gerontol Soc Work* [Internet]. 2006 Dec 20 [cited 2024 Nov 6];48(1–2):193–202. Available from: https://doi.org/10.1300/J083v48n01_13
316. Auger C, Demers L, Swaine B. Making sense of pragmatic criteria for the selection of geriatric rehabilitation measurement tools. *Arch Gerontol Geriatr* [Internet]. 2006 Jul 1 [cited 2024 Nov 6];43(1):65–83. Available from: <https://www.sciencedirect.com/science/article/pii/S0167494305001111>
317. Sonnenblick M. [The geriatric department in a general hospital]. *Harefuah*. 2004 Apr 1;143(4):291–4, 317.
318. McHorney CA. Ten recommendations for advancing patient-centered outcomes measurement for older persons. *Ann Intern Med*. 2003 Sep 2;139(5 Pt 2):403–9.
319. Yeaworth RC, Burke WJ. Frontotemporal dementia: A different kind of dementia. *Arch Psychiatr Nurs* [Internet]. 2000 Oct 1 [cited 2024 Nov 6];14(5):249–53. Available from: [https://www.psychiatricnursing.org/article/S0883-9417\(00\)42780-9/abstract](https://www.psychiatricnursing.org/article/S0883-9417(00)42780-9/abstract)
320. Robben PBM, Olde Rikkert MGM, Hutschemaekers GJM. Beroepsvorming in tweevoud: de klinisch geriater en de verpleeghuisarts, 1945-1990. 63 [Internet]. 1999 [cited 2024 Nov 6]; Available from: <https://repository.ubn.ru.nl/handle/2066/265906>
321. Stessman J, Hammerman-Rozenberg R, Cohen A. Home hospitalization in the spectrum of community geriatric care. *Disabil Rehabil* [Internet]. 1997 Jan 1 [cited 2024 Nov 6];19(4):134–41. Available from: <https://doi.org/10.3109/09638289709166517>
322. Morse R, Jenkinson D. Achieving High Quality Long-Term Care for Elderly People: Consumers' Wishes and Providers' Responsibilities. *J R Coll Physicians Lond* [Internet]. 1995 Jun [cited 2024 Nov 6];29(3):239. Available from: <https://pmc.ncbi.nlm.nih.gov/articles/PMC5401196/>
323. Hall JA, Epstein AM, McNeil BJ. Multidimensionality of health status in an elderly population. Construct validity of a measurement battery. *Med Care*. 1989 Mar;27(3 Suppl):S168-177.
324. Brosseau GE. 'Umbrella' projects are questioning what is, what will be, what should be the status of the aged. *Geriatrics*. 1975 Aug;30(8):34–43.
325. Farhangi MA, Nikniaz L, Nikniaz Z, Dehghan P. Dietary inflammatory index potentially increases blood pressure and markers of glucose homeostasis among adults: findings from an updated systematic review and meta-analysis. *Public Health Nutr*. 2020 Jun;23(8):1362–80.
326. Hua R, Liang G, Yang F. Meta-analysis of the association between dietary inflammatory index (DII) and upper aerodigestive tract cancer risk. *Medicine (Baltimore)*. 2020 Apr;99(17):e19879.
327. Zahedi H, Djalalinia S, Asayesh H, Mansourian M, Esmaeili Abdar Z, Mahdavi Gorabi A, et al. A Higher Dietary Inflammatory Index Score is Associated with a Higher Risk of Incidence and Mortality of Cancer: A Comprehensive Systematic Review and Meta-Analysis. *Int J Prev Med*. 2020;11:15.

328. Jayedi A, Emadi A, Shab-Bidar S. Dietary Inflammatory Index and Site-Specific Cancer Risk: A Systematic Review and Dose-Response Meta-Analysis. *Adv Nutr* [Internet]. 2018 Jul 1 [cited 2024 Nov 3];9(4):388–403. Available from: <https://www.sciencedirect.com/science/article/pii/S2161831322012376>
329. Guo Z, Hong Y, Cheng Y. Dietary inflammatory index and pancreatic cancer risk: a systematic review and dose-response meta-analysis. *Public Health Nutr*. 2021 Dec;24(18):6427–35.
330. Chen QJ, Ou L, Li K, Ou FR. Meta-analysis of the relationship between Dietary Inflammatory Index and esophageal cancer risk. *Medicine (Baltimore)*. 2020 Dec 4;99(49):e23539.
331. Wu PY, Huang KS, Chen KM, Chou CP, Tu YK. Exercise, Nutrition, and Combined Exercise and Nutrition in Older Adults with Sarcopenia: A Systematic Review and Network Meta-analysis. *Maturitas*. 2021 Mar;145:38–48.
332. Emadzadeh M, Sahebi R, Khedmatgozar H, Sadeghi R, Farjami M, Sharifan P, et al. A systematic review and meta-analysis of the effect of Vitamin D-fortified food on glycemic indices. *BioFactors Oxf Engl*. 2020 Jul;46(4):502–13.
333. Zhang Y, Xue Y, Zhang D, Liu Y, Xu Z, Gao J, et al. Effect of Vitamin D Supplementation on Glycemic Control in Prediabetes: A Meta-Analysis. *Nutrients*. 2021 Dec 14;13(12):4464.
334. Tang H, Li D, Li Y, Zhang X, Song Y, Li X. Effects of Vitamin D Supplementation on Glucose and Insulin Homeostasis and Incident Diabetes among Nondiabetic Adults: A Meta-Analysis of Randomized Controlled Trials. *Int J Endocrinol*. 2018;2018:7908764.
335. Li X, Liu Y, Zheng Y, Wang P, Zhang Y. The Effect of Vitamin D Supplementation on Glycemic Control in Type 2 Diabetes Patients: A Systematic Review and Meta-Analysis. *Nutrients*. 2018 Mar 19;10(3):375.
336. Milajerdi A, Ostadmohammadi V, Amirjani S, Kolahdooz F, Asemi Z. The effects of vitamin D treatment on glycemic control, serum lipid profiles, and C-reactive protein in patients with chronic kidney disease: a systematic review and meta-analysis of randomized controlled trials. *Int Urol Nephrol*. 2019 Sep;51(9):1567–80.
337. Qi KJ, Zhao ZT, Zhang W, Yang F. The impacts of vitamin D supplementation in adults with metabolic syndrome: A systematic review and meta-analysis of randomized controlled trials. *Front Pharmacol*. 2022;13:1033026.
338. Teoh SL, Letchumanan V, Lee LH. Can Mindfulness Help to Alleviate Loneliness? A Systematic Review and Meta-Analysis. *Front Psychol*. 2021;12:633319.
339. Hoang P, King JA, Moore S, Moore K, Reich K, Sidhu H, et al. Interventions Associated With Reduced Loneliness and Social Isolation in Older Adults: A Systematic Review and Meta-analysis. *JAMA Netw Open*. 2022 Oct 3;5(10):e2236676.
340. Orr N, Abbott R, Bethel A, Paviour S, Whear R, Garside R, et al. What are the effects of animals on the health and wellbeing of residents in care homes? A systematic review of the qualitative and quantitative evidence. *BMC Geriatr*. 2023 Mar 25;23(1):170.
341. Grillich L, Titscher V, Klingenstein P, Kostial E, Emprechtinger R, Klerings I, et al. The effectiveness of interventions to prevent loneliness and social isolation in the community-dwelling and old population: an overview of systematic reviews and meta-analysis. *Eur J Public Health*. 2023 Apr 1;33(2):235–41.
342. Fu Z, Yan M, Meng C. The effectiveness of remote delivered intervention for loneliness reduction in older adults: A systematic review and meta-analysis. *Front Psychol*. 2022;13:935544.
343. Magni NE, McNair PJ, Rice DA. The effects of resistance training on muscle strength, joint pain, and hand function in individuals with hand osteoarthritis: a systematic review and meta-analysis. *Arthritis Res Ther*. 2017 Jun 13;19(1):131.

344. Kjekken I, Smedslund G, Moe RH, Slatkowsky-Christensen B, Uhlig T, Hagen KB. Systematic review of design and effects of splints and exercise programs in hand osteoarthritis. *Arthritis Care Res*. 2011 Jun;63(6):834–48.
345. Østerås N, Kjekken I, Smedslund G, Moe RH, Slatkowsky-Christensen B, Uhlig T, et al. Exercise for hand osteoarthritis. *Cochrane Database Syst Rev*. 2017 Jan 31;1(1):CD010388.
346. Østerås N, Kjekken I, Smedslund G, Moe RH, Slatkowsky-Christensen B, Uhlig T, et al. Exercise for Hand Osteoarthritis: A Cochrane Systematic Review. *J Rheumatol*. 2017 Dec;44(12):1850–8.
347. Buhler M, Chapple CM, Stebbings S, Sangelaji B, Baxter GD. Effectiveness of splinting for pain and function in people with thumb carpometacarpal osteoarthritis: a systematic review with meta-analysis. *Osteoarthritis Cartilage*. 2019 Apr;27(4):547–59.
348. Hamasaki T, Laprise S, Harris PG, Bureau NJ, Gaudreault N, Ziegler D, et al. Efficacy of Nonsurgical Interventions for Trapeziometacarpal (Thumb Base) Osteoarthritis: A Systematic Review. *Arthritis Care Res*. 2020 Dec;72(12):1719–35.
349. Meireles SM, Jones A, Natour J. Orthosis for rhizarthrosis: A systematic review and meta-analysis. *Semin Arthritis Rheum*. 2019 Apr;48(5):778–90.
350. Leyland KM, Gates LS, Sanchez-Santos MT, Nevitt MC, Felson D, Jones G, et al. Knee osteoarthritis and time-to all-cause mortality in six community-based cohorts: an international meta-analysis of individual participant-level data. *Aging Clin Exp Res*. 2021 Mar;33(3):529–45.
351. Veronese N, Cereda E, Maggi S, Luchini C, Solmi M, Smith T, et al. Osteoarthritis and mortality: A prospective cohort study and systematic review with meta-analysis. *Semin Arthritis Rheum*. 2016 Oct;46(2):160–7.
352. Macêdo MB, Santos VMOS, Pereira RMR, Fuller R. Association between osteoarthritis and atherosclerosis: A systematic review and meta-analysis. *Exp Gerontol*. 2022 May;161:111734.
353. Wang H, Bai J, He B, Hu X, Liu D. Osteoarthritis and the risk of cardiovascular disease: a meta-analysis of observational studies. *Sci Rep*. 2016 Dec 22;6:39672.
354. Namazi N, Larijani B, Azadbakht L. Dietary Inflammatory Index and its Association with the Risk of Cardiovascular Diseases, Metabolic Syndrome, and Mortality: A Systematic Review and Meta-Analysis. *Horm Metab Res Horm Stoffwechselforschung Horm Metab*. 2018 May;50(5):345–58.
355. Bao W, Sun Y, Zhang T, Zou L, Wu X, Wang D, et al. Exercise Programs for Muscle Mass, Muscle Strength and Physical Performance in Older Adults with Sarcopenia: A Systematic Review and Meta-Analysis. *Aging Dis*. 2020 Jul;11(4):863–73.
356. Yoshimura Y, Wakabayashi H, Yamada M, Kim H, Harada A, Arai H. Interventions for Treating Sarcopenia: A Systematic Review and Meta-Analysis of Randomized Controlled Studies. *J Am Med Dir Assoc*. 2017 Jun 1;18(6):553.e1-553.e16.
357. Yin YH, Liu JYW, Välimäki M. Effectiveness of non-pharmacological interventions on the management of sarcopenic obesity: A systematic review and meta-analysis. *Exp Gerontol*. 2020 Jul 1;135:110937.
358. Fealy G, McCarron M, O'Neill D, McCallion P, Clarke M, Small V, et al. Effectiveness of gerontologically informed nursing assessment and referral interventions for older persons attending the emergency department: systematic review. *J Adv Nurs*. 2009 May;65(5):934–5.
359. Karam G, Radden Z, Berall LE, Cheng C, Gruneir A. Efficacy of emergency department-based interventions designed to reduce repeat visits and other adverse outcomes for older patients after discharge: A systematic review. *Geriatr Gerontol Int*. 2015 Sep;15(9):1107–17.

360. Hastings SN, Heflin MT. A systematic review of interventions to improve outcomes for elders discharged from the emergency department. *Acad Emerg Med Off J Soc Acad Emerg Med*. 2005 Oct;12(10):978–86.
361. Hughes JM, Freiermuth CE, Shepherd-Banigan M, Ragsdale L, Eucker SA, Goldstein K, et al. Emergency Department Interventions for Older Adults: A Systematic Review. *J Am Geriatr Soc*. 2019 Jul;67(7):1516–25.
362. Lowthian JA, McGinnes RA, Brand CA, Barker AL, Cameron PA. Discharging older patients from the emergency department effectively: a systematic review and meta-analysis. *Age Ageing*. 2015 Sep;44(5):761–70.
363. Malik M, Moore Z, Patton D, O'Connor T, Nugent LE. The impact of geriatric focused nurse assessment and intervention in the emergency department: A systematic review. *Int Emerg Nurs*. 2018 Mar;37:52–60.
364. Morello RT, Soh SE, Behm K, Egan A, Ayton D, Hill K, et al. Multifactorial falls prevention programmes for older adults presenting to the emergency department with a fall: systematic review and meta-analysis. *Inj Prev J Int Soc Child Adolesc Inj Prev*. 2019 Dec;25(6):557–64.
365. Abbott R, Orr N, McGill P, Whear R, Bethel A, Garside R, et al. How do 'robopets' impact the health and well-being of residents in care homes? A systematic review of qualitative and quantitative evidence. *Int J Older People Nurs*. 2019 Sep;14(3):e12239.
366. Choi M, Kong S, Jung D. Computer and internet interventions for loneliness and depression in older adults: a meta-analysis. *Healthc Inform Res*. 2012 Sep;18(3):191–8.
367. Cohen-Mansfield J, Perach R. Interventions for alleviating loneliness among older persons: a critical review. *Am J Health Promot AJHP*. 2015;29(3):e109-125.
368. Hagan R, Manktelow R, Taylor BJ, Mallett J. Reducing loneliness amongst older people: a systematic search and narrative review. *Aging Ment Health*. 2014;18(6):683–93.
369. Masi CM, Chen HY, Hawkley LC, Cacioppo JT. A meta-analysis of interventions to reduce loneliness. *Personal Soc Psychol Rev Off J Soc Personal Soc Psychol Inc*. 2011 Aug;15(3):219–66.
370. Quan NG, Lohman MC, Resciniti NV, Friedman DB. A systematic review of interventions for loneliness among older adults living in long-term care facilities. *Aging Ment Health*. 2020 Dec;24(12):1945–55.
371. Siette J, Cassidy M, Priebe S. Effectiveness of befriending interventions: a systematic review and meta-analysis. *BMJ Open*. 2017 Apr 26;7(4):e014304.

