

**Contribuciones a los
métodos aplicados en la
investigación en psicología y
ciencias de la comunicación**

Contribuciones a los métodos aplicados en la investigación en psicología y ciencias de la comunicación / Ennio Héctor Carro Pérez y Arturo Secundino Hernández Gómez, coordinadores.— Cd. Victoria, Tamaulipas : Universidad Autónoma de Tamaulipas ; Ciudad de México : Colofón , 2021
166 págs. ; 17 x 23 cm.

1. Psicología – Investigaciones 2. Pruebas psicológicas 3. Salud mental. I. Carro Pérez, Ennio Héctor, coordinador. II. Hernández Gómez, Arturo Secundino, coordinador.

LC: BF76

DEWEY: 150.724

Universidad Autónoma de Tamaulipas
D. R. © 2021
Matamoros SN, Zona Centro Ciudad
Victoria, Tamaulipas C.P. 87000
www.uat.edu.mx

Consejo de Publicaciones UAT
Centro Universitario Victoria
Centro de Gestión del Conocimiento. Tercer Piso
Cd. Victoria, Tamaulipas, México. C.P. 87149
consejopublicacionesuat@outlook.com
Tel. (52) 834 3181-800 • extensión: 2948 • www.uat.edu.mx

 **Fomento Editorial** Una edición del Departamento de Fomento Editorial de la Universidad Autónoma de Tamaulipas

Edificio Administrativo, planta baja, CU Victoria
Ciudad Victoria, Tamaulipas, México
Libro aprobado por el Consejo de Publicaciones UAT
ISBN UAT: 978-607-8750-48-1

Colofón S.A. de C.V.
Franz Hals núm. 130, Alfonso XIII
Delegación Álvaro Obregón C.P. 01460, Ciudad de México
www.colofonlibros.com • colofonedicionesacademicas@gmail.com
ISBN: 978-607-635-223-6

Se prohíbe la reproducción total o parcial de esta obra incluido el diseño tipográfico y de portada, sea cual fuera el medio, electrónico o mecánico, sin el consentimiento del Consejo de Publicaciones UAT.
Impreso en México • *Printed in Mexico* El tiraje consta de 350 ejemplares

Este libro fue dictaminado y aprobado por el Consejo de Publicaciones UAT mediante un especialista en la materia perteneciente al Sistema Nacional de Investigadores (SNI). Asimismo, fue recibido por el Comité Interno de Selección de Obras de Colofón Ediciones Académicas para su valoración en la sesión del segundo semestre 2020, donde se sometió al sistema de dictaminación a “doble ciego” con diagnóstico positivo.

Contribuciones a los métodos aplicados en la investigación en psicología y ciencias de la comunicación

Coordinadores

Ennio Héctor Carro Pérez
Arturo Secundino Hernández Gómez



UAT





Ing. José Andrés Suárez Fernández
PRESIDENTE

Dr. Julio Martínez Burnes
VICEPRESIDENTE

Dr. Héctor Manuel Cappello Y García
SECRETARIO TÉCNICO

C.P. Guillermo Mendoza Cavazos
VOCAL

Dra. Rosa Issel Acosta González
VOCAL

Ing. Rafael Pichardo Torres
VOCAL

Consejo Editorial del Consejo de Publicaciones de la Universidad Autónoma de Tamaulipas

Dra. Lourdes Arizpe Slogher • Universidad Nacional Autónoma de México | **Dr. Amalio Blanco** • Universidad Autónoma de Madrid, España | **Dra. Rosalba Casas Guerrero** • Universidad Nacional Autónoma de México | **Dr. Francisco Díaz Bretones** • Universidad de Granada, España | **Dr. Rolando Díaz Lowing** • Universidad Nacional Autónoma de México | **Dr. Manuel Fernández Ríos** • Universidad Autónoma de Madrid, España | **Dr. Manuel Fernández Navarro** • Universidad Autónoma Metropolitana, México | **Dra. Juana Juárez Romero** • Universidad Autónoma Metropolitana, México | **Dr. Manuel Marín Sánchez** • Universidad de Sevilla, España | **Dr. Cervando Martínez** • University of Texas at San Antonio, E.U.A. | **Dr. Darío Páez** • Universidad del País Vasco, España | **Dra. María Cristina Puga Espinosa** • Universidad Nacional Autónoma de México | **Dr. Luis Arturo Rivas Tovar** • Instituto Politécnico Nacional, México | **Dr. Aroldo Rodríguez** • University of California at Fresno, E.U.A. | **Dr. José Manuel Valenzuela Arce** • Colegio de la Frontera Norte, México | **Dra. Margarita Velázquez Gutiérrez** • Universidad Nacional Autónoma de México | **Dr. José Manuel Sabucedo Cameselle** • Universidad de Santiago de Compostela, España | **Dr. Alessandro Soares da Silva** • Universidad de São Paulo, Brasil | **Dr. Akexandre Dorna** • Universidad de CAEN, Francia | **Dr. Ismael Vidales Delgado** • Universidad Regiomontana, México | **Dr. José Francisco Zúñiga García** • Universidad de Granada, España | **Dr. Bernardo Jiménez** • Universidad de Guadalajara, México | **Dr. Juan Enrique Marcano Medina** • Universidad de Puerto Rico-Humacao | **Dra. Ursula Oswald** • Universidad Nacional Autónoma de México | **Arq. Carlos Mario Yori** • Universidad Nacional de Colombia | **Arq. Walter Debenedetti** • Universidad de Patrimonio, Colonia, Uruguay | **Dr. Andrés Piqueras** • Universitat Jaume I, Valencia, España | **Dr. Yolanda Troyano Rodríguez** • Universidad de Sevilla, España | **Dra. María Lucero Guzmán Jiménez** • Universidad Nacional Autónoma de México | **Dra. Patricia González Aldea** • Universidad Carlos III de Madrid, España | **Dr. Marcelo Urra** • Revista Latinoamericana de Psicología Social | **Dr. Rubén Ardila** • Universidad Nacional de Colombia | **Dr. Jorge Gissi** • Pontificia Universidad Católica de Chile | **Dr. Julio F. Villegas** • Universidad Diego Portales, Chile | **Ángel Bonifaz Ezeta** • Universidad Nacional Autónoma de México

Índice

| | |
|--|-----|
| Prólogo | 9 |
| Introducción | 11 |
| Capítulo 1. La técnica de redes semánticas naturales. Análisis del perfil sanitario de las creencias colectivas sobre salud y enfermedad <i>Arturo Sahagún Morales, Ennio Héctor Carro Pérez y Arturo Secundino Hernández Gómez</i> | 17 |
| Capítulo 2. El método de jueceo e índice de congruencia del ítem de Rovinelli y Hambleton, propuesto por Osterlind: aplicaciones en la validación del contenido de una escala de percepción de riesgo al conducir un automóvil <i>Ennio Héctor Carro Pérez y Arturo Secundino Hernández Gómez</i> | 53 |
| Capítulo 3. Propuesta metodológica para la investigación psicosocial de la calidad de vida en contextos comunitarios <i>Ennio Héctor Carro Pérez y Arturo Secundino Hernández Gómez</i> | 67 |
| Capítulo 4. La evaluación de la calidad científica en ciencias del comportamiento: una aproximación a las escalas y cuestionarios de calidad <i>Ennio Héctor Carro Pérez</i> | 91 |
| Capítulo 5. Revisiones sistemáticas: su definición, directrices y dos casos prácticos en el campo de las ciencias sociales <i>Oscar Eliezer Mendoza de los Santos y Jocelyn Pamela Castelán Félix</i> | 125 |
| Capítulo 6. Relaciones entre confiabilidad y tamaño de efecto en dos muestras de registros empíricos <i>Ennio Héctor Carro Pérez y Amada Ampudia Rueda</i> | 151 |
| Sobre los autores | 165 |

Prólogo

En esta contribución a los métodos aplicados en la investigación, usted se encontrará con diversos estudios realizados en el campo de la Psicología y la Ciencia de la Comunicación, en los cuales encontrará información sobre la elaboración y aplicación de algunas técnicas de redes semánticas, procedimientos de diseños metodológicos, cuestionarios de calidad, depuración de registros empíricos, así como algunas directrices para elaborar una revisión sistemática, además de un método de jueceo y congruencia de reactivos.

En los primeros capítulos encontrará algunos trabajos referentes a las creencias colectivas sobre salud y enfermedad, la percepción de riesgo al conducir un automóvil, así como un abordaje psicosocial de la calidad de vida en contextos comunitarios; después, una serie de investigaciones sobre la evaluación de la calidad científica en ciencias del comportamiento, algunos casos prácticos de la revisión sistemática en el campo de las ciencias sociales y finalmente, un análisis empírico de las relaciones entre confiabilidad y el tamaño de efecto.

Así mismo, en algunos estudios se presentan los instrumentos de medición que se realizaron, lo cual es un complemento de suma importancia para la realización de los trabajos de investigación, o al menos para el lector una guía para comprender mejor el estudio desarrollado por los investigadores.

El presente libro contribuye con algunos fundamentos teóricos, literatura sobre las temáticas estudiadas, diseños metodológicos e instrumentos aplicados a las investigaciones. La información y resultados de las diversas investigaciones, es para compartir las experiencias obtenidas a través de los diversos estudios científicos realizados por los autores y hacer más accesible el conocimiento científico desde la práctica de la investigación.

Esperamos que este libro proporcione información útil y sea una guía para cualquier trabajo de investigación.

Dr. Isaias Martínez Trejo

Introducción

En el ejercicio de la docencia a nivel universitario, a través de múltiples interacciones entre alumnos y compañeros profesores, hemos podido advertir -los autores- que la metodología de la ciencia o de la investigación científica se enseña en la mayoría de las ocasiones como una materia acabada, sin lugar a posibles contribuciones, donde todo está escrito o dicho, lo que ocasiona que el campo del conocimiento se observe estéril, sin posibilidad que en él germine una semilla que constituya una innovación, mejora o alternativa a los métodos ya existentes, y vaya que esto es una gran frustración para todos los que hemos enseñado en cursos avanzados de investigación científica en los grados superiores de licenciatura o en postgrado, cuando se advierte que muchos de los planteamientos de investigación de los alumnos, evitan a la metodología como objeto de indagación y sólo la emplean como una herramienta para dar soporte a sus trabajos.

Lo anterior, sí bien se afirma desde el limitado o estrecho margen que permite un puñado de observaciones asistemáticas acumuladas a lo largo de veinte años o más de experiencia docente en ciencias del comportamiento, permite tener al menos cierta materia empírica para derivar como hipótesis la afirmación de que en el estudio del comportamiento humano, la Metodología de la Investigación es percibida, enseñada y aprendida como un campo fijo, sin movimiento y por lo tanto no sujeto a investigación, lo cual es lamentable considerando la existencia de espacios dedicados por entero o en su mayoría a la publicación de investigaciones en y sobre metodología, como las revistas *Psychological Methods*, *Revista Latinoamericana de Metodología de las Ciencias Sociales*, *Journal of Mathematical Psychology*, *International Journal of Qualitative Methods*, entre otras, sin contar los numerosos trabajos que se publican en una de las áreas del método más populares: la medición. Tópico que tiene un espacio ganado en las ciencias del comportamiento, bajo el nombre de psicometría o sociometría.

De lo anterior y por la firme creencia de los autores de que el método científico es el sustento o estructura para la obtención del conocimiento válido y confiable que llamamos ciencia, hemos establecido un conjunto de iniciativas de investigación que tienen como objetivo contribuir en algunas áreas particulares del desarrollo de instrumentos de medición, de los protocolos de evaluación científica, del proceso general de la investigación y del análisis de datos. Empresas que han arrojado varios productos, de los cuales sólo algunos se muestran en este libro, aquellos que representan de mejor manera el trabajo realizado por el grupo de investigación en el campo de la metodología de las ciencias del comportamiento.

Así, la presente obra integra parte del trabajo de investigación realizado desde el 2015, por el cuerpo académico *Investigaciones Sociales en Contextos Diversos*, que hasta el año del 2016 se llamó *Investigaciones Jurídicas y Sociales*, específicamente en la línea de Alternativas Metodológicas y Tecnológicas en la Observación del Comportamiento Humano, de tal manera, comprende el esfuerzo de sus integrantes y alumnos formados bajo su seno como investigadores y divulgadores de la ciencia, en cada uno de los proyectos de investigación que han desarrollado.

Particularmente este libro se articula a través de la presentación de seis procesos metodológicos, que incluyen formas de recolección de datos a través de mecanismos asociativos como las Redes Semánticas Naturales; procedimientos útiles en la determinación de cualidades esenciales en los procesos de medida como es el cálculo o determinación de la confiabilidad y validez de los instrumentos, así se pone a prueba y se resaltan las bondades del procedimiento propuesto por Osterlind (1997) a través del Índice de congruencia del ítem de Rovinelli y Hambleton (1977), para determinar la validez de contenido de ítems integrantes de una herramienta de medida, método que constituye una alternativa a procesos de jueceo y determinación de coincidencias o acuerdo inter-jueces como los obtenidos a través de los coeficientes Kappa de Cohen o Alfa de Krippendorff (2007); en este mismo sentido, se presenta un trabajo donde se comparan algoritmos tradicionales para obtener la confiabilidad o estabilidad en dos conjuntos de puntajes, con la *d* de Cohen, empleada para establecer el tamaño de efecto, con ello se propone que los indicadores de tamaño de efecto pueden serlo también de confiabilidad, al estar asociada esta última con el efecto; de igual manera se presenta una opción del método en su conjunto a través de una propuesta metodológica mixta para observar la calidad de vida en entornos comunitarios; finalmente se ofrecen dos trabajos que ponen sobre la mesa un área de escaso desarrollo en la investigación en comportamiento humano en México, el análisis de la calidad científica y las síntesis de resultados a través de protocolos de calidad de la investigación y el desarrollo de revisiones sistemáticas y metaanálisis.

De tal manera, en el capítulo uno se analizan las creencias colectivas en materia de salud y enfermedad de habitantes con distinto perfil sanitario de los municipios de Tampico y Madero, Tamaulipas utilizando la Técnica de Redes Semánticas Naturales (TRSN), este estudio parte del proyecto *Rumores y creencias colectivas sobre la salud y la enfermedad en habitantes con distinto perfil*, clave PIFI2014-63, financiado en la convocatoria de investigación UAT-2014.

Para lograr el objetivo del trabajo se realizó una clasificación por perfil sanitario de los habitantes de Tampico y Madero, Tamaulipas. Posteriormente se determinó, mediante la TRSN las creencias colectivas sobre salud y enfermedad

más relevantes en los habitantes observados, finalmente se compararon dichas creencias por los perfiles encontrados. Además de los resultados obtenidos por perfil se discute la utilidad de las TRSN en la exploración de los significados y creencias individuales y colectivas en los individuos en el contexto de la salud.

En el capítulo dos, se muestran las cualidades y uso del Índice de congruencia del ítem de Rovinelli y Hambleton, como un procedimiento apropiado y alternativo para evaluar la validez de contenido de un ítem y de un instrumento o escala en su conjunto de acuerdo a lo expuesto por Osterlind (1997). Se ilustra cómo ejecutar el cálculo del índice y se describe un caso empírico donde se empleó para validar la Escala de Percepción de Riesgo al Conducir, mediante las evaluaciones hechas por una muestra de 50 jueces expertos en metodología e investigación científica. La técnica puede ser empleada para valorar cualquier tipo de ítem y no sólo aquellos confeccionados para medir un aspecto de una variable de manera escalar, se aconseja en cualquier tipo de estudio de carácter experimental o no experimental, al margen de cualquier diseño de investigación de corte cuantitativo o cualitativo. El trabajo formó parte del proyecto *Percepción de riesgo y conductas de riesgo asociadas al conducir un automóvil en zonas urbanas del sur de Tamaulipas y Ciudad de México*, financiado por la Dirección General de Educación Superior Universitaria, PRODEP México, Apoyo de Fomento a la Generación y Aplicación Innovadora del Conocimiento, oficio 511-6/17-8212, entre los años de 2017 a 2019.

En el capítulo tres, se describe una propuesta metodológica para el abordaje psicosocial de la calidad de vida en contextos comunitarios. Esta propuesta surge de la experiencia que han tendido los autores e integrantes del Cuerpo Académico en la investigación en contextos comunitarios, trata de reunir de manera coherente las discusiones y conclusiones que estos han tenido sobre el método en estos imaginarios estudiando el problema de la calidad de vida desde un enfoque psicosocial. Se describen las características más relevantes de las técnicas y procedimientos que integran la propuesta metodológica desde lo cualitativo y lo cuantitativo, por lo que puede calificarse como mixta. Así, se integran técnicas como la identificación y selección de informantes privilegiados con muestreos aleatorios; el desarrollo de un diario de campo con el uso de entrevistas y cuestionarios cualitativos y cuantitativos, sin olvidar el empleo más reciente de medios de georreferencia y la alusión al videoanálisis. La propuesta pretende ser un modelo metodológico plausible a seguir y generar una discusión necesaria en el plano del método para un enriquecimiento del campo en las ciencias del comportamiento.

En el capítulo cuatro, *La evaluación de la calidad científica en ciencias del comportamiento: una aproximación a las escalas y cuestionarios de calidad*, se hace una aportación fundamental para conocer, en los estudios primarios, el nivel de certidumbre

o sesgo de los hallazgos y sus efectos, sin ella no contaríamos con información suficiente para establecer si los esfuerzos destinados en una determinada línea de trabajo se realizan de manera adecuada y fructífera en tanto contribuciones al conocimiento de un fenómeno. De tal manera, en algunas disciplinas científicas, como las médicas, el desarrollo de procedimientos y herramientas en formato de cuestionarios o escalas, para evaluar la calidad de los estudios ha sido abundante, en otras ha sido escaso, esto último es notorio en aquellas que tienen que ver con el estudio del comportamiento humano, y específicamente en el contexto mexicano. El trabajo presenta una revisión preliminar de los protocolos y herramientas de evaluación que han sido desarrollados en las ciencias de la salud y de la psicología, la cual no hubiera sido posible sin el contacto académico entre el autor e integrantes del grupo de investigación *Innovaciones metodológicas en evaluación de programas* de la Universidad de Sevilla, España, particularmente con su líder el doctor Salvador Chacón Moscoso y la doctora Susana Sanduvete Cháves, con quienes se colabora actualmente en una iniciativa de investigación orientada a determinar la presencia de los estudios de revisión y metaanálisis en la producción científica de la psicología en México.

El capítulo cinco, retoma el tema del anterior pero desde el filón aplicado. Así, *Revisiones sistemáticas: su definición, directrices y dos casos prácticos en el campo de las ciencias sociales*, tiene por objetivo caracterizar las revisiones sistemáticas (RS) como un diseño metodológico idóneo en trabajos cuyo objetivo sea la descripción y/o síntesis del conocimiento producto de la investigación científica. Para ello, en primera instancia se ubica el papel de las RS en el plano de la investigación científica presentándose una clasificación de las mismas, se expone también una serie de directrices generales para realizar este tipo de estudios y evaluar su calidad, finalmente se dan a conocer dos casos de investigaciones en las cuales la RS fue el recurso metodológico para determinar la calidad científica de estudios primarios en comunicación mediática, y evaluar la producción de trabajos sobre regulación jurídica en psicología y ciencias de la comunicación. Ambos trabajos empíricos formaron parte del proyecto *Análisis de la investigación científica sobre comunicación mediática y su regulación jurídica en México*, subvencionada por la Subdirección de Educación Superior, de la Dirección General de Educación Superior Universitaria, de la Secretaría de Educación Pública.

Por último, el capítulo seis, *Relaciones entre confiabilidad y tamaño de efecto en dos muestras de registros empíricos*, surge de la colaboración entre integrantes del cuerpo académico y profesores adscritos a la Facultad de Psicología de la Universidad Nacional Autónoma de México (UNAM), gracias al apoyo de la Dirección General de Cooperación e Internacionalización (DGEI) de la UNAM, de la Dirección de Internacionalización y Colaboración Académica, y de la Dirección de Programas

de Apoyo de la Universidad Autónoma de Tamaulipas (UAT), otorgado al proyecto *Determinación de indicadores de conducción vehicular y factores de riesgo asociados a los accidentes de tráfico en zonas urbanas del sur de Tamaulipas*. En esta contribución se puntualiza la necesidad del desarrollo de mecanismos alternativos para obtener indicadores de confiabilidad o estabilidad interna de las medidas, en esta trayectoria se establece que un medio para establecer indicadores de esta cualidad es el tamaño de efecto, de tal forma se analizan los resultados de indicadores tradicionales de confiabilidad como el Alfa de Cronbach y la R de Pearson comparándolas con la d de Cohen, empleando para ello dos muestras de registros simultáneos de tiempo y velocidad. Los resultados, en todas las muestras observadas, evidencian la ocurrencia de valores de confiabilidad que aumentan cuando el efecto es bajo, lo que podría señalar que los indicadores de efecto como la d de Cohen pueden ser útiles en la comprobación de los coeficientes tradicionales de confiabilidad o bien constituirse en alternativa de los mismos, sin embargo se debe seguir demostrando.

De lo hasta aquí comentado, se espera que el presente libro sea una contribución poderosa en el campo de la metodología de la ciencia, y que cumpla con uno de sus cometidos, mostrar el terreno del método como un área en movimiento, inacabada, que es factible como objeto de estudio y no sólo como un conjunto de herramientas finitas.

Referencias

- Hayes, A. F. y Krippendorff, K. (2007). "Answering the Call for a Standard Reliability Measure for Coding Data". *Communication Methods and Measures*. Vol. 1 (1), pp. 77-89. Disponible en línea en <https://doi.org/10.1080/19312450709336664>
- Osterlind, S. J. (1997). *Constructing Test Items: Multiple-Choice, Constructed-Response, Performance and Other Formats*. Boston, United States: Kluwer Academic Publishers.
- Rovinelli, R. J. & Hambleton, R. K. (1977). "On the use of content specialists in the assessment of criterion-referenced test item validity". *Dutch Journal of Educational Research*, 2, pp. 49-60.

Capítulo 1

La técnica de redes semánticas naturales. Análisis del perfil sanitario de las creencias colectivas sobre salud y enfermedad¹

*Sahagún-Morales, Arturo
Carro-Pérez, Ennio Héctor
Hernández-Gómez, Arturo Secundino*

Resumen

El objetivo del estudio fue analizar creencias colectivas de salud y enfermedad en personas con distinto perfil sanitario utilizando la Técnica de Redes Semánticas Naturales. Los participantes fueron mayores de 18 años, residentes de los municipios de Tampico y Madero, Tamaulipas. La muestra fue n=62. Éste es un estudio de campo no experimental de orientación mixta; es de alcance descriptivo y correlacional. Los resultados muestran que las proposiciones asociadas a los estímulos Salud y Enfermedad en un perfil no son las mismas en el otro. De esta aproximación al problema de las creencias colectivas de la salud y enfermedad se han obtenido indicadores para posteriores intervenciones preventivas.

Palabras clave: Perfil sanitario, Creencias colectivas, Redes semánticas naturales, Salud, Enfermedad.

Introducción

En México las líneas de acción en salud han sido desarrolladas a partir de un modelo tradicional biomédico, prescindiendo casi en su totalidad de intervenciones que trabajen sobre factores psicosociales y limitándose a considerar estos últimos como simples eventos coligados a dificultades en la aplicación, seguimiento, difusión y/o aceptación de intervenciones médicas enfocadas en prevenir, diagnosticar o tratar algún padecimiento (Urbina, Mogurl, Muñiz, y Solís, 2006). Si bien es cierto que el modelo biomédico ha sido más que efectivo para el tratamiento, inmunización,

¹ Los datos que se presentan en este trabajo son extraídos de la tesis de licenciatura elaborada en julio de 2016 y defendida el 8 de septiembre de 2016 por Arturo Sahagún-Morales bajo el título de “Rumor y creencias colectivas sobre salud, enfermedad y conductas de riesgo en habitantes de Tampico y Madero”. Fue parte del proyecto *Rumores y creencias colectivas sobre la salud y la enfermedad en habitantes con distinto perfil sanitario* clave PFI2014-63 del cuerpo académico Investigaciones Jurídicas y Sociales, proyecto subvencionado dentro de la convocatoria de apoyo a la investigación UAT 2014, de la Universidad Autónoma de Tamaulipas.

eliminación y hasta la erradicación² de algunas de las más peligrosas enfermedades del siglo XX como la viruela, la poliomielitis, el paludismo, la tuberculosis y la lepra (Urbina, Mogurl, Muñiz, y Solís, 2006), aún existe un amplio campo de trabajo en el terreno de la salud a través del estudio de los factores psicosociales pues, hasta ahora, no ha sido suficiente el hecho de desarrollar nuevos medicamentos y campañas de inmunización, para poder prevenir y controlar (no se diga erradicar) algunos de los más grandes riesgos para la salud como el VIH/SIDA, la obesidad y los accidentes viales.

Para tener una perspectiva general de la salud en México se pueden observar datos de la Organización para la Cooperación y el Desarrollo Económicos (OCDE) que ubican a México como el país con la esperanza de vida más baja (OCDE, 2014). Sólo entre el 2000 y el 2012, en los países de la OCDE, se presentó un incremento en la esperanza de vida de 3 años en promedio (pasando de 77.1 a 80.2 años), mientras que la esperanza de vida en México únicamente creció un año en promedio durante el mismo periodo (pasando de 73.3 a 74.4 años) incrementándose la brecha, en términos de longevidad, entre los países miembros de la OCDE y México (OCDE, 2014). Este incremento se debe a una amplia gama de *comportamientos dañinos* donde destacan los malos hábitos de nutrición y las altas tasas de obesidad. Además, la esperanza de vida se afecta por el incremento en la mortalidad de la población mexicana debido a la diabetes, enfermedades cardiovasculares, accidentes de tráfico y homicidios; sin olvidar que existen otros factores agravantes como la dificultad para acceder a servicios médicos de alta calidad (OCDE, 2014).

Todos estos factores de riesgo mencionados que impiden el crecimiento en la esperanza de vida y que aumentan el índice de mortalidad de la población mexicana se enmarcan dentro de las denominadas Enfermedades No Transmisibles (ENT) y la OMS las clasifica en 4 grandes grupos: 1) enfermedades cardiovasculares, 2) cáncer, 3) enfermedades respiratorias crónicas y 4) diabetes (OMS, 2015). Según la OMS, las ENT (hasta el 2015) causan 38 millones de personas por año. Además, del 100 % de muertes por ENT, 28 millones (75 %) se producen en países de ingresos bajos y medios (OMS, 2015).

² En epidemiología las palabras eliminación y erradicación se utilizan para diferentes situaciones. La eliminación de una enfermedad conlleva la ausencia del microbio o de los casos clínicos de algún padecimiento determinado en un lugar y tiempo específicos sin garantía de que no surgirán nuevos brotes a futuro, en cambio, la erradicación implica la desaparición definitiva del virus lo cual garantiza la ausencia de casos a nivel global en el futuro.

Lo anterior no sólo representa una alerta significativa para países como México, sino que, a nivel mundial, implica la necesidad de trabajar sobre factores de riesgo. Tal ha sido el incremento de muertes por ENT que ha propiciado la creación de diversos mecanismos preventivos a nivel internacional como el *Plan de acción mundial para la prevención y el control de las ENT 2013-2020* (OMS, 2013), apoyado por más de 190 países que buscan reducir el número de muertes por ENT en 25 % proyectando dicha disminución para el 2025. Aunado a lo anterior, existen más factores que ponen en peligro la salud directamente atribuibles a comportamientos de riesgo y que requieren de atención tales como la contaminación ambiental, el cambio climático y el estrés, estos se asocian con los niveles de morbilidad y mortalidad por cáncer, enfermedades cardiovasculares y neumopatías crónicas. De igual manera, la exposición a productos como los gases producidos por la combustión del diésel incrementa el riesgo de contraer algún tipo de cáncer, y el estrés crónico que se vincula con el trabajo o el desempleo relacionado con el padecimiento de enfermedades cardiovasculares (OMS, 2013).

Todo lo anterior permite observar que los problemas de salud no se remiten sólo al contagio de agentes infecciosos, ya sea por vía de vectores vivos (mosquitos, cucarachas, garrapatas, moscas, etc.), por vía aérea o por ingesta de alimentos infectados o por otra forma de exposición. La salud, en una concepción integral, se encuentra estrechamente relacionada con los comportamientos de las personas, sus actitudes y sus creencias. Luego entonces, el objetivo de este estudio es analizar las creencias colectivas en materia de salud y enfermedad de habitantes con distinto perfil sanitario de los municipios de Tampico y Madero, Tamaulipas utilizando la Técnica de Redes Semánticas Naturales (TRSN). Para lograr este objetivo se realiza una clasificación por perfil sanitario de los habitantes de Tampico y Madero, Tamaulipas. Posteriormente se determina, mediante la TRSN, cuáles son las creencias colectivas sobre salud y enfermedad en los habitantes con distinto perfil sanitario, para finalmente comparar por perfil dichas creencias colectivas.

Ahora bien, es necesario exponer algunos puntos importantes sobre el enfoque teórico de este trabajo así como de la TRSN.

Se trabajó desde el marco de la Teoría de las Representaciones Sociales (Moscovici, 1979). Aquí la definición de representación social tentativa, que para efectos de este trabajo se considera más apropiada, es la que ofrecen Gaskell y Fraser, (1990 citado en León y Páez, 2002) donde la refieren como “las funciones sociales de las actitudes y creencias socialmente compartidas sobre un objeto social” (p. 217). Analicemos brevemente los conceptos que componen esta definición y procedamos, a partir de dicho análisis, a ofrecer una definición precisa de representaciones sociales.

El primer término que observamos en la definición de anterior es el de *actitud* mismo que se puede definir como “la tendencia o predisposición relativamente duradera para evaluar de un determinado modo a una persona, suceso o situación a partir de los significados que se les da y a actuar en consonancia con esta evaluación” (Consuegra, 2010, p. 3). Una actitud socialmente compartida es, por tanto, una predisposición relativamente duradera que adoptan los individuos que integran un determinado grupo social para evaluar de un determinado modo a personas, sucesos o situaciones a partir de los significados que se comparten en el grupo y a actuar en concordancia con esta evaluación. Las creencias, en este trabajo, se entenderán como conjunto de proposiciones, referentes a un objeto, que no implican una comprobación objetiva previa y que son aceptadas como verdaderas; y las creencias socialmente compartidas o creencias colectivas son, entonces, el conjunto de proposiciones referentes a un objeto social, y aceptadas como verdaderas sin comprobación objetiva previa que se comparten dentro de un determinado grupo social. El concepto de objeto social en este trabajo se entenderá como cualquier objeto concreto o conceptual que resulte de interés para un grupo social determinado. Finalmente, la definición de representación social que ofrecen Gaskell y Fraser hace referencia al término “funciones sociales” entendido como la finalidad que tienen, para los individuos dentro de un grupo social, determinadas actitudes y creencias socialmente compartidas sobre un objeto.

Realizadas las aclaraciones previas puede decirse que estudiar una representación social de la salud y la enfermedad es estudiar la finalidad que tienen, para un determinado grupo social, las actitudes (en tanto predisposición para evaluar y actuar) y las creencias colectivas (en tanto objetos conceptuales específicos) respecto a la salud y la enfermedad (en tanto objetos concretos de interés social).

Respecto a la TRSN se destaca que ésta es una importante herramienta de medición de significados desarrollada en la psicología cognitiva (Figuroa, González y Solís, 1976) y que se encuentra en creciente uso (García y Jiménez, 1996; Valle, 2003; Acosta, 2010; Sánchez, de la Garza y López, 2013; Bermúdez et al. 2014). Existen antecedentes de la medición de representaciones sociales utilizando la TRSN (Aguilera, 2010; Denegri, et al., 2010; Montano-Alarcón, 2015; Pérez, 2015; Urrego, 2016; Castañeda, 2016) al igual que estudios que han buscado su perfeccionamiento (Reyes, 1993). En este estudio se atienden las sugerencias de (Hinojosa, 2008) referentes al tratamiento estadístico de los datos obtenidos mediante redes semánticas naturales.

Luego de lo anterior podemos establecer la pregunta de investigación ¿existen diferencias entre las creencias colectivas sobre salud y enfermedad en habitantes con distinto perfil sanitario de los municipios de Tampico y Madero, Tamaulipas? Ante

esta pregunta se plantea la hipótesis de trabajo (H1) de que *las creencias colectivas sobre salud y enfermedad en habitantes con perfil sanitario saludable difieren de las creencias colectivas de los habitantes con perfil sanitario no saludable en habitantes de los municipios de Tampico y Madero, Tamaulipas.*

Método

Participantes

Los participantes fueron personas mayores de 18 años residentes de los municipios de Tampico y Madero, Tamaulipas. Se requirió que los participantes habitaran en alguna de las colonias seleccionadas para el estudio, lo anterior responde al hecho de que se buscó trabajar con personas que vivan en colonias consideradas en riesgo, esto es, colonias con bajos recursos y alta densidad poblacional.

Muestra

Se determinó trabajar con una muestra no probabilística propositiva (Kerlinger y Lee, 2002) debido a que se eligieron áreas geográficas y unidades de análisis que se presume, por conocimientos previos de la población a estudiar, son típicas del universo de estudio, en este caso integrado por personas que viven en situaciones de riesgo para su salud. La muestra planeada originalmente fue de 120 participantes ($n = 120$) sin embargo, el total de encuestas posibles fue de 116 ($n = 116$) debido a dificultades de índole práctica durante el levantamiento de datos. Tras realizar la depuración de los datos perdidos en la matriz con los datos de los 116 participantes se obtuvo una muestra final de 62 participantes ($n = 62$) sobre la cual se realizaron los análisis de datos. En dicha muestra de 62 participantes se encontró una distribución por sexo de 38.70 % hombres y 61.30 % mujeres, mientras que por municipio fue de 50 % participantes de Tampico y 50 % de Madero.

Instrumentos

Se elaboró un cuestionario enfocado en conocer el perfil sanitario de los participantes así como obtener información referente a las creencias de los mismos sobre su propia salud y a los conceptos de salud y de enfermedad. Para su elaboración se consultó una batería de 4 instrumentos mostrados en la Tabla 1.

El cuestionario final constó de 3 apartados I) Datos Generales, II) Perfil Sanitario y Percepción de la Salud y III) Creencias Salud-Enfermedad (ver Apéndice A).

Tabla 1. Instrumentos utilizados para la elaboración del cuestionario para determinar el perfil sanitario y las creencias sobre salud, enfermedad y conductas de riesgo

| Instrumento | Autor(es) |
|---|--|
| Encuesta Nacional de Salud y Nutrición 2015. Cuestionario de adultos de 20 años o más | Secretaría de Salud (SSA) |
| Cuestionario de Salud SF-36 | Ware JE, Jr, Sherbourne CD. The MOS 36 item short-form health survey (SF-36) (I). Conceptual framework and item selection. <i>Med Care.</i> 1992;30:473-83 |
| Cuestionario de factores de riesgo | Fundación BEST y Comisión Federal para la Protección contra Riesgos Sanitarios (Cofepris) |
| Cuestionario de Fagerstrom | Versión del Centro de Integración Juvenil, A.C. Tampico, Tamaulipas, México |

Fuente: Elaboración propia.

Diseño de investigación

El enfoque del estudio es mixto debido a que es una investigación de campo no experimental de un grupo con orientación cuantitativa y al mismo tiempo es cualitativo, porque busca conocer la perspectiva de participantes que viven una determinada situación, con un diseño etnográfico transversal (Joyceen, 1994 citado en Álvarez-Gayou, 2003). También, es de alcance descriptivo y correlacional.

Procedimiento

Para determinar la población y muestra del estudio primeramente se examinó el Sistema para la Consulta de Información Censal 2010 Versión 05/2012 (INEGI, 2012) perteneciente al Instituto Nacional de Estadística y Geografía (INEGI). Se trabajó con las Áreas Geo-Estadísticas Básicas (AGEB) urbanas con mayor densidad poblacional en los municipios de Tampico y Madero, Tamaulipas. Las AGEB son delimitaciones territoriales desarrolladas por el INEGI para clasificar y referenciar correctamente información censal a través del Marco Geo-Estadístico Nacional. Éstas pueden ser de tres niveles diferentes: 1) Estatal (AGEE), Municipal (AGEM) y Básica (AGEB); y, a su vez, pueden ser urbanas o rurales. Se utilizaron AGEB urbanas porque delimitan la mayor parte de una cabecera municipal en conjuntos que abarcan de 25 a 50 manzanas, por lo que representan la mejor aproximación posible a la población real de los municipios y facilitan la identificación de las colonias que los componen (INEGI, s.f.).

Se identificaron 32 AGEB urbanas con alta densidad poblacional (entre 3461 y 5319 habitantes),³ 14 de ellas se ubican en el municipio de Tampico y 18 en el municipio de Madero. Posteriormente se identificaron colonias cuya extensión territorial se encuentra total o parcialmente dentro de los límites de las 32 AGEB seleccionadas; para ello se usó Google Maps.⁴ Después se seleccionaron 12 colonias en riesgo las cuales se muestran en la Tabla 2.

Tabla 2. Colonias en riesgo de los municipios de Tampico y Madero, Tamaulipas

| Colonias del municipio de Tampico | Colonias del municipio de Madero |
|-----------------------------------|----------------------------------|
| 1. Enrique Cárdenas González | 1. Revolución verde |
| 2. Laguna de la Puerta | 2. Emiliano Zapata |
| 3. Tanco | 3. Luna Luna |
| 4. Vicente Guerrero | 4. Candelario Garza |
| 5. Cascajal | 5. Francisco Villa |
| | 6. Obrera |
| | 7. Árbol Grande |

Fuente: Elaboración propia.

Una vez seleccionadas las colonias se determinó una muestra no probabilística propositiva (Kerlinger y Lee, 2002) de 120 participantes (n=120) con una distribución de 50 % hombres y 50 % mujeres. El muestreo no probabilístico propositivo facilita la realización del estudio al elegir de forma intencional áreas y participantes que se presume son representativos de la población específica (Kerlinger y Lee, 2002).

El levantamiento de datos se realizó contemplando 10 aplicaciones del cuestionario por colonia (5 hombres y 5 mujeres). Las ubicaciones específicas de cada aplicación se determinaron con anterioridad utilizando, nuevamente, la información del Sistema para la Consulta de Información Censal 2010 Versión 05/2012 del INEGI y de Google Maps. El criterio de ubicación fue el siguiente: se distribuyeron las 10 aplicaciones por colonia de tal manera que se cubriera

³ El rango de 3461 a 5319 habitantes está predefinido en el Sistema para la Consulta de Información Censal 2010 Versión 05/2012 del INEGI como el de mayor densidad poblacional.

⁴ La información obtenida del Sistema para la Consulta de Información Censal 2010 Versión 05/2012 del INEGI discrepaba en algunos casos con la información obtenida de Google Maps, dicha discrepancia se debe a la diferencia en la actualización de los datos pues, mientras la información del INEGI en el momento de obtener la información del estudio era del 2012, la información de Google Maps era del 2009.

de la forma *más homogénea posible* la extensión territorial de la colonia, esto es, una aplicación por calle en calles específicas para abarcar toda la colonia. Una vez determinada la ubicación de cada una de las 120 aplicaciones se determinó que, al llegar a la calle de una aplicación en concreto, se elegiría de forma aleatoria la casa en donde se aplicaría el cuestionario.

El levantamiento de datos se efectuó durante 3 días por 5 aplicadores previamente capacitados, lográndose un total de 116 aplicaciones. El cuestionario se aplicó por parte del entrevistador y no de forma auto-aplicada. Los datos fueron analizados en el software SPSS Statistics v.20.

Resultados

Clasificación de los participantes según el Perfil Sanitario

Para clasificar a los participantes del estudio en función del Perfil Sanitario y poder realizar comparaciones entre los distintos perfiles, se realizó un análisis bietápico de conglomerados, debido a que permite la exploración de agrupaciones naturales de datos categóricos y continuos (Pérez, 2004). Para la realización del mismo se usaron variables nominales del apartado II del cuestionario:

- Ítem 9 ¿Algún médico le ha dicho que tiene diabetes o azúcar en la sangre?
- Ítem 10 ¿Algún médico le ha dicho que tiene la presión alta?
- Ítem 11 ¿Le ha dicho el médico que usted tiene alguna afección cardiaca?
- Ítem 12 ¿Alguna vez un médico le ha dicho que padece alguna enfermedad de riñón?
- Ítem 13 ¿Algún médico le ha dicho que tiene colesterol alto?
- Ítem 14 ¿Algún médico le ha dicho que tiene triglicéridos altos?
- Ítem 15 ¿Le ha dicho su médico si ha tenido una embolia cerebral?
- Ítem 16 ¿Alguna vez le ha dicho un médico o dentista que tiene o tuvo cáncer o tumor maligno?

Tabla 3. Resultados de la Razón de verosimilitud para probar la independencia de las variables, del ítem 9 al 16

| Ítem | | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|------|----------------|-------|--------|-------|-------|--------|-------|-------|
| 9 | X ² | 0.786 | 1.099 | 1.830 | 2.782 | 1.962 | 0.242 | 3.101 |
| | P | 0.375 | 0.295 | 0.176 | 0.095 | 0.161 | 0.623 | 0.078 |
| 10 | X ² | --- | 4.132 | 0.020 | 3.418 | 5.165 | 3.736 | 1.841 |
| | P | --- | 0.042* | 0.886 | 0.065 | 0.023* | 0.053 | 0.175 |
| 11 | X ² | | --- | 0.738 | 0.446 | 0.261 | 0.100 | 0.518 |
| | P | | --- | 0.390 | 0.504 | 0.610 | 0.752 | 0.472 |

| Ítem | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|------|----------------|----|-----|-------|--------|-------|-------|
| 12 | X ² | | --- | 2.782 | 5.095 | 0.242 | 0.351 |
| | P | | --- | 0.095 | 0.024* | 0.623 | 0.554 |
| 13 | X ² | | | --- | 12.312 | 0.394 | 0.018 |
| | P | | | --- | 0*** | 0.530 | 0.892 |
| 14 | X ² | | | | --- | 0.475 | 0.003 |
| | P | | | | --- | 0.491 | 0.956 |
| 15 | X ² | | | | | --- | 0.170 |
| | P | | | | | --- | 0.680 |

*p < 0.05 ***p < 0.001

Fuente: Elaboración propia.

El criterio de elección de los ítems para la clasificación del Perfil Sanitario fue el hecho de que dichos ítems obtienen información sobre una indicación médica. Para la exploración del supuesto de independencia de las variables, necesario en el análisis, se utilizó la Razón de verosimilitud, obteniéndose los resultados de la Tabla 3).

Se obtuvo un resultado de dos conglomerados con una buena medida de silueta de cohesión y separación (0.6). El conglomerado 1 se denominó Perfil No Saludable, conforma el grupo más pequeño ($n_1=29$). Es el perfil caracterizado por la mayor cantidad de respuestas afirmativas a los ítems clasificatorios, esto implica que, al menos una vez, los participantes que conforman el conglomerado han recibido la indicación de un médico de que su salud ha estado deteriorada de alguna manera en los rubros especificados según el ítem (ver Tabla 4).

Tabla 4. % de respuestas afirmativas y negativas a ítems clasificatorios entre los conglomerados 1 (Perfil No Saludable) y 2 (Perfil Saludable)

| Ítem clasificatorio en orden de importancia para el modelo | Perfil No Saludable ^a | | Perfil Saludable ^b | |
|--|----------------------------------|---------------------------|-------------------------------|---------------------------|
| | % de respuestas afirmativas | % de respuestas negativas | % de respuestas afirmativas | % de respuestas negativas |
| Ítem 14 ¿Algún médico le ha dicho que tiene triglicéridos altos? | 44.8 % (13) | 55.2 % (16) | 0 % (0) | 100 % (33) |
| Ítem 13 ¿Algún médico le ha dicho que tiene colesterol alto? | 37.9 % (11) | 62.1 % (18) | 0 % (0) | 100 % (33) |

| Ítem clasificatorio en orden de importancia para el modelo | Perfil No Saludable ^a | | Perfil Saludable ^b | |
|---|----------------------------------|---------------------------|-------------------------------|---------------------------|
| | % de respuestas afirmativas | % de respuestas negativas | % de respuestas afirmativas | % de respuestas negativas |
| Ítem 10 ¿Algún médico le ha dicho que tiene la presión alta? | 34.5 % (10) | 65.5 % (19) | 0 % (0) | 100 % (33) |
| Ítem 12 ¿Alguna vez un médico le ha dicho que padece alguna enfermedad de riñón? | 24.1 % (7) | 75.9 % (22) | 0 % (0) | 100 % (33) |
| Ítem 9 ¿Algún médico le ha dicho que tiene diabetes o azúcar en la sangre? | 24.1 % (7) | 75.9 % (22) | 0 % (0) | 100 % (33) |
| Ítem 16 ¿Alguna vez le ha dicho un médico o dentista que tiene o tuvo cáncer o tumor maligno? | 17.2 % (5) | 82.8 % (24) | 0 % (0) | 100 % (33) |
| Ítem 11 ¿Le ha dicho el médico que usted tiene alguna afección cardiaca? | 10.3 % (3) | 89.7 % (26) | 0 % (0) | 100 % (33) |
| Ítem 15 ¿Le ha dicho su médico si ha tenido una embolia cerebral? | 3.4 % (1) | 96.6 % (28) | 0 % (0) | 100 % (33) |

a) n₁ = 29 b) n₂ = 33

Fuente: Elaboración propia.

El conglomerado 2 se denominó Perfil Saludable, conforma el grupo más grande (n₂=33). Es el perfil en el que el 100 % de los participantes respondió negativamente a todos los ítems clasificatorios, lo que significa que los participantes en este conglomerado no han sido diagnosticados oficialmente por un médico en alguna de las áreas especificadas en los ítems (ver Tabla 4).

Para ver una descripción detallada de las diferencias y similitudes de los 2 conglomerados obtenidos se puede consultar el Apéndice B.

Interpretación de la Técnica de Redes Semánticas Naturales

Para el análisis del apartado III del cuestionario se recurrió a las recomendaciones de Hinojosa (2008) para el tratamiento cuantitativo de la TRSN. Los pasos del análisis realizado en este trabajo se describen a continuación:

1. Se analizó un total de 4 redes referentes a las palabras estímulo *Salud y Enfermedad*.

Las redes resultantes se clasifican de la siguiente manera:

- a. CSPS: Creencias Salud de Perfil Saludable (n = 33)
- b. CEPS: Creencias Enfermedad de Perfil Saludable (n = 33)
- c. CSPNS: Creencias Salud de Perfil No Saludable (n = 29)
- d. CEPNS: Creencias Enfermedad de Perfil No Saludable (n = 29)

2. Se pidió a los participantes que mencionaran 5 conceptos que vinieran a su mente cuando escuchaban la palabra estímulo correspondiente a cada red. Una vez mencionadas las palabras se pidió a los participantes que las ordenaran en función de la importancia que para ellos guarda cada palabra, asignando el número 1 a la palabra más importante, 2 a la que sigue en importancia, 3 a la siguiente y continuando sucesivamente hasta jerarquizar las 5 palabras.
3. Los datos de las redes fueron analizados en una hoja de cálculo de Microsoft Office Excel (2013).
4. Se determinó el tamaño de la red (o Valor J),⁵ es decir, el número de palabras definidoras mencionadas por todos los participantes de la red.
5. En cada red semántica se elaboró una relación de palabras definidoras⁶ y su frecuencia. Para realizar análisis estadísticos a partir de las frecuencias se consideraron sólo palabras definidoras con una frecuencia mayor a uno.
6. Se calculó la mediana de las posiciones jerárquicas que los participantes asignaron a cada palabra definidora. Esto significa que si una palabra fue mencionada 5 veces en la primera posición jerárquica (la más alta, o sea, la posición número 1), 3 veces en la posición jerárquica 2 y 3 veces en la posición jerárquica 3, la mediana obtenida, para esa palabra, es de 2 debido a que en el ordenamiento se obtiene 1 1 1 1 2 2 2 3 3 3. El ordenamiento final de las palabras atendió los criterios de frecuencia descendente y mediana ascendente.
7. Los porcentajes se calcularon obteniendo el cociente entre la frecuencia de la palabra definidora y el número de participantes de la red, multiplicando el resultado por cien. Este es un indicador de la importancia de la palabra para el conjunto de participantes de la red.
8. Se realizaron 2 tipos de comparaciones de grupos para determinar el grado de consenso intra-grupo de cada red. La primera comparación se realizó estableciendo como criterio para elegir conjuntos SAM (Semantic Association

⁵ Tradicionalmente, en el uso de redes semánticas naturales, se calcula el valor J sumando todas las palabras que son mencionadas por los participantes, de tal forma que si, por ejemplo, diez participantes mencionaron un total de 110 palabras entonces el valor de J es 110 ($J = 110$). Algunos autores consideran este valor como “la riqueza o densidad semántica de la red”, no obstante, partiendo de las observaciones realizadas por Hinojosa (2008), en este trabajo consideraremos al valor J como un indicador del tamaño de la red.

⁶ Las palabras definidoras se agruparon según dos criterios: 1) la repetición exacta de una palabra, por ejemplo, el sujeto 1 menciona la palabra *Cáncer* y el sujeto 2 menciona la palabra *Cáncer* obteniéndose una frecuencia de 2; y 2) la presencia de palabras sinónimas, por ejemplo, los participantes 1 y 2 dicen la palabra *Cáncer* y el sujeto 3 dice *Cuando se tiene cáncer*, en este caso la frecuencia de la palabra *Cáncer* es de 3.

Memory⁷) porcentajes $\geq 10\%$ y porcentajes $\geq 15\%$. Los grupos con mayor cantidad de palabras que satisfacen los criterios son aquellos con mayor consenso intra-grupo. Asimismo, la segunda comparación para determinar el grado de consenso intra-grupo consistió en calcular los índices C10 y C15 (Consenso Grupal 10 y Consenso Grupal 15 respectivamente) los cual representan el porcentaje promedio de las primeras 10 palabras (C10) con mayor frecuencia y de las primeras 15 palabras (C15) con mayor frecuencia; en este indicador el porcentaje mayor representa mayor consenso intra-grupo.

9. Para determinar si las diferencias en cuanto al grado de consenso en los grupos son estadísticamente significativas se realizó la prueba U de Mann Whitney.
10. Para saber si existe correlación entre las palabras definidoras con frecuencia ≥ 2 se realizó el análisis Rho de Spearman.⁸
11. Finalmente se realizó una descripción no cuantitativa de las redes semánticas naturales.

A continuación se muestran los resultados de cada red semántica natural comenzando por CE_{PS} y CE_{PNS} y su comparación para, posteriormente, mostrar lo propio con CS_{PS} y CS_{PNS} .

La red CE_{PS} corresponde a $n = 33$ y está conformada por un total de 138 palabras definidoras ($J = 138$). Por otro lado, la red CE_{PNS} corresponde a $n = 29$ y está conformada por un total de 117 palabras definidoras ($J = 117$). Las Tablas 5 y 6 muestran un resumen de las redes semánticas naturales, la frecuencia (F), mediana (M), porcentaje (%) e indicadores de consenso intra-grupo de las redes CEPS y CEPNS respectivamente.

⁷ Los conjuntos Semantic Association Memory son indicadores de fortaleza en las asociaciones y de la estructura de significados dadas en un determinado conjunto de individuos debido a que la memoria semántica es el punto de partida sobre el que se fundamenta el uso de redes semánticas. Luego, las palabras que integran los conjuntos SAM muestran el núcleo central de la red de significados que giran en torno a un concepto concreto, en este caso: Salud y Enfermedad. Para más información sobre el fundamento teórico de las redes semánticas se recomienda consultar los trabajos de Hernández (1998) y Vera-Noriega, Batista de Albuquerque y Pimentel (2005).

⁸ Para poder realizar análisis de correlación es necesario igualar los grupos. En la correlación de las palabras definidoras con frecuencia ≥ 2 fue necesario igualar los grupos debido a que no todas las palabras de una red aparecen en otra. Por ejemplo, la palabra *Sobrepeso* es una palabra definidora con frecuencia ≥ 2 en la red CE_{PNS} pero sólo alcanza una frecuencia de 1 en la red CE_{PS} . En este caso se asigna la frecuencia de 1 para la palabra *Sobrepeso* de la red CE_{PS} . Asimismo, se asignó una frecuencia de 0 para las palabras que no se presentaron en un grupo pero que en el otro tuvieron frecuencia ≥ 2 .

Tabla 5. Red Semántica Natural CE_{ps} y sus indicadores de consenso intra-grupo

| # | Palabra definidora | F | M | % | Conjunto SAM % ≥10 % | Conjunto SAM % ≥15 % |
|----|---------------------------|----|-----|----------------------|-------------------------|-------------------------|
| 1 | Mal humor | 6 | 3.5 | 18.18 | SÍ | SÍ |
| 2 | Cansancio | 5 | 1 | 15.15 | SÍ | SÍ |
| 3 | Cáncer | 5 | 1 | 15.15 | SÍ | SÍ |
| 4 | Mala alimentación | 5 | 2 | 15.15 | SÍ | SÍ |
| 5 | Alcohol y cigarros | 5 | 2 | 15.15 | SÍ | SÍ |
| 6 | Depresión | 5 | 2 | 15.15 | SÍ | SÍ |
| 7 | Malestar | 5 | 3 | 15.15 | SÍ | SÍ |
| 8 | Tristeza | 5 | 3 | 15.15 | SÍ | SÍ |
| 9 | Diabetes | 4 | 1.5 | 12.12 | SÍ | - |
| 10 | Presión alta y baja | 4 | 2.5 | 12.12 | SÍ | - |
| 11 | Decaído | 3 | 2 | 9.09 | - | - |
| 12 | Enfermedad de los riñones | 3 | 3 | 9.09 | - | - |
| 13 | Estrés | 3 | 5 | 9.09 | - | - |
| 14 | Agotamiento | 3 | 5 | 9.09 | - | - |
| 15 | Dolor | 2 | 1.5 | 6.06 | - | - |
| 16 | Sentirse mal | 2 | 2 | 6.06 | - | - |
| 17 | Falta de ejercicio | 2 | 2.5 | 6.06 | - | - |
| 18 | Dificultades | 2 | 2.5 | 6.06 | - | - |
| 19 | Debilidad | 2 | 2.5 | 6.06 | - | - |
| 20 | Gripa | 2 | 2.5 | 6.06 | - | - |
| 21 | Azúcar | 2 | 2.5 | 6.06 | - | - |
| 22 | Inyecciones | 2 | 3.5 | 6.06 | - | - |
| | Suma | 77 | | 233.33 | | |
| | Índice C10 = 14.84 % | | | Índice C15 = 12.72 % | | |

Nota: los porcentajes no suman 100 % debido a que las palabras definidoras no son categorías mutuamente excluyentes.

Fuente: Elaboración propia.

Tabla 6. Red Semántica Natural CE_{PNS} y sus indicadores de consenso intra-grupo

| # | Palabra definidora | F | M | % | Conjunto SAM % ≥10 % | Conjunto SAM % ≥15 % |
|----|---------------------------|----|-----|----------------------|-------------------------|-------------------------|
| 1 | Cansancio | 7 | 3 | 24.14 | SÍ | SÍ |
| 2 | Mala alimentación | 7 | 3 | 24.14 | SÍ | SÍ |
| 3 | Diabetes | 6 | 2 | 20.69 | SÍ | SÍ |
| 4 | Dolor | 5 | 2 | 17.24 | SÍ | SÍ |
| 5 | No poder trabajar | 5 | 4 | 17.24 | SÍ | SÍ |
| 6 | Cáncer | 4 | 1.5 | 13.79 | SÍ | - |
| 7 | Chequeo médico | 3 | 2 | 10.34 | SÍ | - |
| 8 | Discapacidad | 3 | 2 | 10.34 | SÍ | - |
| 9 | Medicina | 3 | 3 | 10.34 | SÍ | - |
| 10 | Colesterol | 3 | 3 | 10.34 | SÍ | - |
| 11 | Enfermedad de los riñones | 3 | 5 | 10.34 | SÍ | - |
| 12 | Sobrepeso | 2 | 1 | 6.90 | - | - |
| 13 | Malestar | 2 | 2 | 6.90 | - | - |
| 14 | Falta de ejercicio | 2 | 2.5 | 6.90 | - | - |
| 15 | Tristeza | 2 | 3.5 | 6.90 | - | - |
| 16 | Dificultades | 2 | 3.5 | 6.90 | - | - |
| 17 | Obesidad | 2 | 3.5 | 6.90 | - | - |
| | Suma | 61 | | 210.34 | | |
| | Índice C10 = 15.86 % | | | Índice C15 = 13.10 % | | |

Nota: los porcentajes no suman 100 % debido a que las palabras definidoras no son categorías mutuamente excluyentes.

Fuente: Elaboración propia.

Es posible observar que, respecto al primer indicador de consenso intra-grupo que son los conjuntos $SAM \% \geq 10 \%$ y $\% \geq 15 \%$, el grado de consenso es mayor en la red CE_{PNS} respecto al conjunto $SAM \% \geq 10 \%$ pero es menor respecto al conjunto $SAM \% \geq 15 \%$. Para determinar el consenso intra-grupo con mayor precisión hay que observar los índices C10 y C15 en donde, como es notorio, los valores son mayores para ambos índices en la red CE_{PNS} . Evaluando las diferencias entre los índices C10 y C15 con la prueba U de Mann Whitney, se observa que, para las palabras de C10 se obtuvo $U = 50$, $p = 1.000$, *dos colas*, $r = 0$. El resultado

para las palabras de C15 en la misma prueba fue de $U = 109$, $p = 0.883$, *dos colas*, $r = -0.03$. Lo anterior indica que las diferencias en el consenso intra-grupo no son estadísticamente significativas ni en C10 ni en C15, es decir, el grado de consenso intra-grupo es igual. Para realizar el análisis de correlación Rho de Spearman se igualaron los grupos como se muestra en la Tabla 7.

Tabla 7. Grupos de palabras definidoras de la palabra estímulo Enfermedad igualados a partir de las frecuencias correspondientes para cada Perfil Sanitario

| # | % CE _{PS} | F CE _{PS} | Palabra definidora | F CE _{PNS} | % CE _{PNS} |
|----|--------------------|--------------------|---------------------------|---------------------|---------------------|
| 1 | 15.15 | 5 | Mala alimentación | 7 | 24.14 |
| 2 | 15.15 | 5 | Cansancio | 7 | 24.14 |
| 3 | 12.12 | 4 | Diabetes | 6 | 20.69 |
| 4 | 15.15 | 5 | Cáncer | 4 | 13.79 |
| 5 | 15.15 | 5 | Tristeza | 2 | 6.90 |
| 6 | 15.15 | 5 | Malestar | 2 | 6.90 |
| 7 | 18.18 | 6 | Mal humor | 1 | 3.45 |
| 8 | 6.06 | 2 | Dolor | 5 | 17.24 |
| 9 | 3.03 | 1 | No poder trabajar | 5 | 17.24 |
| 10 | 9.09 | 3 | Enfermedad de los riñones | 3 | 10.34 |
| 11 | 12.12 | 4 | Presión alta y baja | 1 | 3.45 |
| 12 | 15.15 | 5 | Alcohol y cigarros | 0 | 0 |
| 13 | 15.15 | 5 | Depresión | 0 | 0 |
| 14 | 6.06 | 2 | Dificultades | 2 | 6.90 |
| 15 | 3.03 | 1 | Chequeo médico | 3 | 10.34 |
| 16 | 3.03 | 1 | Discapacidad | 3 | 10.34 |
| 17 | 3.03 | 1 | Medicina | 3 | 10.34 |
| 18 | 6.06 | 2 | Falta de ejercicio | 2 | 6.90 |
| 19 | 9.09 | 3 | Estrés | 1 | 3.45 |
| 20 | 3.03 | 1 | Colesterol | 3 | 10.34 |
| 21 | 9.09 | 3 | Agotamiento | 1 | 3.45 |
| 22 | 3.03 | 1 | Sobrepeso | 2 | 6.90 |
| 23 | 6.06 | 2 | Sentirse mal | 1 | 3.45 |
| 24 | 6.06 | 2 | Inyecciones | 1 | 3.45 |

| # | % CE _{PS} | F CE _{PS} | Palabra definidora | F CE _{PNS} | % CE _{PNS} |
|----|--------------------|--------------------|--------------------|---------------------|---------------------|
| 25 | 6.06 | 2 | Gripa | 1 | 3.45 |
| 26 | 9.09 | 3 | Decaído | 0 | 0.00 |
| 27 | 6.06 | 2 | Debilidad | 1 | 3.45 |
| 28 | 0.00 | 0 | Obesidad | 2 | 6.90 |
| 29 | 3.03 | 2 | Azúcar | 0 | 3.45 |

Fuente: Elaboración propia.

Se consideró pertinente el cálculo de correlación con el coeficiente Rho de Spearman debido a que es el apropiado para medidas jerárquicas y lo que se expresa con los datos de la Tabla 7 es precisamente la jerarquía que cada palabra tiene dentro de cada perfil sanitario. La correlación se calculó con las frecuencias. El coeficiente Rho de Spearman obtuvo un valor $r = -0.139$ siendo éste no significativo. Este resultado sugiere que los grupos pertenecen a poblaciones distintas lo cual implica que las variaciones en uno de los grupos no pueden predecirse a partir de variaciones en el otro.

Analizando de forma no cuantitativa las palabras definidoras de cada red semántica podemos encontrar algunas peculiaridades en cada una de las redes semánticas naturales. Tanto para el Perfil Saludable como para el Perfil No Saludable la *Mala alimentación* y la *Falta de ejercicio* son cosas que se asocian con la enfermedad aunque la mala alimentación muestra una asociación considerablemente más fuerte en el Perfil No Saludable que en el Perfil Saludable. La falta de ejercicio es algo que ambos perfiles asocian y ambos de forma poco intensa con la enfermedad, Observando la tabla de diferencias y similitudes de perfiles (ver Apéndice B) podemos ver que ambos perfiles muestran una tendencia a tener poca o nula actividad física. Asimismo, entre los factores de riesgo que se asocian a la enfermedad es de destacar que el *Alcohol* y el *cigarro* tienen igual grado de asociación en el Perfil Saludable que la mala alimentación mientras que en el perfil No Saludable ni siquiera figuran como algo que se asocie a la enfermedad. Más aún, la *Presión alta* y *baja* parece ser importantes en las palabras que los participantes del Perfil Saludable asocian a la enfermedad pero tampoco figura en las asociaciones del Perfil No Saludable. Tomando en cuenta que los participantes del Perfil Saludable nunca han sido diagnosticados con problemas de presión alta puede considerarse que su comportamiento está más enfocado en evitar dichos problemas que en los participantes del Perfil No Saludable.

El *Azúcar* en la sangre parece ser algo que sólo el Perfil Saludable asocia con la enfermedad aunque de forma tangencial. Y en ambos casos la *Diabetes* es algo que

se asocia con la enfermedad; esto último es más fuerte en el Perfil No Saludable lo cual es de esperarse si se considera que hay participantes dentro del Perfil No Saludable que han sido diagnosticados con azúcar en la sangre lo cual es un indicador de diabetes. En este mismo sentido el *Sobrepeso* y la *Obesidad*, dos factores de riesgo para la diabetes, son algo que precisamente el Perfil No Saludable asocia con la enfermedad (aunque cabe señalar que no son asociaciones muy fuertes). Al respecto del sobrepeso y la obesidad cabe destacar que los participantes del Perfil Saludable se consideran sanos en ese sentido mientras que los participantes del Perfil No Saludable se reconocen como personas con sobrepeso u obesidad; es precisamente en ese punto donde puede radicar la explicación de por qué sólo para los participantes de este perfil el sobrepeso y la obesidad guarda una asociación con la enfermedad. Otra cosa que es de destacar es que, aunque los participantes del Perfil Saludable se consideran en buen estado en este perfil (y en el perfil No Saludable también) la mayoría de los hombres y mujeres tienen cinturas que miden arriba de lo saludable (≤ 90 cm en hombres y ≤ 80 cm en mujeres). Otro factor que sólo el Perfil No Saludable asocia con la enfermedad es el *Colesterol*. Sobre esto es necesario señalar que precisamente hay participantes del Perfil No Saludable que han recibido diagnósticos de colesterol alto.

Una tendencia de ambos perfiles, aunque en distinto grado (y bajo contextos distintos), es asociar con la palabra Enfermedad a distintos tipos de enfermedades específicas. En ese sentido, los conceptos *Cáncer*, *Diabetes* y *Enfermedad de los riñones* son asociados con la enfermedad en ambos grupos, aunque sólo los participantes del Perfil No Saludable han recibido diagnósticos en esas tres enfermedades. Además, aunque de forma débil, el Perfil Saludable asocia la *Gripa* con la enfermedad, de nuevo otra enfermedad específica asociada.

Otro punto importante en las asociaciones dadas por los perfiles sanitarios radica en el hecho de que para ambos perfiles el concepto *Cansancio* se asocia con la enfermedad y, en el caso del Perfil Saludable, el *Agotamiento* también guarda una asociación con la palabra estímulo pero de menor intensidad. De forma similar, el concepto de *Dificultades* se asocia en ambos perfiles con la palabra estímulo Enfermedad. No obstante, en el Perfil No Saludable la asociación es menos intensa. Sobre este punto hay que considerar que alrededor de la mitad de los participantes que conforman el Perfil No Saludable se encuentran limitados por su salud para realizar tareas tales como subir un solo piso de escalera, agacharse o arrodillarse y caminar distancias de un kilómetro o más. En el Perfil Saludable esta condición no se presenta y sin embargo la asociación con el concepto dificultades es mayor. Lo anterior puede estar relacionado con otros dos conceptos que son mencionados los cuales son *Debilidad* y *Discapacidad*. El concepto debilidad es asociado sólo por

los participantes del Perfil Saludable mientras que el concepto discapacidad es asociado solamente por los participantes del Perfil No Saludable.

Una posible explicación de estas asociaciones radica en que para los participantes del Perfil Saludable no hay “discapacidades” que les impidan moverse o realizar sus actividades cotidianas, a lo sumo podrían considerar debilidades a las áreas de mayor dificultad para desempeñarse. En cambio, para el Perfil No Saludable el cual está conformado por participantes que sí se encuentran limitados por su salud, las dificultades para la realización de las actividades cotidianas pudieran representar para ellos una discapacidad o hacerlos sentir tal condición. También se puede observar, en las distintas palabras asociadas a la palabra estímulo, que hay asociaciones con estados de malestar no especificados, esto se observa al considerar las palabras *malestar* y *dolor*, las cuales se presentan en ambos perfiles sanitarios, así como en la presencia de la palabra *sentirse mal*, aunque ésta sólo se presenta en el Perfil Saludable.

En lo referente a posibles asociaciones de estados psicológicos con la palabra Enfermedad, se observa que éstos tienen gran presencia en las asociaciones del Perfil Saludable pues se presentan las palabras *Mal humor*, *Depresión*, *Tristeza*, *Decaído* y *Estrés*. Estos conceptos se asocian también con el concepto mal humor es el que se encuentra más fuertemente asociado a la palabra Enfermedad, en el Perfil Saludable. Al respecto de esto se destaca que para el Perfil No Saludable no existen dichas asociaciones, salvo por el concepto *tristeza* aunque aún en este caso se habla de una asociación débil. Un punto importante que no se puede dejar de lado es el hecho de que los conceptos que tienen que ver con medicinas o médicos parecen tener presencia sólo en el Perfil No Saludable con las palabras *Chequeo médico* y *Medicinas*. En ese respecto, en el Perfil Saludable solamente figura la palabra *Inyecciones* siendo, además, la palabra con la asociación más débil de todas.

La última asociación con la palabra Enfermedad que se presenta es la de *No poder trabajar*. Ésta asociación sólo se presenta en el Perfil No Saludable y su grado de asociación es de los más altos. Resulta importante recordar que es precisamente este perfil el que, debido a su salud, se puede llegar a ver impedido para realizar su trabajo mientras que para el Perfil Saludable las circunstancias son distintas.

Las razones por las que los participantes de los distintos perfiles asocian algo con la palabra enfermedad pueden ser varias. Una de ellas es que relacionan aquello que conocen como enfermedades que han padecido o condiciones que viven. Esto último puede ser también el motivo por el que las asociaciones de algunos conceptos como diabetes, sobrepeso y obesidad. No obstante, resulta difícil concluir que ese motivo predomina en todas las asociaciones; por ejemplo, el concepto Cáncer se

asocia a la enfermedad en ambos perfiles pero los participantes del Perfil Saludable lo asocian más fuertemente que los participantes del Perfil No Saludable cuando es en este último perfil donde hay participantes que han sido diagnosticados con cáncer. De igual forma, es notoria la fuerza con la que conceptos como *depresión*, *estrés* y *tristeza* se manifiestan en los participantes del Perfil Saludable mientras que su presencia es prácticamente nula en el Perfil No Saludable.

La red CS_{PS} corresponde a $n = 33$ y está conformada por un total de 147 palabras definidoras ($J = 147$). Asimismo, la red CS_{PNS} corresponde a $n = 29$ y está conformada por un total de 116 palabras definidoras ($J = 116$). Las Tablas 8 y 9 muestran un resumen de las redes semánticas naturales y de los indicadores de consenso intra-grupo de las redes CS_{PS} y CS_{PNS} respectivamente.

Tabla 8. Red Semántica Natural CS_{PS} y sus indicadores de consenso intra-grupo

| # | Palabra definidora | F | M | % | Conjunto SAM % ≥10 % | Conjunto SAM % ≥15 % |
|----|---------------------|----|-----|-------|-------------------------|-------------------------|
| 1 | Buena alimentación | 12 | 2 | 36.36 | SÍ | SÍ |
| 2 | Saludable | 11 | 3 | 33.33 | SÍ | SÍ |
| 3 | Ejercicio | 9 | 2 | 27.27 | SÍ | SÍ |
| 4 | Bienestar | 9 | 3 | 27.27 | SÍ | SÍ |
| 5 | No tener enfermedad | 6 | 3.5 | 18.18 | SÍ | SÍ |
| 6 | Sentirse bien | 5 | 2 | 15.15 | SÍ | SÍ |
| 7 | Acudir al médico | 5 | 4 | 15.15 | SÍ | SÍ |
| 8 | Alegría | 5 | 4 | 15.15 | SÍ | SÍ |
| 9 | Tomar agua | 4 | 1.5 | 12.12 | SÍ | - |
| 10 | Ánimo | 4 | 3 | 12.12 | SÍ | - |
| 11 | Actividad | 3 | 2 | 9.09 | - | - |
| 12 | Trabajo | 3 | 2 | 9.09 | - | - |
| 13 | Dormir bien | 3 | 3 | 9.09 | - | - |
| 14 | Vitaminas | 3 | 4 | 9.09 | - | - |
| 15 | Medicinas | 3 | 5 | 9.09 | - | - |
| 16 | Caminar | 2 | 1.5 | 6.06 | - | - |
| 17 | Enfermedad | 2 | 2 | 6.06 | - | - |
| 18 | Fuerzas | 2 | 3 | 6.06 | - | - |

| # | Palabra definidora | F | M | % | Conjunto SAM % ≥10 % | Conjunto SAM % ≥15 % |
|----|----------------------|-----|-----|--------|-------------------------|-------------------------|
| 19 | Actitud positiva | 2 | 3.5 | 6.06 | - | - |
| 20 | Deporte | 2 | 3.5 | 6.06 | - | - |
| 21 | Tranquilidad | 2 | 4 | 6.06 | - | - |
| 22 | Estar delgado | 2 | 4 | 6.06 | - | - |
| 23 | No fumar | 2 | 5 | 6.06 | - | - |
| | Suma | 101 | | 306.06 | | |
| | Índice C10 = 21.21 % | | | | Índice C15 = 17.17 % | |

Nota: los porcentajes no suman 100 % debido a que las palabras definidoras no son categorías mutuamente excluyentes.

Fuente: Elaboración propia.

Tabla 9. Red Semántica Natural CS_{PNS} y sus indicadores de consenso intra-grupo

| # | Palabra definidora | F | M | % | Conjunto SAM % ≥10 % | Conjunto SAM % ≥15 % |
|----|---------------------|----|-----|-------|-------------------------|-------------------------|
| 1 | Bienestar | 15 | 3 | 51.72 | SÍ | SÍ |
| 2 | No tener enfermedad | 13 | 3 | 44.83 | SÍ | SÍ |
| 3 | Buena alimentación | 12 | 3 | 41.38 | SÍ | SÍ |
| 4 | Ejercicio | 8 | 2 | 27.59 | SÍ | SÍ |
| 5 | Alegría | 5 | 2 | 17.24 | SÍ | SÍ |
| 6 | Sentirse bien | 5 | 3 | 17.24 | SÍ | SÍ |
| 7 | Acudir al médico | 5 | 3 | 17.24 | SÍ | SÍ |
| 8 | Nietos | 2 | 1 | 6.90 | - | - |
| 9 | Actividad | 2 | 1.5 | 6.90 | - | - |
| 10 | Tranquilidad | 2 | 1.5 | 6.90 | - | - |
| 11 | Hijos | 2 | 1.5 | 6.90 | - | - |
| 12 | Saludable | 2 | 2.5 | 6.90 | - | - |
| 13 | Trabajo | 2 | 2.5 | 6.90 | - | - |
| 14 | Tomar agua | 2 | 3 | 6.90 | - | - |
| 15 | Enfermedad | 2 | 3 | 6.90 | - | - |
| 16 | Actitud positiva | 2 | 3 | 6.90 | - | - |
| 17 | Dormir bien | 2 | 4 | 6.90 | - | - |
| 18 | Control | 2 | 4 | 6.90 | - | - |

| # | Palabra definidora | F | M | % | Conjunto SAM % ≥10 % | Conjunto SAM % ≥15 % |
|---|----------------------|----|---|--------|-------------------------|-------------------------|
| | Suma | 85 | | 293.10 | | |
| | Índice C10 = 23.79 % | | | | Índice C15 = 18.16 % | |

Nota: los porcentajes no suman 100 % debido a que las palabras definidoras no son categorías mutuamente excluyentes.

Fuente: Elaboración propia.

En el primer indicador de consenso intra-grupo que es el de conjuntos SAM % $\geq 10\%$ y % $\geq 15\%$, el grado de consenso es mayor en la red CS_{PS} respecto al conjunto SAM % $\geq 10\%$ y también respecto al conjunto SAM % $\geq 15\%$. No obstante, al observar los índices C10 y C15 se destaca que ambos índices son mayores en la red CS_{PNS} que en la red CS_{PS} . Evaluando las diferencias entre los índices C10 y C15 con la prueba U de Mann Whitney, se observa que, para las palabras de C10 se obtuvo $U = 47, p = 0.820, dos\ colas, r = -0.05$ mientras que el resultados para las palabras de C15 en la misma prueba es $U = 88, p = 0.303, dos\ colas, r = -0.19$. Lo anterior indica que las diferencias en lo relativo al consenso intra-grupo no son estadísticamente significativas ni en C10 ni en C15, es decir, el grado de consenso intra-grupo es igual para las redes semánticas del concepto de Salud. Para realizar el análisis de correlación Rho de Spearman se igualaron los grupos, de la misma forma que con las redes CE_{PS} y CE_{PNS} , como se muestra en la Tabla 10.

Tabla 10. Grupos de palabras definidoras de la palabra estímulo Salud igualados a partir de las frecuencias correspondientes para cada Perfil Sanitario

| # | % CE_{PS} | F CE_{PS} | Palabra definidora | F CE_{PNS} | % CE_{PNS} |
|----|-------------|-------------|---------------------|--------------|--------------|
| 1 | 27.27 | 9 | Bienestar | 15 | 51.72 |
| 2 | 36.36 | 12 | Buena alimentación | 12 | 41.38 |
| 3 | 18.18 | 6 | No tener enfermedad | 13 | 44.83 |
| 4 | 27.27 | 9 | Ejercicio | 8 | 27.59 |
| 5 | 33.33 | 11 | Saludable | 2 | 6.90 |
| 6 | 15.15 | 5 | Alegría | 5 | 17.24 |
| 7 | 15.15 | 5 | Sentirse bien | 5 | 17.24 |
| 8 | 15.15 | 5 | Acudir al médico | 5 | 17.24 |
| 9 | 12.12 | 4 | Tomar agua | 2 | 6.90 |
| 10 | 9.09 | 3 | Actividad | 2 | 6.90 |

| # | % CE _{PS} | F CE _{PS} | Palabra definidora | F CE _{PNS} | % CE _{PNS} |
|----|--------------------|--------------------|--------------------|---------------------|---------------------|
| 11 | 9.09 | 3 | Trabajo | 2 | 6.90 |
| 12 | 9.09 | 3 | Dormir bien | 2 | 6.90 |
| 13 | 6.06 | 2 | Tranquilidad | 2 | 6.90 |
| 14 | 6.06 | 2 | Enfermedad | 2 | 6.90 |
| 15 | 6.06 | 2 | Actitud positiva | 2 | 6.90 |
| 16 | 9.09 | 3 | Medicina | 1 | 3.45 |
| 17 | 9.09 | 3 | Vitaminas | 1 | 3.45 |
| 18 | 12.12 | 4 | Ánimo | 0 | 0 |
| 19 | 3.03 | 1 | Control | 2 | 6.90 |
| 20 | 6.06 | 2 | Fuerzas | 1 | 3.45 |
| 21 | 6.06 | 2 | Caminar | 1 | 3.45 |
| 22 | 0 | 0 | Nietos | 2 | 6.90 |
| 23 | 0 | 0 | Hijos | 2 | 6.90 |
| 24 | 6.06 | 2 | Deporte | 0 | 0 |
| 25 | 6.06 | 2 | Estar delgado | 0 | 0 |
| 26 | 6.06 | 2 | No fumar | 0 | 0 |

Fuente: Elaboración propia.

La correlación se calculó de la misma forma que para las redes CE_{PS} y CE_{PNS}. El resultado se presenta a continuación. El coeficiente Rho de Spearman fue $r=0.589$ con $p=0.002$. El resultado sugieren que los grupos pertenecen a poblaciones similares aunque no logra ser una correlación alta. Esto indica que las variaciones en uno de los grupos pueden predecirse hasta cierto punto a partir de variaciones en el otro, a diferencia de lo ocurrido con las redes CE_{PS} y CE_{PNS} donde las asociaciones dadas en un perfil no se correlacionan con las del otro. Estos resultados pueden indicar que, de manera general, los distintos perfiles sanitarios asocian conceptos con la palabra estímulo Salud de forma más consistente que con la palabra Enfermedad.

Ahora bien, analizando de forma no cuantitativa las palabras definidoras de cada red semántica podemos percatarnos de más características en las asociaciones dadas. Primeramente, se puede observar que en ambos perfiles existen asociaciones similares referentes a la *Buena alimentación, el Ejercicio, Acudir al médico, Tomar agua y Dormir bien* aunque dichas asociaciones tienen una mayor jerarquía en el Perfil Saludable. Aunado a lo anterior, destaca el hecho de que sólo en el Perfil Saludable se presentaron asociaciones con los conceptos de *Caminar, Deporte y No fumar*. En términos conductuales este tipo de diferencias en las asociaciones pueden

representar cambios significativos en los esquemas de conducta. Una tendencia que parece presentarse en ambos perfiles es la de asociar la Salud con estados de *Bienestar*, con *Sentirse bien* o simplemente con estar *Saludable*. Aunque en ambos perfiles las asociaciones con estos conceptos son altas, en el Perfil No Saludable el concepto Bienestar es el que guarda la asociación más fuerte. De igual manera, se destaca que para ambos perfiles la Salud se asocia con *No tener enfermedad*. Sobre esto mismo es necesario hacer notar que las creencias sobre Salud parecen contemplar fuertemente el bienestar y no solamente la ausencia de enfermedad, como convencionalmente se ha entendido desde una perspectiva estrictamente biomédica. En lo relacionado a estados de ánimo, se puede observar que ambos perfiles contemplan la *Alegría y la Actitud positiva* como conceptos que se asocian a la Salud. También, para el Perfil Saludable el concepto *Ánimo* guarda una relación con la palabra estímulo. Otro resultado que se puede observar es que los conceptos de *Actividad y Trabajo* guardan relación con la salud en ambos perfiles y, en el caso del Perfil Saludable, también la palabra *Fuerzas* guarda asociación. Los conceptos *Vitaminas y Medicinas* se presentan en el Perfil Saludable aunque con poca intensidad. Por otro lado, en ambos perfiles se asocia la palabra *Enfermedades* con la palabra estímulo Salud.

Finalmente, destacan dos aspectos particulares que se presentan en las redes analizadas y es el hecho de que sólo en el Perfil No Saludable se presenta asociación con los conceptos de *Nietos e Hijos*. Estos conceptos presentan una asociación media. Asimismo, para ambos perfiles sanitarios parece ser importante la *Tranquilidad* en lo relativo a la salud, aunque esto último es más claro para el Perfil No Saludable.

Discusión

La TRSN permite indagar en asociaciones que los grupos de participantes realizan referentes a una palabra estímulo en concreto. En este sentido, las redes semánticas naturales representan un instrumento útil para el estudio de las creencias colectivas pero resulta insuficiente para estudiar todo el proceso de comunicación de información y, por ende, las conclusiones sobre las posibles causas de las representaciones sociales obtenidas no pueden explorarse con dicha técnica. No obstante, a manera de contrapeso ante tal limitación es de destacar que durante este proyecto ha sido de vital importancia la elaboración teórica referente a las creencias. Con los planteamientos teóricos que se manejan en este trabajo se aclaran algunos puntos importantes y se abre la posibilidad de trabajos posteriores, de índole empírica y de mayor profundidad.

Dicho lo anterior, es posible contextualizar los resultados en función de los objetivos del trabajo así como de la teoría de las representaciones sociales y el modelo psicológico de la salud que se maneja en esta tesis. Primeramente, la

hipótesis de trabajo H1: *Las creencias colectivas sobre salud, enfermedad y conductas de riesgo en habitantes con perfil sanitario saludable difieren de las creencias colectivas de los habitantes con perfil sanitario no saludable en habitantes de los municipios de Támico y Madero, Tamaulipas*; se acepta de forma parcial. Puede observarse en los resultados de las redes semánticas naturales, sobre todo en las redes CE_{PS} y CE_{PNS} donde no se presenta correlación entre las creencias de ambos perfiles, que hay fundamentalmente dos tipos de diferencias en las creencias colectivas sobre Salud y Enfermedad; 1) las palabras o proposiciones asociadas a las palabras estímulo Salud y Enfermedad en un perfil no son exactamente las mismas en el otro perfil, y 2) la intensidad con la que se presentan las asociaciones en un perfil no son iguales en el otro aunque se trate de la misma palabra. Si volvemos a la definición de representación social que se esbozó en un principio como *la finalidad que tienen determinadas actitudes y creencias socialmente compartidas sobre un objeto social para los individuos dentro de un determinado grupo social*; observamos las características de cada Perfil Sanitario y recordamos la forma en la que se integra dicho concepto de representaciones sociales en el modelo de la dimensión psicológica de la salud sobre la cual se está trabajando, es posible notar que las creencias del Perfil No Saludable dan poco o nulo lugar a conceptos que podrían estar ligados con actitudes saludables. Concretamente, el Perfil No Saludable no muestra creencias que indiquen un accionar preventivo en su conducta. Más aún, tampoco el Perfil Saludable muestra creencias que puedan considerarse favorables en términos de prevención de enfermedades o accidentes. Las redes semánticas naturales ofrecen información enriquecedora en ese sentido pues no sólo las creencias colectivas reportadas resultan esclarecedoras sobre el panorama semántico de los participantes sino que, además, términos como seguridad, prevención, planificación de la salud, etc., simplemente no aparecen, lo cual también es un indicador bastante sólido de hacia qué dirección van las creencias colectivas de ambos perfiles.

Otro punto importante para dar una interpretación adecuada a los resultados obtenidos radica en las características de los perfiles. Si se observa a detalle es posible notar que, pese a la denominación utilizada para este trabajo, los participantes del Perfil Saludable bien podrían considerarse participantes en riesgo debido a que en muchos aspectos no presenta diferencias con el Perfil No Saludable. Este hecho también puede explicar parcialmente por qué las creencias sobre salud en ambos perfiles mostraron una correlación significativa, en realidad en ambos casos hay participantes con indicadores de riesgo a la salud, por ejemplo, el ancho de la cintura (ver diferencias y similitudes de los perfiles en el Apéndice B). Ahora bien, evaluar las diferencias cualitativas dadas por cada perfil en las redes semánticas naturales resulta una tarea difícil si se contempla la forma tradicional de interpretación de las

redes semánticas. Es recomendable que, en aproximaciones ulteriores, se contemple la posibilidad de analizar la información proporcionada por las redes semánticas naturales de una forma más precisa, con metodologías cuasi-experimentales o experimentales, puesto que el hecho de que existan diferencias cuantitativas entre redes de distintos grupos no implica necesariamente una diferencia cualitativa importante. Finalmente, la extrapolación de los resultados a otras poblaciones así como las comparaciones con otros trabajos pueden resultar riesgosas si no se parte de una misma base conceptual. En este respecto, es importante aclarar que la base conceptual de la que se parte aquí NO está presente en otros trabajos, lo que implica que, en relación a los estudios de representaciones sociales (creencias colectivas) sobre salud y enfermedad existentes, no hay conmensurabilidad. En este caso el trabajo se limita a ofrecer una descripción del perfil sanitario y las creencias de una determinada población y a explorar algunas correlaciones, no obstante, se busca proponer una forma de trabajar las representaciones sociales, a saber, trabajar en términos de creencias y actitudes, y definir estos, a su vez, en términos de proposiciones y disposiciones para actuar respectivamente. Lo anterior no es trivial si se reconocen las marcadas divergencias que hay entre los trabajos sobre representaciones sociales respecto a cómo se definen las representaciones sociales y, por ende, qué es lo que en verdad dicen sobre la realidad.

De este trabajo se pueden concluir cosas en dos sentidos, uno de índole teórico-metodológico y otro de índole práctico. En el primer sentido ya se ha mencionado la necesidad de elaboraciones teóricas no ambiguas ni multívocas. No obstante, es menester señalar la necesidad de una aproximación más precisa para manejar variables tan complejas como las creencias colectivas. Una aproximación mediante diseños de investigación experimentales así como mediante formas de análisis multivariado permitirá tener mayor claridad sobre el fenómeno de las creencias y de cómo éstas afectan la conducta aunque de esta aproximación descriptiva mediante un diseño no experimental transversal se han obtenido buenos indicadores para posteriores intervenciones preventivas. En el segundo sentido, se puede concluir que las creencias de la población estudiada no parecen contemplar elementos importantes que pudieran facilitar la adquisición de hábitos saludables en la población. Las discrepancias entre el estado de salud real de los participantes y lo que ellos reportan de sí mismos también resulta alarmante si se contempla que un porcentaje alto de participantes que ha recibido diagnósticos oficiales de algún padecimiento médico no se considera en situación de riesgo. También es destacable la prioridad que se le da a los estados de ánimo o estados subjetivos de bienestar en relación con los conceptos de salud y enfermedad. Asociaciones de este tipo pueden representar un área de oportunidad sobre la cual trabajar debido a que la

inacción en lo relativo al autocuidado de la salud puede deberse en gran medida a que los participantes de hecho no se “sientan” enfermos.

Referencias

- Acosta, L. (2010). *Conceptualización del aborto en el gremio médico a través de la técnica de redes semánticas naturales* (Tesis). Universidad Nacional Autónoma de México, México, D. F.
- Aguilera, H. E. (2010). “El uso de las redes semánticas naturales en las representaciones sociales de la responsabilidad”. *Revista Internacional de Psicología*, 11(2), 1-13.
- Álvarez-Gayou, J. L. (2003). *Cómo hacer investigación cualitativa. Fundamentos y metodología*. México, D.F.: Paidós.
- Bermúdez, P., Medina-Mora, M. E., Berenzon, S., García, B., Amador, N. y Pérez, C. (2014). “Significado psicológico que estudiantes de secundaria de la Ciudad de México atribuyen al concepto drogas”. *Revista Intercontinental de Psicología y Educación*, 16(1), 11-36.
- Castañeda, A. (2016). “Las redes semánticas naturales como estrategia metodológica para conocer las representaciones sociales acerca de la investigación en el contexto de la formación profesional de los comunicadores”. *Estudios sobre las Culturas Contemporáneas*, XXII(43), 123-168.
- Consuegra, N. (2010). *Diccionario de psicología*. Bogotá, Colombia: Ecoe Ediciones.
- Denegri, M., Cabezas, D., Sepúlveda, J., Del Valle, C., González, Y. y Miranda, H. (2010). “Representaciones sociales sobre pobreza en estudiantes universitarios chilenos”. *LIBERABIT*, 16(2), 161-170.
- Figueroa, J. P., González, E. G. y Solís, V. M. (1976). “An approach to the trouble of meaning”. *Journal of Psycholinguistic Research*, 5(2), 107-117.
- García, B. y Jiménez, S. (1996). “Redes semánticas de los conceptos de presión y flotación en estudiantes de bachillerato”. *Revista Mexicana de Investigación Educativa*, 1(2), 343-361.
- Hernández, F. (1998). “Aplicación de la Técnica de Redes Semánticas a Tres Conceptos Asociados a las Relaciones de Pareja: Amor, Celos y Temor” (Tesis de licenciatura). Recuperado de http://www.ciad.mx/archivos/desarrollo/publicaciones/Tesis_asesoradas/Licenciatura/19.pdf
- Hinojosa, G. (2008). “El tratamiento estadístico de las redes semánticas naturales”. *Revista Internacional de Ciencias Sociales y Humanidades, SOCIOTAM*, XVIII(1) 133-154. Recuperado de <http://www.redalyc.org/articulo.oa?id=65411190007>
- Instituto Nacional de Estadística y Geografía. (s.f.). “Marco Geoestadístico Nacional. Fundamentos y orígenes”. Recuperado de <http://www.INEGI.org.mx/geo/contenidos/geoestadistica/catorcen.aspx>
- _____ (2012). “Sistema Para la Consulta de Información Censal 2010” (Versión 05/2012) [Mapa interactivo en línea]. Recuperado de <http://gaia.INEGI.org.mx/scince2/viewer.html>

- International Business Machines (Versión 20) [Software de computación]. Armonk, N.J., Estados Unidos: Statistical Package for the Social Sciences.
- Kerlinger, F. N. y Lee, H. B. (2002). *Investigación del Comportamiento*. México, D.F.: McGraw-Gill.
- León, M. y Páez, D. (2002). “Representaciones Sociales de la Enfermedad de Chagas en Comunidades en Riesgo: Creencias, Actitudes y Prevención”. *Revista Interamericana de Psicología* 36(1-2), 215-236. Recuperado de <http://www.redalyc.org/articulo.oa?id=28436213>
- Microsoft. (Versión 2013) [Software de computación]. Redmond, Washington, Estados Unidos: Office Excel.
- Montano-Alarcón, R. A. (2015). “Representaciones sociales en torno a educación ambiental para la sustentabilidad en alumnos de pedagogía de la universidad veracruzana”. En J. Agüero, B. Torres, (eds.) *Educación Ambiental desde la Innovación, la Transdisciplinariedad e Interculturalidad, Tópicos Selectos de Educación Ambiental* (pp. 300-310). Veracruz: ECORFAN.
- Moscovici, S. (1979). *El psicoanálisis, su imagen y su público*. Buenos Aires, Argentina: Editorial Huemul.
- Organización Mundial de la Salud. (2013). “Plan de acción mundial para la prevención y el control de las enfermedades no transmisibles 2013-2020”. Recuperado de http://www.who.int/cardiovascular_diseases/15032013_updated_revised_draft_action_plan_spanish.pdf
- Organización Mundial de la Salud. (2015). “Enfermedades no transmisibles” (Nota descriptiva). Recuperado de <http://www.who.int/mediacentre/factsheets/fs355/es/>
- Organización para la Cooperación y el Desarrollo Económicos. (2014). “Estadísticas de la OCDE sobre la salud 2014 México en comparación”. Recuperado de <http://www.oecd.org/els/health-systems/Briefing-Note-MEXICO-2014-in-Spanish.pdf>
- Pérez, C. (2004). *Técnicas de Análisis Multivariante de Datos. Aplicaciones con SPSS*. Madrid, España: Pearson Educación.
- Pérez, O. A. (2015). “Redes semánticas naturales: anotaciones metodológicas para el análisis de las representaciones sociales”. *Revista Contribuciones a las Ciencias Sociales*, 27. Recuperado de: <http://www.eumed.net/rev/ccss/2015/01/redes-semanticas.html>
- Reyes, I. (1993). “Las Redes Semánticas Naturales, su Conceptualización y su Utilidad en la Construcción de Instrumentos”. *Revista de Psicología Social y Personalidad*, IX(1), 81-97.
- Sánchez, M. P., De la Garza, A. y López, E. O. (2013). “Simulaciones computacionales sobre cuestiones ambientales en dos grupos de contraste”. *Liberabit. Revista de Psicología*, 19(2), 223-233.
- Urbina, M., Mogurl, A., Muñiz, M. E. y Solís, J. A. (Coord.). (2006). *La experiencia mexicana en salud pública oportunidad y rumbo para el tercer milenio*. México, D.F.: Fondo de Cultura Económica.

- Urrego, M. A. (2016). “Las redes semánticas naturales como un método alternativo para el diagnóstico de imagen de marca”. En C. Atehortua., C. A. Castaño., L. Ortegón., O. J. Robayo & S. P. Rojas. (2016). *Aportes y reflexiones. Gerencia estratégica de Mercadeo* (565-581). Bogotá: Editorial Politécnico Gran colombiano.
- Valle, M. C. (2003). “Redes semánticas en la elección de carrera en alumnos de tercero de secundaria” (Tesis). Universidad Pedagógica Nacional, México, D.F.
- Vera-Noriega, J. Á., Batista de Albuquerque, F. J. y Pimentel, C. E. (2005). “Redes semánticas: aspectos teóricos, técnicos, metodológicos y analíticos”. *Ra Ximhai*, 1(3) 439-451. Recuperado de <http://sociales.redalyc.org/articulo.oa?id=46110301>

Apéndice A. Cuestionario para conocer el perfil sanitario y las creencias sobre los conceptos de salud y de enfermedad



UAT



Unidad Académica
de Ciencias Jurídicas
y Sociales
Universidad Autónoma de Tamaulipas

Fecha de aplicación: ___/___/___ Folio: ___

La presente encuesta es para recopilar información sobre el perfil de salud para el proyecto *Rumor y creencias colectivas sobre salud y enfermedad en habitantes con distinto perfil sanitario en los municipios de Tampico y Madero*. La información que nos proporcione servirá para diseñar soluciones a problemas de salud en la región. Es una encuesta anónima por lo que se solicita contestar con la mayor honestidad posible. De antemano agradecemos su participación voluntaria.

Apartado I. Datos generales

Sexo: Masculino ___ Femenino ___ **Edad:** ___

Estado civil:

Soltero/a ___ Casado/a ___ Viudo/a ___

Divorciado/a ___ Unión libre ___

Escolaridad:

Primaria ___ Secundaria ___

Preparatoria/Bachillerato ___

Universidad ___ Posgrado ___ Ninguno ___

Dirección:

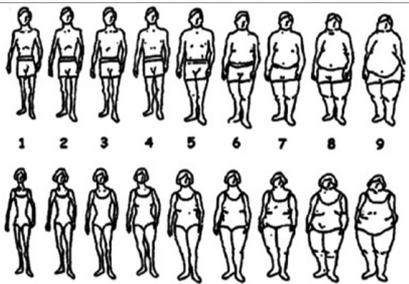
Municipio: Tampico ___ Madero ___ Colonia: _____

Teléfono: _____ Celular: _____

Aplicador: _____

A continuación se presenta una serie de preguntas enfocadas a conocer su perfil de salud, responda de la manera más honesta posible recordando que la información que proporcione es completamente anónima.

Apartado II. Perfil de salud

| Pregunta | Respuesta | |
|---|---|----|
| Actividad física | | |
| 1. Durante los últimos 7 días, ¿cuántos días realizó usted alguna actividad física moderada? | No realiza actividades moderadas | |
| | 1 2 3 4 5 6 7 | |
| 2. ¿Practica algún deporte? | Sí | No |
| 3. ¿Pasa más de 5 horas diarias frente al televisor, videojuego o PC? | Sí | No |
| 4. Su salud actual, ¿le impide correr, levantar objetos pesados, o participar en deportes agotadores? | Sí | No |
| Obesidad y sobrepeso | | |
| 5. Ahora le voy a mostrar unas figuras corporales, por favor dígame qué figura siente que más se parece a usted en este momento. Respuesta_____ | | |
|  | | |
| 6. ¿Considera usted que... | | |
| tiene sobrepeso? | Sí | No |
| tiene obesidad? | Sí | No |
| está bajo de peso? | Sí | No |
| está bien de peso? | Sí | No |
| 7. ¿Cuánto mide su cintura (Saludable es <80cm en mujeres y <90cm en hombres)? | Respuesta_____ | |
| Diabetes mellitus | | |
| 8.- ¿Sigue algún tratamiento para controlar su peso? | Sí | No |
| 9. ¿Algún médico le ha dicho que tiene diabetes o azúcar en la sangre? | Sí | No |
| | Sí, durante el embarazo (Diabetes gestional). | |

| Pregunta | Respuesta | |
|---|---|----|
| Hipertensión arterial | | |
| 10. ¿Algún médico le ha dicho que tiene la presión alta? | Sí | No |
| Enfermedad cardiovascular | | |
| 11. ¿Le ha dicho el médico que usted tiene alguna afección cardíaca? | Sí | No |
| Enfermedad renal, hipercolesterolemia y neuromuscular | | |
| 12. ¿Alguna vez un médico le ha dicho que padece alguna enfermedad del riñón? | Sí | No |
| 13. ¿Algún médico le ha dicho que tiene colesterol alto? | Sí | No |
| 14. ¿Algún médico le ha dicho que tiene los triglicéridos altos? | Sí | No |
| 15. ¿Le ha dicho su médico si ha tenido una embolia cerebral? | Sí | No |
| Cáncer | | |
| 16. ¿Alguna vez le ha dicho un médico o un dentista que tiene (o tuvo) cáncer o un tumor maligno de cualquier tipo? | Sí | No |
| 17. ¿Cuál es la situación actual del cáncer? | Con enfermedad | |
| | Sin enfermedad | |
| | El cáncer está cediendo, va mejor (en remisión) | |
| Accidentes | | |
| 18. ¿Sufrió usted algún daño a su salud a causa de un accidente en los últimos 12 meses? | Sí | No |
| 19. ¿Cómo fue que se accidentó? | Choque de o entre vehículos de transporte | |
| | Atropellamiento | |
| | Otros accidentes de transporte: | |
| | Caída de un mismo nivel | |
| | Exposición a fuerzas mecánicas inanimadas | |
| | Caída de un nivel a otro | |
| | Exposición a fuerzas mecánicas animadas | |

| | |
|---------------------------------|---|
| 19. ¿Cómo fue que se accidentó? | Exposición a corriente eléctrica, radiación, temperatura y presión atmosférica extrema. |
| | Exposición a fuego, humo y llamas |
| | Contacto con calor o sustancias calientes |
| | Exposición a fuerzas de la naturaleza |
| | Contacto traumático con animales y plantas venenosos |
| | Envenenamiento accidental |
| | Ahogamiento o sumersión accidental |
| | Otros accidentes que obstruyen la respiración |
| | Exposición accidental a otros factores |

Apartado III. Creencias salud-enfermedad

3.1 A continuación se te presentan los conceptos de Salud y Enfermedad, escribe bajo cada uno de ellos cinco palabras que asocies o piensas sobre cada uno, posteriormente asígnales un número del 1 al 5 a partir de qué tanto representan al concepto en cuestión, de tal manera el número 1 será para el que mejor represente lo que pienses sobre la Salud o la Enfermedad.

| Salud | Enfermedad |
|----------|------------|
| _____ () | _____ () |
| _____ () | _____ () |
| _____ () | _____ () |
| _____ () | _____ () |
| _____ () | _____ () |

3.2 ¿Qué es para usted la salud?

3.3 ¿Qué significa para usted estar sano (a)?

3.4 ¿Qué es para usted la enfermedad?

3.5 ¿Qué significa para usted estar enfermo (a)?

Apéndice B. Cuadro comparativo entre las características del Perfil No Saludable y el Perfil Saludable.

A continuación se muestra un cuadro comparativo en donde se describen las características de cada perfil en las áreas donde se encontraron diferencias estadísticamente significativas al 0.05. Las pruebas estadísticas fueron la t de student para comparación con variables intervalares y de razón que sí cumplieron los supuestos de parametría; la prueba U de Mann-Whitney para la comparación de las variables ordinales y para las variables intervalares y de razón que no cumplieron los supuestos de parametría; la prueba Ji cuadrada de contingencias para las variables nominales en donde al menos una de las variables de la comparación tenía más de 2 niveles; la prueba Ji cuadrada de contingencias con corrección de Yates para todas las comparaciones con variables dicotómicas cuyas casillas tuvieron frecuencias esperadas mayores a 5; la prueba exacta de Fisher para comparaciones en donde al menos una de las casillas obtuvo frecuencias esperadas menores a 5. Cabe aclarar que los resultados que no fueron estadísticamente significativos en el análisis de los ítems del apartado *Perfil Sanitario y Percepción de la Salud* NO carecen de valor para el presente trabajo, simplemente que, para efectos de clasificar y describir a los participantes en función del Perfil Sanitario, es necesario considerar las diferencias entre ambos perfiles más que las similitudes. No obstante, en la tabla comparativa que se presenta a continuación, también se incluyen una síntesis de los resultados que no mostraron ser estadísticamente significativos en la clasificación y descripción de los participantes, es decir, aquellos resultados en los que el comportamiento de los participantes fue el mismo independientemente del Perfil Sanitario en el que se encuentran clasificados (similitudes).

Cuadro comparativo entre las características del Perfil No Saludable y el Perfil Saludable

| Sub- área | Perfil sanitario | | |
|--------------|---|--|---|
| | Perfil No Saludable (n ₁ = 29) | Perfil Saludable (n ₂ = 33) | |
| A F | D | -La mayoría de los participantes está limitado por su salud para realizar esfuerzos intensos como correr, levantar objetos pesados o participar en deportes agotadores | -Son participantes que, en su mayoría, no se encuentran limitados por su salud para realizar esfuerzos intensos como correr, levantar objetos pesados o participar en deportes agotadores |
| | S | <ul style="list-style-type: none"> - En ambos perfiles predominan los participantes que, en los 7 días previos a ser encuestados, no realizaron actividades físicas moderadas. - En ambos perfiles predominan los participantes que no practican deporte alguno (más del 80 % en ambos perfiles). - En ambos perfiles predominan los participantes que pasan más de 5 horas diarias frente al televisor, videojuego o PC (más del 90 % en ambos casos). | |

| Sub- área | Perfil sanitario | | |
|--------------|---|--|--|
| | Perfil No Saludable (n ₁ = 29) | Perfil Saludable (n ₂ = 33) | |
| OS | D | <ul style="list-style-type: none"> - Los participantes se identifican con figuras de dimensiones más robustas. - La mayoría de los participantes considera que tiene sobrepeso. - Alrededor de la mitad de los participantes considera que tiene obesidad. | <ul style="list-style-type: none"> - Los participantes se identifican con figuras de dimensiones más esbeltas. - La mayoría de los participantes no considera que tenga sobrepeso. - La mayoría de los participantes no considera que tenga obesidad. |
| | S | <ul style="list-style-type: none"> - En ambos perfiles más del 80 % de los participantes considera que está bajo de peso. - En ambos perfiles más de la mitad de los participantes considera que no está bien de peso (entre 50 % y 62 % de los participantes). - En ambos perfiles más del 65 % de los participantes masculinos tiene una cintura que mide más de 90 centímetros (lo saludable es <= a 90 centímetros de cintura en los hombres). - En ambos perfiles más del 75 % de los participantes femeninos tienen una cintura que mide más de 80 centímetros (lo saludable es <= a 80 centímetros en mujeres). | |
| DM | D | - Hay participantes que han recibido un diagnóstico médico de diabetes o azúcar en la sangre. | - Los participantes nunca han recibido un diagnóstico médico de diabetes o azúcar en la sangre. |
| | S | - En ambos perfiles más del 87 % de los participantes no sigue ningún tratamiento para controlar su peso. | |
| HA | D | - Hay participantes que han recibido un diagnóstico médico de presión alta. | - Los participantes nunca han recibido un diagnóstico médico de presión alta. |
| | S | - No hay similitudes en esta sub-área. | |
| EC | D | No hay similitudes en esta sub-área. | |
| | S | - En ambos perfiles más del 89 % de los participantes no ha recibido un diagnóstico médico sobre alguna afección cardíaca. | |
| ERHN | D | <ul style="list-style-type: none"> - Hay participantes que han recibido un diagnóstico médico de alguna enfermedad del riñón. - Hay participantes que han recibido un diagnóstico médico de colesterol alto. - Hay participantes que han recibido algún diagnóstico médico de triglicéridos altos. | <ul style="list-style-type: none"> - Los participantes nunca han recibido un diagnóstico médico de alguna enfermedad del riñón. - Los participantes nunca han recibido algún diagnóstico médico de colesterol alto. - Los participantes nunca han recibido algún diagnóstico médico de triglicéridos altos. |
| | S | - En ambos perfiles más del 96 % de los participantes nunca ha recibido un diagnóstico médico de embolia cerebral. | |

| Sub- área | Perfil sanitario | | |
|--------------|---|---|--|
| | Perfil No Saludable (n ₁ = 29) | Perfil Saludable (n ₂ = 33) | |
| C | D | - Hay participantes que han recibido algún diagnóstico médico de cáncer o de un tumor maligno de algún tipo. | - Los participantes nunca han recibido algún diagnóstico médico de cáncer o de un tumor maligno de algún tipo. |
| | S | - No hay similitudes en esta sub-área. | |
| A | D | - No hay diferencias en esta sub-área. | |
| | S | - En ambos perfiles más del 96 % de los participantes aseguró no haber sufrido daño a su salud a causa de un accidente en los 12 meses previos a ser encuestados. | |
| FR | D | No hay similitudes en esta sub-área. | |
| | S | <ul style="list-style-type: none"> - En ambos perfiles cerca del 50 % de los participantes reportó haber fumado tabaco alguna vez en su vida. - En ambos perfiles más del 84 % de los participantes reporta no fumar en la actualidad. - En ambos perfiles más del 65 % de los participantes asegura haber tomado alcohol alguna vez en su vida. - En ambos perfiles más del 69 % de los participantes reporta no tomar alcohol en la actualidad. | |
| MCPAC | D | <p><i>Movilidad:</i></p> <ul style="list-style-type: none"> - Alrededor de la mitad de los participantes se encuentra limitado por su salud para subir un solo piso por la escalera. - Alrededor de la mitad de los participantes se encuentra limitado por su salud para agacharse o arrodillarse. - Alrededor de la mitad de los participantes se encuentra limitado por su salud para caminar un kilómetro o más. <p><i>Cuidado personal:</i></p> <ul style="list-style-type: none"> - No hay diferencias en esta sub-área. <p><i>Actividades cotidianas:</i></p> <ul style="list-style-type: none"> - No hay diferencias en esta sub-área. | <p><i>Movilidad:</i></p> <ul style="list-style-type: none"> - Predominan ampliamente los participantes en los que su salud no es una limitante para subir un solo piso por la escalera. - Predominan ampliamente los participantes en los que su salud no es una limitante para agacharse o arrodillarse. - Predominan ampliamente los participantes en los que su salud no representa una limitante para caminar un kilómetro o más. <p><i>Cuidado personal:</i></p> <ul style="list-style-type: none"> - No hay diferencias en esta sub-área. <p><i>Actividades cotidianas:</i></p> <ul style="list-style-type: none"> - No hay diferencias en esta sub-área. |
| | S | <p><i>Movilidad:</i></p> <ul style="list-style-type: none"> - No hay similitudes en esta sub-área. <p><i>Cuidado personal:</i></p> <ul style="list-style-type: none"> - En ambos perfiles más de 96 % de los participantes reportó no estar limitado por su salud para bañarse o vestirse por sí mismos. | |

| Sub- área | Perfil sanitario | |
|---|---|--|
| | Perfil No Saludable (n ₁ = 29) | Perfil Saludable (n ₂ = 33) |
| MCPAC | S | <p><i>Actividades cotidianas:</i></p> <ul style="list-style-type: none"> - En ambos perfiles más del 89 % de los participantes reportó no haber tenido problemas para realizar sus actividades cotidianas el día en que fue encuestado. - En ambos perfiles más del 72 % de los participantes reportó que, durante las 4 semanas anteriores a la encuesta, no tuvo que reducir el tiempo a su trabajo o a sus actividades cotidianas a causa de su salud física. - En ambos perfiles más del 72 % de los participantes reportó que, durante las 4 semanas anteriores a la encuesta, no hizo menos de lo que hubiera querido hacer a causa de su salud física. |
| | D | <ul style="list-style-type: none"> - Alrededor de la mitad de los participantes califica su salud con un puntaje de entre 5 y 9: la mayoría se califica con salud regular o mala (cercaos al 1). - Alrededor de la mitad de los participantes califica su salud con un puntaje de entre 8 y 9: la mayoría se califica con buena salud (cercaos al 10). |
| PES | S | <ul style="list-style-type: none"> - En ambos perfiles alrededor de la mitad de los participantes reportó que, en comparación con los 12 meses anteriores a la encuesta, su salud (hasta el día de la encuesta) está más o menos igual, algo mejor o mucho mejor (mediana de 3 donde 3 significa más o menos igual). - En ambos perfiles la mayoría de los participantes se calificó en cuanto a qué tan enfermo considera estar, en escala de 0 a 10, con puntuaciones de 3 o menores. |
| | | |
| D: Diferencias S: Similitudes AF: Actividad física OS: Obesidad y sobrepeso DM: Diabetes mellitus HA: Hipertensión arterial EC: Enfermedad cardiovascular | | ERHN: Enfermedad renal, hipercolesterolemia y neuromuscular C: Cáncer A: Accidentes FR: Factores de riesgo MCPAC: Movilidad, cuidado prsonal y actividades cotidianas PES: Percepción del estado de salud |

Capítulo 2

El método de jueceo e índice de congruencia del ítem de Rovinelli y Hambleton, propuesto por Osterlind: aplicaciones en la validación del contenido de una escala de percepción de riesgo al conducir un automóvil¹

*Carro-Pérez, Ennio Héctor
Hernández-Gómez, Arturo Secundino*

Resumen

Se muestra el Índice de congruencia del ítem de Rovinelli y Hambleton, como un indicador apropiado de corte numérico para tasar la validez de contenido de un ítem y de un instrumento o escala en su conjunto de acuerdo al método expuesto por Osterlind (1997). Se ilustra cómo ejecutar el cálculo del índice y se describe un caso empírico donde se empleó para validar la Escala de Percepción de Riesgo al Conducir, mediante las evaluaciones hechas por una muestra de 50 jueces expertos en metodología e investigación científica. A través de este procedimiento se obtuvo una escala reducida de 11 ítems de los 26 originales, cuyos índices de Rovinelli y Hambleton o de Osterlind eran mayores a 0.4. El método constituye una alternativa robusta y sencilla de cuantificar la validez de contenido de un instrumento.

Palabras clave: validez de contenido, índice de Rovinelli y Hambleton, jueceo, desarrollo de medidas, percepción de riesgo al conducir.

Introducción

Con el propósito de determinar la validez de un instrumento cualquiera en las ciencias del comportamiento humano, se han definido varios procesos con los que se pueden obtener indicios de los diferentes tipos de validez referidos en la literatura sobre metodología general o psicometría (Kerlinger y Lee, 2002; Nunnally, 2013; Clark-Carter, 2002). Entre los tipos de validez más referidos se encuentran: a) de constructo, b) criterio, c) discriminación y d) de contenido (Cronbach y Meehl, 1955; Nadelsticher, 1983); sin excluir la patente o de *facie* (facial) (Clark-Carter, 2002) que hace referencia a la similitud de significado o percepción proporcionada por los usuarios ante ítems verbales o en otras palabras que los ítems o las mediciones

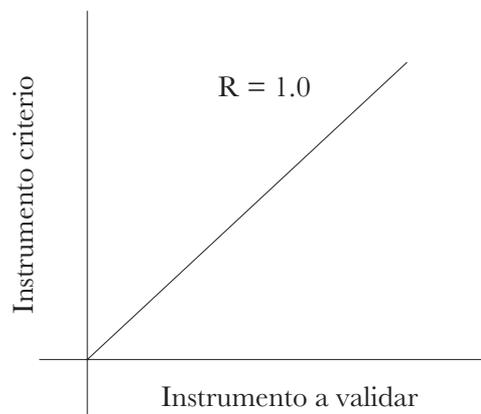
¹ Capítulo realizado gracias al apoyo otorgado por la Dirección General de Educación Superior Universitaria, PRODEP México, Apoyo de Fomento a la Generación y Aplicación Innovadora del Conocimiento, mediante oficio 511-6/17-8212 al proyecto *Percepción de riesgo y conductas de riesgo asociadas al conducir un automóvil en zonas urbanas del sur de Tamaulipas y ciudad de México*.

sean entendidas como deben ser entendidas por los usuarios, incluyendo a los administradores.

Los procesos referentes al cálculo u obtención de cada una de ellas tienen cierta divergencia de acuerdo al manual que se consulte o bien el nivel de medida de las variables, sin embargo, al parecer hay cierto acuerdo en que la validez de constructo, siempre y cuando el nivel de medida de las variables sea escalar, se puede obtener mediante un Análisis Factorial Confirmatorio, e inclusive Exploratorio (Arribas, 2004; Pérez-Gil, Chacón y Moreno, 2000), encontrando divergencias quizá en los tipos de rotación a emplear, el método de extracción o la sustentación de la construcción factorial ya por matrices de correlaciones o de covariación.

En cuanto a la validez de criterio, el proceso establecido requiere un criterio o referente de comparación respecto al instrumento de medida, por tal motivo, en niveles de medición escalar, los criterios suelen ser instrumentos que miden la misma variable y en condiciones similares, o bien medidas que teóricamente pueden encontrarse relacionadas con la o las variables del instrumento que se quiere validar. En este sentido los indicadores matemáticos de la validez suelen tener como base coeficientes de correlación como el R de Pearson o equivalentes de acuerdo al nivel de medida de las variables (Arribas, 2004; Cohen y Swerdlik, 2001). En los que puntajes cercanos a 1, reflejarían un alto valor de relación con el criterio empleado, mientras que un valor cercano a cero indicaría una escasa relación, y en consecuencia una validez de criterio baja.

Figura 1. Relación del cien por ciento ($R=1.0$) entre instrumento y criterio

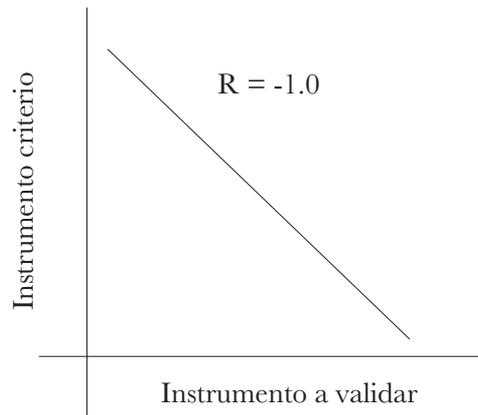


En términos de los valores que pueden producir este tipo de coeficientes se espera que puntajes con valor positivo se den en condiciones de comparación con

instrumentos que miden la misma variable u otra con la que se tiene una correlación directamente proporcional, mientras que los valores negativos con variables que se encuentren asociados de manera inversamente proporcional. Así, un instrumento sobre actitud hacia el aborto (instrumento a validar), debería correlacionar de forma positiva con otro que mida exactamente lo mismo (criterio), con lo que se podría describir una gráfica como la siguiente (ver Figura 1).

Si por el contrario el criterio fuera un instrumento cuya variable a través de la evidencia empírica y teórica correlaciona de manera negativa con la medida a validar se esperaría una gráfica distinta a la anterior (ver Figura 2). En nuestro caso, se podría considerar un instrumento sobre intolerancia al aborto, de tal manera que respondientes que puntuaran altos en intolerancia, tuvieran una puntuación baja o una actitud desfavorable en el instrumento de actitud hacia el aborto y viceversa.

Figura 2. Relación negativa o inversamente proporcional del cien por ciento ($R = -1.0$) entre instrumento y criterio



En lo relativo a la validez de discriminación, Nadelsticher (1983) propone un procedimiento basado en el estadístico T de Student para muestras o grupos independientes, donde las medias de los grupos se sustituyen por las medias del 25 % de puntuaciones más altas y el 25 % más bajas dadas para cada ítem de un instrumento. De tal manera la fórmula original para t se adaptaría a t_i :

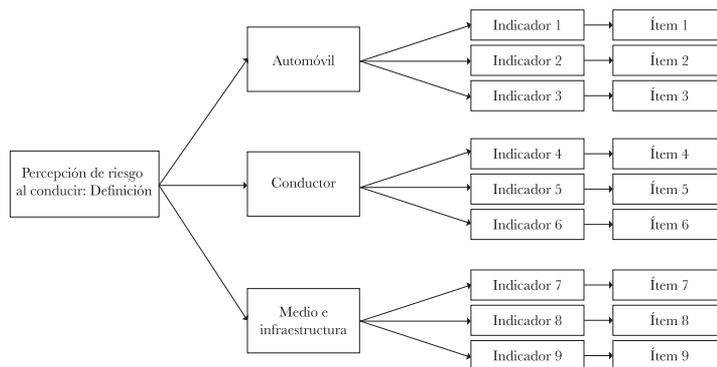
$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad \bullet \bullet \quad t_i = \frac{\bar{x}_a - \bar{x}_b}{\sqrt{\frac{\sigma_a^2}{n_a} + \frac{\sigma_b^2}{n_b}}}$$

Donde $\bar{X}a$ es el promedio del 25 % de puntuaciones más altas dadas a un ítem, $\bar{X}b$ el promedio del 25 % de las puntuaciones más bajas dadas al ítem en cuestión, σ^2 es la varianza tanto para el grupo de más altos puntajes (σ^2a) y el grupo de más bajos puntajes (σ^2b), na y nb la cantidad de puntuaciones en el grupo a y el grupo b respectivamente.

El anterior procedimiento de contraste entre puntajes altos y bajos deberá practicarse para cada ítem del instrumento, de tal manera, sí el instrumento de medida tiene 10 ítems deberán ejecutarse 10 cálculos a través de T de Student, el resultado de la T, de acuerdo con Nadelsticher (1983) deberá ser mayor o igual a 1.75 para que el ítem sea discriminativo, de tal manera todos aquellos ítems con puntajes T menores a 1.75 no son discriminantes y deberán ser eliminados del instrumento de medida, una alternativa a este criterio es considerar la significación asociada al valor de T, en este sentido una significación menor o igual a .05 indicaría que el ítem es discriminante y una mayor a .05 lo contrario.

En cuanto a la validez de contenido se describe como proceso normal el desarrollar los ítems a partir de la definición conceptual de la variable medida por el instrumento, de tal manera que cada uno de los ítems contenga alguno de los aspectos o cualidades que se mencionan en la definición, en este sentido cada ítem representa una porción de la variable (Clark-Carter, 2002; Osterlind, 1997), por ejemplo sí una variable como la *Percepción de riesgo al conducir un automóvil*, por definición, tiene tres cualidades o dimensiones de riesgo: el automóvil, las conductas de riesgo del conductor, y el medio donde se conduce. Un instrumento que la mida deberá incluir ítems para cada dimensión (ver Figura 3).

Figura 3. Modelo de desarrollo de ítems a partir de la definición conceptual de la variable



En la imagen anterior se describe el proceso que se debe seguir en el desarrollo de nueve ítems a partir de la definición de Percepción de riesgo al conducir, que consiste

en la evaluación que emite el conductor sobre la posibilidad de tener un accidente de tránsito por aspectos relacionados con el automóvil, a las conductas que realiza mientras conduce, y el medio circundante a la vía y el estado de la misma. También se puede apreciar un paso intermedio entre la definición de la variable, sus dimensiones y los ítems, el establecimiento de indicadores empíricos que determinarán el contenido de estos últimos (Escobar-Pérez y Cuervo-Martínez, 2008).

Tradicionalmente este proceso podría agotarse aquí, particularmente después de haber confeccionado una gran cantidad de ítems por dimensión y no sólo tres como se ha mostrado en el ejemplo anterior, sobre todo si se considera la existencia de una relación entre la cantidad de ítems y la confiabilidad del instrumento (Nadelsticher, 1983). Sin embargo, aparte de realizar una suficiente cantidad de ítems por dimensión para que la totalidad del concepto se encuentre reflejado en el instrumento cual espejo, siempre existen dudas al respecto: ¿los ítems son efectivamente representativos de la dimensión de la variable?, ¿los respondientes o sujetos responderán a ellos de acuerdo al contenido del mismo ítem?, ¿los ítems son útiles para medir la variable o una dimensión de esta?

Para responder estas preguntas, se puede recurrir a un método de jueceo y cálculo de la congruencia o consistencia de las evaluaciones que los jueces emiten hacia cada ítem del instrumento, proceso propuesto por Osterlind (1997), que parte del Índice de Congruencia del Ítem de Rovinelli y Hambleton (1977), que proporciona un valor entre 1 y -1, que puede interpretarse como un indicador escalar de validez de contenido del ítem y en su conjunto de la totalidad de un instrumento, donde 1, será el máximo puntaje de congruencia con una dimensión o variable, y el -1 la incongruencia o desajuste máximo del ítem con una dimensión o aspecto.

Método de Osterlind para obtener un indicador numérico de la validez de contenido del ítem

Para realizar el procedimiento de validación de contenido de los ítems de un instrumento o escala a través del método por Osterlind, como primer paso, debemos tener el conjunto de ítems de la escala organizados de tal manera que los jueces puedan evaluar su congruencia en las dimensiones o aspectos que se requieran, Sanduvete-Chaves (2008), empleó para este proceso tres áreas que se consideran apropiadas para esta labor: la *representatividad* o qué tan representativo (R) es el ítem de la variable a medir; la *utilidad* (U) del ítem para medir a la variable o dimensión de esta; y por último la posibilidad de que la respuesta dada al ítem realmente corresponda al aspecto medido o *factibilidad* (F), que también puede nombrarse como *viabilidad* (V). Una forma de ordenar los ítems es mediante un tabulado impreso o

electrónico en donde los ítems se encontrarán en los renglones y en columnas las tres dimensiones señaladas (ver Tabla 1).

Tabla 1. Organización de los ítems para ser evaluados por jueces

| | | | |
|--|---|---|---|
| Instrucciones. A continuación se presentan un conjunto de ítems, los cuales deberán ser valorados por usted en cuanto a su representatividad (R), utilidad (U), y viabilidad (V), para medir la variable X. La cual debe entenderse como... Para lo anterior podrá emplear una puntuación del 1, 2 o 3. Donde el 1 será el menor valor de R, U o V, y 3 el mayor. | | | |
| Items | R | U | V |
| Ítem 1... | | | |
| Ítem n... | | | |

Fuente: Elaboración propia.

En la tabla anterior se puede observar un formato de instrucciones para los jueces, este puede ser diferente dependiendo del estilo de redacción del investigador y sus necesidades, lo que no se puede omitir es proporcionar el significado de la variable a los jueces, ya que será el referente a partir del cual ellos valoraran qué tan representativos, útiles y viables son los ítems, además se debe indicar la puntuación que pueden emplear para realizar la evaluación, en el ejemplo anterior se ha usado un intervalo de 1 a 3, sin embargo, Osterlind (1997), propone el intervalo de -1 a 1, pasando por 0. Se puede emplear un intervalo u otro siempre y cuando se ajuste la fórmula de congruencia del ítem que se mostrará más adelante. En el presente trabajo se ha preferido emplear las puntuaciones de 1, 2 y 3, por ser más habitual para los individuos emplear números positivos que negativos.

Como segundo paso se debe contar con una muestra de jueces, ésta debe estar integrada por lo menos con 30 jueces expertos en la temática o área de medición para cumplir con criterios de normalidad estadística (Clark-Carter, 2002), sin embargo, puede ocurrir que exista una escasez de jueces expertos en un tópico de medida, de tal manera se pueden emplear jueces especializados en metodología de la investigación científica o bien sólo emplear la cantidad de jueces que existan o respondan de manera voluntaria a una convocatoria, ésta puede realizarse por invitación personal, telefónica o por correo electrónico. Respecto a la cantidad de jueces, Osterlind (1997) no es concluyente, él ilustra el uso del índice con una cantidad menor a 10 jueces, mientras en un estudio de Vila, Hijós y Holgado (2013) se emplearon 12 jueces, y en otro de Sanduvete-Chaves, Chacón-Moscoso, Sánchez-Martín y Pérez-Gil (2013) se utilizaron 105 jueces.

A los jueces que hayan confirmado su participación se les debe proporcionar los ítems organizados para su evaluación, tal como se ha descrito en la Tabla 1. A partir de los puntajes otorgados por los jueces a los ítems se realizará el tercer paso, el cual consiste en calcular los índices de congruencia o índice de Osterlind (IO), para cada ítem en cada una de las dimensiones R, U y F o V. De tal manera, un conjunto de 10 ítems, tendrán un total de 30 IO, es decir tres por ítem.

La obtención del IO se realizará mediante el empleo de la siguiente fórmula cuando se use el intervalo originalmente propuesto de -1, 0, y 1 (Sanduvete-Chaves, Chacón-Moscoso, Sánchez-Martín y Pérez-Gil, 2013).

$$I_{ik} = \frac{(N-1)\sum_{j=1}^n X_{ijk} + N\sum_{j=1}^n X_{ijk} - \sum_{j=1}^n X_{ijk}}{2(N-1)n}$$

Donde:

- N = número de dominios.
- X_{ijk} = el valor que cada juez j, otorga a un ítem k, en la dimensión o dominio i
- n = número de jueces.

Sí se utilizan las puntuaciones 1, 2, y 3, para evaluar a los ítems, se propone un ajuste de la fórmula para quedar como:

$$I_{ik} = \left[\frac{(N-1)\sum_{j=1}^n X_{ijk} + N\sum_{j=1}^n X_{ijk} - \sum_{j=1}^n X_{ijk}}{2(N-1)n} \right] - 2$$

Para ilustrar el cálculo de un IO para un aspecto o dominio como R o cualquier otro, consideremos que un conjunto de 10 jueces proporcionaron las siguientes puntuaciones al primer ítem de una escala, utilizando el intervalo de 1 a 3:

$$R: \{2, 1, 2, 3, 3, 2, 3, 3, 3, 2\}$$

Empleando la fórmula ajustada del IO, deberemos obtener el término $\sum_{j=1}^n X_{ijk}$, el cual es una sumatoria de las puntuaciones otorgadas por los jueces al aspecto R, así el resultado de la suma de estas puntuaciones es:

$$\sum_{j=1}^n X_{ijk} = 2 + 1 + 2 + 3 + 3 + 2 + 3 + 3 + 3 + 2 = 24$$

Considerando en la fórmula el número de dominios o N, igual a 3, por ser esta cantidad los aspectos a evaluar para cada ítem (R, U, F o V), sustituyendo se tiene:

$$I_{ik} = \left[\frac{(3 - 1) 24 + 3(24) - 24}{2(3 - 1)10} \right] - 2$$

Así,

$$I_{ik} = \left[\frac{48 + 72 - 24}{40} \right] - 2 = [2.4] - 2 = 0.4$$

Siguiendo lo establecido por Sanduvete-Chaves (2008), Sanduvete-Chaves, Chacón-Moscoso, Sánchez-Martín y Pérez-Gil (2013), las puntuaciones del IO iguales o mayores a 0.5 se consideran adecuadas, en el ejemplo propuesto el IO fue de 0.4, de tal manera la R de este ítem no sería apropiada.

El anterior procedimiento se deberá repetir para obtener los IO para la R, U, y V de cada ítem de la escala o instrumento, lo deseable es que cada ítem tenga IO iguales o superiores a 0.5 en cada elemento o dimensión de la validez de contenido (R, U, V), sin embargo si una de las dimensiones o dominios no cumple con este criterio, se sugiere como alternativa calcular el promedio de los mismos, esperando que este sea igual o superior al 0.5 para considerar al ítem como válido.

La media será producto de:

$$\bar{IO} = \sum IO_{RUV}$$

Finalmente, es importante realizar una observación a la fórmula propuesta por Osterlind, el producto de la misma, a reserva de probarlo con mayor amplitud, es igual a la media de las puntuaciones otorgadas por los jueces al ítem k para la dimensión o dominio i . Así, considerando las cifras de nuestro ejemplo para el dominio R, y la fórmula usada:

$$\bar{X}_{ijk} = \frac{\sum_{j=1}^n X_{ijk}}{n} = \frac{2+1+2+3+3+2+3+3+3+2}{10} = 24/10 = 2.4 - 2 = 0.4$$

Como se puede advertir, al resultado final se le han restado dos puntos y es idéntico al obtenido a través del IO. Esta resta de dos puntos, sólo operaría cuando se utilice el intervalo de evaluación de 1 a 3 y no de -1 a 1.

Como una alternativa al IO, Turner, Mulvenon, Thomas y Balkin (2002), y Turner y Carlson (2003), presentan otra versión del Índice de Congruencia del Ítem

propuesto por Rovinelli y Hambleton (1977), la cual comparte varios términos con la propuesta de Osterlind:

$$I_{ik} = \frac{(N-1) \sum_{j=1}^n X_{ijk} - \sum_{j=1}^n X_{ijk} + \sum_{j=1}^n X_{ijk}}{2(N-1)n}$$

La principal diferencia con la fórmula de Osterlind, está en el numerador, donde se usa el término $\sum_{i=1}^N$. Para Turner, Mulvenon, Thomas y Balkin (2002), $\sum_{i=1}^N$, es la suma de los promedios de las calificaciones que los jueces han dado al ítem en todos los objetivos o dimensiones evaluables, en los que podría ser congruente o no el ítem, considerando que sólo en uno podría ser altamente congruente. Todos los demás símbolos representan lo mismo que en la fórmula de Osterlind, y el resultado también oscila entre -1 a 1. Para mayor información sobre el proceso de aplicación de este índice se remite al lector al texto de Turner y colaboradores (2002).

El proceso para obtener la validez de contenido a través del procedimiento de Osterlind, está conformado por los siguientes pasos:

1. Organización de los ítems para su evaluación por parte de los jueces.
2. Convocar una muestra de jueces expertos en la variable medida y evaluación de los ítems por parte de estos.
3. Cálculo del índice de congruencia del ítem o de Osterlind (IO).
4. Identificar los ítems con un IO igual o mayor a 0.5 en cada dimensión o dominio (R, U, V o F), o con una validez de contenido adecuada.

Aplicación del IO en la validación de una Escala de Percepción del Riesgo al Conducir un Automóvil

Dentro del marco de un proyecto de investigación para determinar la percepción de riesgo y conductas de riesgo al conducir un automóvil por parte de automovilistas, en la Ciudad de México y la zona conurbada del sur de Tamaulipas, se empleó el procedimiento anteriormente descrito, en la validación de una escala para medir la percepción de riesgo a tener un accidente de tránsito al conducir un automóvil.

La Escala de Percepción de Riesgo al Conducir un Automóvil (EPRC), originalmente estuvo conformada por 26 ítems relacionados con aspectos de seguridad del vehículo, conductas de riesgo del conductor, y características del medio por donde circula el vehículo.

En nuestro caso se solicitó a 50 jueces, profesores-investigadores, convocados por correo electrónico que evaluaran los ítems elaborados para la escala de percepción de riesgo mediante una plantilla electrónica en internet de la aplicación de Google Drive ®, entre los meses de marzo a septiembre de 2018. Los valores

usados para la evaluación de los ítems en las dimensiones R, U y V, por parte de los jueces fueron del 1 al 3, de tal manera, para ser congruentes con el planteamiento original de Osterlind, se restó a los valores obtenidos mediante el Índice de Osterlind (IO) dos puntos, o bien se empleó la fórmula anteriormente mostrada:

$$I_{ik} = \left[\frac{(N-1) \sum_{j=1}^n X_{ijk} + N \sum_{j=1}^n X_{ijk} - \sum_{j=1}^n X_{ijk}}{2(N-1)n} \right] - 2$$

Los resultados obtenidos para cada dimensión se muestran a continuación, así como los promedios de IO. Se han resaltado los ítems que muestran valores promedio de Osterlind mayores o iguales a 0.4, valor que se sugirió para este estudio aunque Sanduvete-Chaves, Chacón-Moscoso, Sánchez-Martín y Pérez-Gil (2013) sugieren .5, en virtud de que al menos se consideró que alguna de las dimensiones R, U o V, tuvieran o superaran el .5, los ítems que cumplen con esta condición se han marcado con un asterisco (*) (ver Tabla 2). En este sentido Osterlind (1997), señala que el valor criterio para considerar un ítem válido o no puede ser mayor a .5 dependiendo del investigador y jueces.

Tabla 2. Valores de IO para cada dimensión y promedio de IO por ítem de la escala

| Items | IO _R | IO _U | IO _V | (\bar{IO}) | ≥0.4 |
|---|-----------------|-----------------|-----------------|----------------|------|
| 1. Velocidad a la que se conduce el automóvil | 0.6 | 0.34 | 0.48 | 0.47333333 | Sí* |
| 2. Usar el celular mientras se conduce | 0.74 | 0.66 | 0.62 | 0.67333333 | Sí |
| 3. Conducir sin usar el cinturón de seguridad | 0.5 | 0.44 | 0.4 | 0.44666667 | Sí* |
| 4. La distancia que se tiene con el automóvil de adelante | 0.44 | 0.4 | 0.46 | 0.43333333 | No |
| 5. Conducir con las dos manos al volante | -0.18 | -0.1 | -0.12 | -0.13333333 | No |
| 6. Conducir cuando se ha consumido alcohol | 0.74 | 0.76 | 0.78 | 0.76 | Sí |
| 7. Conducir sin usar las señales de luz del automóvil | 0.44 | 0.56 | 0.42 | 0.47333333 | Sí* |
| 8. Conducir sin acatar las señales de tránsito | 0.68 | 0.66 | 0.64 | 0.66 | Sí |
| 9. Las competencias o habilidades del conductor | 0.36 | 0.44 | 0.38 | 0.39333333 | No |
| 10. El funcionamiento general del automóvil | 0.38 | 0.4 | 0.18 | 0.32 | No |
| 11. Qué tan probable es que pueda sufrir un accidente de tránsito | -0.16 | -0.18 | -0.22 | -0.18666667 | No |

| Items | IO _R | IO _U | IO _V | (\bar{IO}) | ≥ 0.4 |
|---|-----------------|-----------------|-----------------|----------------|------------|
| 12. El funcionamiento de las luces de señalización del automóvil | 0.34 | 0.46 | 0.26 | 0.35333333 | No |
| 13. El estado de los frenos | 0.56 | 0.76 | 0.58 | 0.63333333 | Si |
| 14. El desgaste de los neumáticos | 0.38 | 0.46 | 0.32 | 0.38666667 | No |
| 15. La presión del aire de los neumáticos | 0.24 | 0.32 | 0.18 | 0.24666667 | No |
| 16. El estado de los cinturones de seguridad | 0.04 | 0.24 | 0.16 | 0.14666667 | No |
| 17. Los sistemas de visibilidad del automóvil (cristales y espejos) | 0.5 | 0.54 | 0.32 | 0.45333333 | Si* |
| 18. Los fluidos del motor | 0.22 | 0.38 | 0.16 | 0.25333333 | No |
| 19. El intervalo en que se da mantenimiento al automóvil | 0.24 | 0.22 | 0.1 | 0.18666667 | No |
| 20. El encargo de verificación o dar mantenimiento al automóvil | 0.14 | 0.04 | -0.08 | 0.03333333 | No |
| 21. El mantenimiento del automóvil | 0.28 | 0.24 | 0.16 | 0.22666667 | No |
| 22. La carretera, calles o caminos por donde se transita | 0.32 | 0.36 | 0.36 | 0.34666667 | No |
| 23. Los otros automovilistas | 0.48 | 0.38 | 0.34 | 0.4 | No |
| 24. Los peatones | 0.22 | 0.18 | 0.12 | 0.17333333 | No |
| 25. Las señales de tránsito | 0.4 | 0.42 | 0.38 | 0.4 | No |
| 26. De acuerdo a cómo conduzco actualmente un automóvil, el riesgo que tengo de sufrir un accidente de tránsito es: | 0.04 | -0.02 | -0.04 | -0.00666667 | No |

Fuente: Elaboración propia.

Los valores de IO adecuados (≥ 0.4) sugieren que el ítem proporcionará mediciones válidas al menos en el campo del contenido o representación de la variable estudiada.

En función de lo anterior la EPRC quedó validada en términos de contenido, y redujo su tamaño de 26 ítems a sólo ocho ítems (ver Tabla 3).

Tabla 3. Ítems que integraron la EPRC final

| Items | IO _R | IO _U | IO _V | (\bar{IO}) | ≥ 0.4 |
|---|-----------------|-----------------|-----------------|----------------|------------|
| 1. Velocidad a la que se conduce el automóvil | 0.6 | 0.34 | 0.48 | 0.47333333 | Si* |
| 2. Usar el celular mientras se conduce | 0.74 | 0.66 | 0.62 | 0.67333333 | Si |
| 3. Conducir sin usar el cinturón de seguridad | 0.5 | 0.44 | 0.4 | 0.44666667 | Si* |

| Items | IO _R | IO _U | IO _V | (\bar{IO}) | ≥ 0.4 |
|--|-----------------|-----------------|-----------------|----------------|------------|
| 4. Conducir cuando se ha consumido alcohol | 0.74 | 0.76 | 0.78 | 0.76 | Sí |
| 5. Conducir sin usar las señales de luz del automóvil | 0.44 | 0.56 | 0.42 | 0.47333333 | Sí* |
| 6. Conducir sin acatar las señales de tránsito | 0.68 | 0.66 | 0.64 | 0.66 | Sí |
| 7. El estado de los frenos | 0.56 | 0.76 | 0.58 | 0.63333333 | Sí |
| 8. Los sistemas de visibilidad del automóvil (cristales y espejos) | 0.5 | 0.54 | 0.32 | 0.45333333 | Sí* |

Fuente: Elaboración propia.

Como se puede advertir en la anterior tabla, los ítems con mejor IO, se pueden agrupar en los factores que tradicionalmente se encuentran asociados con los accidentes de tránsito: las conductas de riesgo que realiza el conductor (ítems 1 al 6), que van desde el uso del cinturón de seguridad, distractores, la velocidad o la falta de seguimiento de las reglas de tránsito; y aspectos relativos al automóvil, como el sistema de frenado (ítems 7 y 8).

Conclusiones

La validez de un instrumento de medida es fundamental en el proceso de la medición, sin evidencia de la misma no se puede contar con la certidumbre apropiada de que las mediciones que se efectúan en un proyecto de investigación, de intervención o bien en una situación cualesquiera sean correctas. Se debe considerar que la validez, por definición, es la condición de que el instrumento mida lo que debe medir, en otras palabras sea congruente con la variable para la cual fue confeccionado.

Hay distintos tipos de validez, un instrumento apropiado debe contar con indicadores de cada una de ellas, por lo que se puede considerar a ésta como una sumatoria, así su total es el producto de la suma de los subtipos constructo, criterio, discriminación y contenido.

Sí bien cada validez por separado se obtiene de diferente manera, ésta se podría ponderar o convertir a otro tipo de unidades homologables, por ejemplo aquellas que se obtienen a partir de coeficientes de correlación pueden leerse como porcentajes, e inclusive la validez de constructo obtenida por análisis factorial, puede ser traducida aún promedio de porcentajes a través de las cargas factoriales o bien el total de la varianza explicada de los factores extraídos. En cuanto a la validez de contenido, el proceso visto en el presente capítulo es una alternativa fácil y robusta para cuantificarla a partir de las evaluaciones de jueces en tres áreas: representatividad del ítem respecto a la variable, utilidad del ítem para medir la variable, y la viabilidad

de obtener una respuesta no sesgada de parte del respondiente. Sin embargo, puede emplearse para medir la congruencia del ítem con cualquier dimensión o componente de una variable y no sólo las observadas aquí.

También es una herramienta versátil, ya que la técnica puede ser modificada en lo relativo a la escala con la que evalúan los jueces, por ejemplo, de la propuesta original de -1 a 1 puede elegirse de 1 a 3, inclusive ampliarse el rango a 5 puntos (Sanduvete-Chaves, Chacón-Moscoso, Sánchez-Martín y Pérez-Gil, 2013), lo cual sólo llevaría a realizar ajustes a la fórmula. Así mismo, pueden reducirse las dimensiones de evaluación de los ítems a dos, en caso que sólo se desee considerar la utilidad y representatividad de las preguntas o afirmaciones integrantes de un cuestionario o escalamiento.

La técnica puede ser empleada para valorar cualquier tipo de ítem y no sólo aquellos confeccionados para medir un aspecto de una variable de manera escalar, de tal manera el nivel de medida de los ítems no es un impedimento, se puede calcular el IO de aquellos que son ordinales o nominales, o con distintas formas de presentación a la escrita, como ítems gráficos.

Como punto pendiente, sólo quedaría por confirmar en diferentes circunstancias, que la técnica no se reduzca a una simple media de las puntuaciones, en este caso, sería más útil el cálculo directo de la media y una medida de dispersión como la desviación estándar, que resolver un algoritmo más complejo.

Por lo anterior la técnica se aconseja en cualquier tipo de estudio de carácter experimental o no experimental, al margen de cualquier diseño de investigación de corte cuantitativo o cualitativo.

Referencias

- Arribas, M. (2004). "Diseño y validación de cuestionarios". *Matronas Profesión*. 5 (17). pp. 23-29. Recuperado el 20 de mayo de 2020, de http://evidencia.com/wp-content/uploads/2014/07/validacion_cuestionarios.pdf
- Clark-Carter, D. (2002). *Investigación cuantitativa en psicología*. México: Oxford.
- Cohen, R. J. y Swerdlik, M. E. (2001). *Pruebas y evaluación psicológicas. Introducción a las pruebas y a la medición* [4ª. Edición]. México, D.F.: McGraw-Gill.
- Cronbach, L y Meehl, P. (1955). "Construct validity in Psychological Test". *Psychological Bulletin*, LII, pp. 281-302. En Wainerman, Catalina H. (1976). *Escalas de medición en ciencias sociales*. Ediciones Nueva Visión: Buenos Aires.
- Escobar-Pérez, J. y Cuervo-Martínez, A. (2008). "Validez de contenido y juicio de expertos: una aproximación a su utilización". *Avances en Medición*, Vol. 6 (1), pp. 27-36. https://www.researchgate.net/profile/Jazmine_EscobarPerez/publication/302438451_Validez_de_contenido_y_juicio_de_expertos_Una_aproximacion_a_su

- utilizacion/links/59a8daecaca27202ed5f593a/Validez-de-contenido-y-juicio-de-expertos-Una-aproximacion-a-su-utilizacion.pdf
- Indices of Item Congruence for Test Development Validity Assessments. *SAS Conference Proceedings: SAS Users Group International 27*. 14-17 April 2002, Orlando, Florida, USA, 1-5.
- Kerlinger, F. N. y Lee, H. B. (2002). *Investigación del Comportamiento*. México, D.F.: McGraw-Gill
- Nadelsticher M., A. (1983). *Técnicas para la construcción de cuestionarios de actitudes y opción múltiple*. México: Instituto Nacional de Ciencias Penales.
- Nunnally, J.C. (2013). *Teoría psicométrica* [reimpresión]. México: Trillas.
- Osterlind, S. J. (1997). *Constructing Test Items: Multiple-Choice, Constructed-Response, Performance and Other Formats*. Boston, United States: Kluwer Academic Publishers.
- Pérez-Gil, J. A.; Chacón M., S. y Moreno R., R. (2000). “Validez de constructo: el uso de análisis factorial exploratorio-confirmatorio para obtener evidencias de validez”. *Psicothema*, Vol. 12 (Suplemento 2), pp. 442 - 446.
- Rovinelli, R. J. y Hambleton, R. K. (1977). “On the use of content specialists in the assessment of criterion-referenced test item validity”. *Dutch Journal of Educational Research*, 2, 49-60.
- Sanduvete-Chaves, S. (2008). “Innovaciones metodológicas en la evaluación de la formación continua”. [Tesis Doctoral Inédita]. Universidad de Sevilla, Sevilla. Recuperado el 14 de diciembre de 2020, en https://idus.us.es/bitstream/handle/11441/15496/Y_TD_PS-295.pdf?sequence=1&isAllowed=y
- Sanduvete-Chaves, S., Chacón-Moscoso, S., Sánchez-Martín, M. y Pérez-Gil, J. A. (2013). “The Revised Osterlind Index. A Comparative Analysis in Content Validity Studies”. *Acción Psicológica*. 10 (2), pp. 19-26.
- Turner, R. C. y Carlson, L. (2003). “Indexes of Item-Objective Congruence for Multidimensional Items”. *International Journal of Testing*, 3 (2), 163-171.
- Turner, R. C., Mulvenon, S. W., Thomas, S. P. y Balkin, R. S. (2002). *Computing*
- Vila A., E., Hijós L., L. A., y Holgado T., F. P. (2013). “Adaptación y validez de contenido del cuestionario de envidia infantil a población adulta”. *Acción Psicológica*, 10(2), 59-72. <https://dx.doi.org/10.5944/ap.10.2.11824>

Capítulo 3

Propuesta metodológica para la investigación psicosocial de la calidad de vida en contextos comunitarios

*Carro-Pérez, Ennio Héctor
Hernández-Gómez, Arturo Secundino*

Resumen

La calidad de vida es una variable compleja que requiere de múltiples fuentes de información para su evaluación, a menudo se encuentra asociada o se subsume a otros términos como desarrollo social, desarrollo humano, bienestar, entre otros. Considerando lo anterior se presenta una propuesta metodológica para abordar el estudio de la calidad de vida en contextos comunitarios, la cual integra herramientas de las tradiciones cualitativas y cuantitativas en investigación. La ejecución de la misma, en su fase operativa se da en tres momentos después de la planificación inherente a todo proyecto, el primer momento involucra el contacto inicial con el entorno comunitario, ubicación de espacios representativos e informantes privilegiados; el segundo momento profundiza en la calidad de vida a través de entrevistas no estructuradas a los residentes; el cierre se da en el tercer momento mediante el levantamiento de registros objetivos sobre la calidad de vida.

Palabras clave: calidad de vida, metodología básica, investigación científica, alternativas metodológicas, contexto comunitario.

Introducción

Antes de iniciar con la exposición de la presente propuesta es importante señalar que en la literatura científica, las capacidades que tienen los grupos y los individuos para solventar sus necesidades en general han sido enmarcadas bajo distintos términos como desarrollo social, bienestar personal, desarrollo comunitario, o calidad de vida, en nuestra aproximación hemos preferido sintetizar a estas capacidades bajo el término de calidad de vida, sin omitir sus múltiples relaciones con otros conceptos, particularmente con el desarrollo y bienestar social o comunitario. La elección del término se da especialmente por la posibilidad de trabajar con la calidad de vida desde una aproximación psicosocial, una vez que en ella se pueden distinguir tanto componentes de índole material u objetivos, como subjetivos y socio cognitivos, producto de la percepción interpersonal de la realidad compartida por los individuos miembros de una comunidad (Arita, 2005; Ardila, 2003).

Antecedentes de desarrollo social

El desarrollo social puede ser vislumbrado como la capacidad que tiene un conjunto de personas para satisfacer sus necesidades o en otras palabras para tener un nivel satisfactorio de vida, en nuestro caso el desarrollo social se puede verificar de manera inmediata en una comunidad, entendida ésta como un espacio de interacción de individuos y elementos medio ambientales o materiales, que les permiten generar conjuntos de creencias, significados, formas particulares de comunicación, prácticas y necesidades compartidas. Entender a la comunidad de esta manera nos permitirá hablar entonces de desarrollo comunitario y de un término aparejado: calidad de vida (Arita, 2005).

El desarrollo de las comunidades se encuentra relacionado con un amplio conglomerado de factores, los cuales se pueden clasificar en categorías de acuerdo con la fuente de las que emanan, así al realizar un ejercicio taxonómico a partir de una lógica deductiva, podemos encontrar dos grandes grupos: aquellos factores de procedencia externa a la comunidad, y los factores internos a ella. Dentro de los factores externos podemos distinguir a los que derivan de la acción de instituciones de gobierno, tanto de índole federal como estatal, encontrándose en este último rubro las municipales, a parte de las instituciones anteriormente referidas, se encontraran todas aquellas agrupaciones que por sus características se constituyen en instituciones u organizaciones independientes o privadas (ONG u otras) cuya actividad también afecta a la comunidad, de tal manera se pueden encontrar los prestadores de servicios, comercios, iglesias, escuelas, entre otros.

Dentro de los agentes o factores internos debemos establecer que de inicio son todas las acciones que los integrantes de la comunidad, las personas, efectúan, de tal manera, que el desarrollo comunitario se debe también a lo que realizan los elementos sustanciales y naturales de la misma. En este tipo de factores podemos distinguir entre aquellas acciones estrictamente individuales y las de corte interindividual que se materializan en ejecuciones que realizan las personas como parte de un grupo destinadas a la modificación del entorno para su mejoría o mantenimiento, dentro de estos grupos se distinguen primordialmente a las familias, grupos de acción o ayuda, y las asociaciones políticas formales.

La determinación de los aspectos que permiten el desarrollo comunitario, nos permite vislumbrar que la comunidad es un todo complejo, multivariado, donde su evolución ha sido abordada desde diversas ópticas o discursos disciplinares que han permitido, en su momento, comprender aspectos particulares de las comunidades, sin embargo no han contribuido a tener una visión holística o integrada de los imaginarios comunitarios, sobre todo que contemple los aspectos individuales e interindividuales de los mismos.

Para ejemplificar lo anterior es necesario recurrir, a los aspectos tradicionales del abordaje del desarrollo comunitario, en particular del bienestar social y calidad de vida de las comunidades, así desde una visión económico-social el desarrollo de las comunidades en términos de bienestar y calidad de vida ha estado asociado a la capacidad que tienen los individuos integrantes de una comunidad de satisfacer necesidades básicas para la vida, como el poseer una vivienda, alimentación, o acceso a la educación, de tal manera, el bienestar se ha observado, particularmente, desde sus elementos exclusivamente materiales (Ardila, 2003). Lo anterior ha propiciado que el estudio del mismo se efectúe como una simple suma de condiciones materiales, como la cantidad de habitaciones, focos, baños que tiene la vivienda o la percepción salarial, tratándose muy poco los aspectos perceptuales o de valoración subjetiva de los mismos por parte de las personas. De tal manera un indicador del bienestar social usado por organismos como la Organización de las Naciones Unidas (ONU) (Prieto, 1982), como lo es el Índice de Desarrollo Humano (IDH), valora aspectos como el acceso a la educación, a sistemas de salud adecuados y una remuneración por arriba del nivel de pobreza. Este bienestar social, también puede verificarse a través de la calidad de vida de los integrantes de una comunidad en el sentido más amplio.

Calidad de vida

Así el interés de estudiar la calidad de vida surge de lo que se encuentra asociado a ella, autores como Fernández-Ballesteros (1998) consideran a la calidad de vida como un conjunto de condiciones de carácter ambiental, psicosocial y socioeconómico, que involucran entre otras cosas los aspectos materiales u objetivos de la subsistencia de los individuos como el acceso al empleo, a la educación, a sistemas de salud o de vivienda según necesidades de las personas, pero también elementos no materiales, subjetivos, como la elaboración cognoscitiva que efectúan las personas de sus propias condiciones de vida y las expectativas de bienestar personal y social. De tal manera, la calidad de vida se constituye en un indicador de las necesidades que han sido cubiertas por parte de los individuos, además de los requerimientos no alcanzados que pueden traducirse en malestar individual o social, y en consecuencia comportamientos atípicos o anómalos como los asociados a la delincuencia, violencia interindividual o familiar.

En este sentido una propuesta metodológica para observar la calidad de vida en contextos comunitarios debe ser, por la naturaleza del objeto de estudio, versátil, heterogénea y compleja, no por esto desordenada o menos científica que la cultivada desde las tradiciones más ortodoxas ancladas en el paradigma experimental.

Lejos de las discusiones epistemológicas concernientes al método científico, la realidad comunitaria exige un abordaje plural desde el punto de vista técnico, de tal manera, la propuesta que se describe en el presente trabajo intenta conjugar procedimientos cultivados en las tradiciones cuantitativas y cualitativas, lo que algunos autores como Creswell y Plano (2011) han dado en llamar diseños o enfoques metodológicos mixtos, y que a juicio de los autores se debería de calificar como una metodología de campo, independientemente de los sectarismos filosóficos.

De tal manera, la observación en la comunidad requiere de observaciones precisas que ningún instrumento o herramienta de recogida de datos por sí solo puede dar, ninguno es lo suficiente extenso y sensible, al menos por el momento, para captar la complejidad de la realidad comunitaria, aun cuando se delimiten las variables como en el caso de la calidad de vida.

Propuesta metodológica para el estudio de las comunidades

Herramientas metodológicas

En esta metodología de campo se pueden distinguir aquellos procedimientos o herramientas provenientes de la clasificación tradicional del método cualitativo y cuantitativo referidos en textos como los de Hernández (2018), Clark-Carter,(2002), Kerlinger y Lee, (2002), Taylor y Bogdan, (2013), Denzin y Lincoln (2011), Flick (2018), entre otros, además de herramientas que son empleadas por investigadores de ambas tradiciones así como las que emergen a raíz del desarrollo de nuevas tecnologías, particularmente en el campo de la informática o computación. A continuación, se listan algunos de ellos, los cuales son referidos por su mayor visibilidad al momento de ser planteada la presente propuesta metodológica, sin embargo, no constituyen una revisión exhaustiva por lo que pueden variar con el tiempo, en este sentido el acento debe estar siempre en la integración de las herramientas:

- Cualitativos: detección de informantes privilegiados, diario de campo, entrevistas no estructuradas o semi estructuradas, registros biográficos, cuestionarios cualitativos.
- Cuantitativos: muestreos probabilísticos por racimos o entidades geográficas, cuestionarios objetivos (cerrados o mixtos).
- Comunes: registros fotográficos, redes semánticas.
- Emergentes: georeferenciación digital y por computador.

Sin embargo, pese a la procedencia de diferentes tradiciones metodológicas y epistemológicas, las anteriores técnicas y herramientas pueden al conjuntarse en un mismo estudio, aportar información más rica, que usando sólo una de las vías de recolección o análisis de la información. Por ejemplo, una fotografía puede

proporcionar datos de una variable como “condiciones de vida” diferentes a los obtenidos a través de un cuestionario, pero la integración de ambos captará información que por separado no podrían. A continuación, se describen las características más relevantes de las técnicas y procedimientos que integran la propuesta de acuerdo con su procedencia metodológica:

a) Cualitativas

- *Detección de informantes privilegiados*: constituye una técnica de muestreo y desarrollo de redes de informantes. En los textos de metodología, particularmente los de investigación cualitativa (Naupas, Valdivia, Pañacios y Romero, 2018; Taylor y Bogdan, 2013), se refiere como uno de las actividades iniciales que se deben practicar en el proceso de acercamiento al campo o comunidad objetivo, la finalidad es establecer un red de informantes que por su condición dentro del grupo social investigado, pueden proporcionar información sobre aspectos relativos al inicio, desarrollo, interacciones o costumbres más significativas de la comunidad.
- *Diario de campo*: constituye un registro, día a día, de eventos, características o circunstancias más relevantes del o sobre el contexto comunitario. La información registrada es de gran utilidad como referente interpretativo o complementario de otras fuentes de recolección de información. En el abordaje de la calidad de vida en comunidades, es útil en tanto compendio de notas de apreciación sobre la geografía de los espacios comunitarios, o del ambiente social de los mismos.
- *Entrevistas y cuestionarios cualitativos*: la definición más simple y llana refiere la interacción de dos partes, el entrevistador y el entrevistado, el que pregunta y el que responde, sin embargo, en el trabajo de campo, más que una situación unilateral, es un intercambio de ideas o creencias sobre diversos aspectos del objeto de estudio, este intercambio, debe estar modulado por los intereses del estudio, pero sin que constituyan un cerco para la libre expresión, sobre todo, del habitante de la comunidad. De tal manera, se aconseja contar con un guion sobre temas a indagar, más que una compilación de preguntas rígidas.
- *Registros biográficos*: se debe contar con un medio para captar la narración biográfica de los habitantes de la comunidad, esta captación se puede realizar a través de lápiz y papel, o grabaciones de voz a través de dispositivos de registro de audio portátil, o incluso la videograbación, actualmente se puede efectuar hasta por dispositivos electrónicos o virtuales, como el caso del computador, teléfonos móviles o el servicio de mensajería por internet. Independientemente del medio, lo importante es el registro de la vida de las personas, que en el transcurso de las mismas han experimentado la comunidad y por ser una narración de vida se puede identificar aspectos de la calidad de la misma, de su evolución.

b) Cuantitativas

- *Muestras probabilísticas por racimos o entidades geográficas*: la posibilidad de relacionar el muestreo por informantes privilegiados, con la selección de habitantes de la comunidad y entidades físicas (casas, residencias, espacios de recreo, entre otros), dota de riqueza al método, superando a los abordajes tradicionales con sólo un método de muestreo. La unidad de selección aquí la constituye el espacio físico que se habita o se ocupa para algún fin en la comunidad. Los conteos de los espacios deben practicarse en vivo o *in situ*, para ello se puede utilizar el vagabundeo (Bautista, 2015; Taylor y Bogdan, 2013) o el tránsito a pie por las calles, callejones, veredas, sendas o avenidas de la comunidad. La precisión en el conteo de los espacios se encuentra en el registro total de viviendas o espacios a pie, sin embargo, con fines de ahorro se pueden seleccionar de forma aleatoria la cantidad de sendas, calles o avenidas, para posteriormente visitarlas, realizar los registros, y obtener una estimación de la cantidad promedio de casas o espacios por calle o senda, y después de toda la comunidad. En esta actividad es útil el registro de viviendas desocupadas o en ruina que se debe efectuar en el diario de campo, además de la asistencia de un mapa de la zona o el uso de la versión más actual del *Google Earth*, que dependiendo de la misma puede ser útil en la marcación de referencias geográficas como ubicación de informantes privilegiados (ver Figura 1) y ubicación de las calles, manzanas o espacios de la comunidad para la realización de muestreos (ver Figura 2).

Figura 1. Ubicación de informantes privilegiados en una colonia mediante *Google Earth*.



Figura 2. Ubicación de espacios públicos representativos de una colonia mediante *Google Earth*.



- *Cuestionarios objetivos (cerrados o mixtos)*: existe una gran variedad de formatos sobre instrumentos de medición objetiva referidos en la literatura, que van desde cuestionarios cerrados, mixtos, hasta escalas, donde las más tradicionales son las de Likert y Osgood (Hogg y Vaughan, 2014, Hernández, 2018; Nadelsticher, 1983). Bajo la propuesta de estudio de la calidad de vida, los instrumentos que son útiles para la medición de los factores objetivos y subjetivos de la calidad de vida pueden determinarse fácilmente con preguntas cerradas, especialmente para evaluar las condiciones de la vivienda y de los espacios, así las preguntas comunes sobre la cantidad de habitaciones, focos, habitantes, dimensiones de la vivienda o el terreno son útiles y fáciles de responder. En cuanto a los aspectos subjetivos de la calidad de vida, como la apreciación o la satisfacción de las condiciones de la vivienda o de algún servicio, se puede realizar mediante escalamiento Likert, empleando afirmaciones del tipo, *las dimensiones de mi vivienda cubren mis necesidades de espacio*, y series de adjetivos bipolares tipo Osgood (Hogg y Vaughan, 2014), un ejemplo puede ser la siguiente sentencia sobre la vivienda con sus respectivos adjetivos: *mi casa es: grande-pequeña, fea-bonita, fuerte-débil, espaciosa-estrecha*.

Para el desarrollo de un cuestionario objetivo o escalamiento en formato Likert u Osgood, se recomienda consultar a Nadelsticher (1983) y Osterlind (1997), este último describe un procedimiento de validación de ítems, sustentado en el Índice de Congruencia del Ítem de Rovinelli y Hambleton (1977), versátil en lo referente al tipo de ítems que se pueden probar mediante la utilización de jueces, que califican el ajuste o congruencia del ítem a una dimensión o variable. Esta técnica de validación

ha sido referida en uno de los capítulos del presente libro. El método de Osterlind (1997) se sintetiza en la siguiente fórmula:

$$I_{ik} = \frac{(N-1) \sum_{j=1}^n X_{ijk} + N \sum_{j=1}^n X_{ijk} - \sum_{j=1}^n X_{ijk}}{2(N-1)n}$$

Donde:

N= número de dominios o dimensiones

X_{ijk} = el valor que cada juez j , otorga a un ítem k , en la dimensión o dominio i

n = número de jueces

De igual manera, se recomienda para la validación de contenido del ítem, el uso del Índice de Congruencia del Ítem propuesto por Rovinelli y Hambleton (1977) mostrado por Turner, Mulvenon, Thomas y Balkin (2002) y Turner y Carlson (2003):

$$I_{ik} = \frac{(N-1) \sum_{j=1}^n X_{ijk} - \sum_{t=1}^N \sum_{j=1}^n X_{ijk} + \sum_{j=1}^n X_{ijk}}{2(N-1)n}$$

El significado de los símbolos de esta fórmula, es el mismo que los presentados en la propuesta de Osterlind (1997).

c) Comunes

- *Registros fotográficos*: la fotografía e incluso la grabación en video pueden aportar datos valiosos en el estudio de los fenómenos de corte psicosocial. El análisis fotográfico y el videoanálisis se han constituido en técnicas valiosas para la obtención de datos y su posterior interpretación de registros visuales y auditivos. Para una revisión sobre las ventajas del videoanálisis se puede consultar a Knoblauch, Schnettler, Raab y Soeffner (2006). En el estudio de la calidad de vida, la fotografía de los espacios comunitarios, de las viviendas, de los comportamientos cotidianos y de los eventos significativos –fiestas, ritos y otros- aportan una riqueza informativa difícil de obtener mediante otros medios convencionales, como los cuestionarios. El registro fotográfico tiene dos momentos que son fundamentales para que su uso sea relevante al trabajo: *la toma de la fotografía y el análisis de la imagen*. En cuanto al primer aspecto se debe considerar que elemento fotografiar: una vivienda, una persona, una fiesta, un conjunto de comportamientos. Esto deberá ser guiado por los intereses del estudio, así mismo por los objetos de estudio o sus variables. Por lo que respecta al análisis de la fotografía, éste se puede realizar de manera manual o con asistencia de la

computadora a través de un software, la intención es ir codificando de manera inductiva o deductiva los fragmentos de información visual que aporten datos sobre la calidad de vida de la comunidad.

- *Redes semánticas*: en la evaluación de los elementos subjetivos de la calidad de vida, algo que puede ser de gran útil es el uso de las redes semánticas naturales, en dos sentidos, como herramienta para obtener indicadores de calidad de vida para el posterior desarrollo de instrumentos de medida objetivos, y el segundo, como es que se ha planteado en la presente propuesta, como un registro de los significados sobre “calidad de vida”. De tal manera, se puede explorar los perfiles semánticos de la calidad de vida, como se significa en cada contexto comunitario, controlando un posible sesgo del observador o investigador cuando define *a priori* en un conjunto de preguntas a la “calidad de vida”.

d) Emergentes: geo-referenciación digital y por computador

El uso de dispositivos de ubicación geográfica, como los sistemas de posicionamiento geosatelital o de posicionamiento global (GPS), se han ido acreditando paulatinamente dentro de la investigación en campo, en particular cuando se requiere de las referencias geográficas de algunos puntos de la comunidad que pueden ser de difícil ubicación y acceso, de tal manera las coordenadas son un dato que puede eliminar debates sobre las delimitaciones o posiciones comunitarias. Si bien el GPS es útil, en nuestra propuesta apostamos al uso de sistemas de georreferencia por computador, en particular por algunos sistemas de ubicación geográfica disponibles en internet de forma gratuita, como es el caso de Google Earth ® (GE), dispositivo mediante el cual se pueden obtener fotografías satelitales de los espacios comunitarios, si bien la actualización de tales imágenes depende de la versión del GE la calidad de las mismas permite realizar labor de marcaje de lugares de interés para el estudio así como realizar notas en cuanto a las condiciones de vida de un determinado sector de la comunidad, inclusive puede incentivar el desarrollo de teoría de la transmisión, propagación, evolución o desarrollo geográfico de los procesos psicosociales o de diversa índole, entre ellos la formación o transmisión de la idea de bienestar social o calidad de vida en una comunidad. Como dato adicional, algunos de los programas de análisis cualitativo que existen actualmente en el mercado permiten producir y trabajar con datos georreferenciados, como es el caso del MAXQDA (Huchim y Reyes, 2013).

En la Figura 3 y 4 se muestran dos formas de emplear el GE para georreferenciar espacios estratégicos de una colonia o área residencial, o bien marcar zonas con distinta calidad de vida. Ambas imágenes provienen de estudios no publicados por los autores del presente capítulo.

Figura 3. Puntos estratégicos de una colonia o área residencial



Puntos relevantes en la colonia: 1: Acceso a la colonia; 2 y 3: Base de transporte público; 5: Capilla; 6: Módulo de policía; 7: Escuela secundaria; 8: Campo de fútbol; 9: Escuela de oficios.

Figura 4. Mapa de la calidad de vida en la comunidad



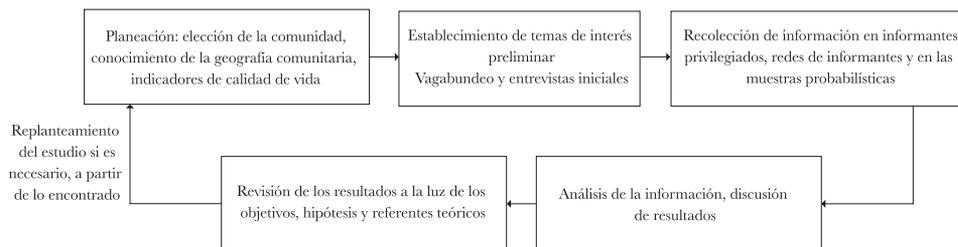
Regiones de calidad de vida percibida: 1: buena calidad de vida; 2: calidad de vida media; 3: baja calidad de vida.

Forma de iniciar el abordaje o técnica de acercamiento al campo

Considerando el objetivo del estudio, y los aspectos de diseño implicados en cualquier trabajo de investigación, que van desde los cuestionamientos, variables, hasta las

técnicas de análisis de datos, el evaluar la calidad de vida o grado de satisfacción percibida de las condiciones de vida por parte de los integrantes comunitarios, nos plantea la forma de abordar al objeto de estudio, la que se puede describir a partir de la siguiente Figura 5:

Figura 5. Replanteamiento del estudio



Podemos observar que, de manera general, una vez planteado el problema de estudiar la calidad de vida, el abordaje psicosocial de la comunidad es una secuencia de pasos que inicia con la elección de la comunidad meta, ésta debe ser conocida de forma preliminar por el investigador con la asistencia de diversas herramientas, como el uso del GE o el registro fotográfico cuando se visite a la comunidad a pie, con el objeto de realizar un estudio detallado de la geografía comunitaria. En cuanto a la incursión inicial del investigador en la comunidad se sugiere que ésta se realice a pie, salvo en casos excepcionales donde las características del terreno o la zona hagan peligroso el tránsito de la manera recomendada. El uso de otros sistemas cartográficos, como los disponibles por instituciones gubernamentales como el INEGI o el Registro Federal de Electores pueden resultar particularmente útiles en la empresa.

Posteriormente, en una segunda incursión se debe iniciar la ubicación de los informantes privilegiados y el establecimiento de redes de informantes, esto también es conocido como técnica de *bola de nieve* (Taylor y Bogdan, 2013; Fuentelsaz, 2004). La duración de esta etapa puede ser variable en función de la facilidad con que se detecten a los habitantes significativos de la comunidad, entre ellos pueden figurar los presidentes de colonia, representantes de manzanas, profesores, ministros de culto, y algunas profesiones que por el servicio estratégico que brindan a los habitantes acumulan una gran cantidad de información sobre la dinámica de la comunidad y gozan de reconocimiento de la misma, como es el caso del médico de la colonia.

La ubicación y entrevista de los informantes privilegiados conduce a una siguiente etapa, menos artesanal y mayormente técnica en el sentido del método rígido de la perspectiva cuantitativa, que se inaugura con la posibilidad de

realizar muestreos por racimos, conglomerados (Fuentelsaz, 2004) o por entidades geográficas, como las manzanas y viviendas, donde la selección aleatoria de las unidades de análisis es posible, de tal manera, la aplicación de entrevistas más estructuradas y cuestionarios objetivos se convierte en una de las actividades prioritarias. El auxilio, cuando sea posible, de cintas métricas metálicas o de material flexible para obtener medidas precisas de los espacios puede ser valioso.

La obtención de datos orales, escritos, y visuales, deben ser analizados a través de la asistencia del computador, para ello existe una gran variedad de software, por citar los más comerciales para el análisis de información codificada a priori en un instrumento objetivo o por técnicas como las redes semánticas, se puede utilizar el SPSS, STATA, MINITAB, PSPP o EPI INFO, en cuanto a la información proveniente de las entrevistas a profundidad o semiestructuradas, y registros fotográficos, el software es igualmente variado, entre ellos destacan el ATLAS TI, NVIVO, ETNOGRAPH, o MAXQDA, se debe aclarar que de ninguna manera los softwares referidos constituyen una lista exhaustiva de los actuales desarrollos en análisis de datos por computadora, sólo una muestra.

Finalmente, los datos llevarán a la obtención de resultados y la discusión de los mismos, que al ser comparados con las preguntas y objetivos que dieron origen al estudio de la calidad de vida, podrán sugerir modificaciones y replanteamientos, con el propósito de realizar futuros estudios.

Hasta el momento se ha descrito de forma superficial las fases contenidas en el esquema, lo cual obedece al objetivo del presente capítulo, que no es realizar una revisión profunda del procedimiento de cada etapa, sino centrarse en una de ellas, la de recolección de información, que contempla la aplicación de las técnicas de recogida de datos y de cómo éstas al encontrarse integradas bajo el marco de un mismo estudio se complementan para obtener mayores beneficios informativos.

Técnicas de recolección de información

Vagabundeo:

El vagabundeo es una técnica de acercamiento al campo o contexto físico donde se realizarán las observaciones y medidas (Rodríguez, Gil y García, 1996). La utilidad reside en que permite contar con un conocimiento previo del lugar donde se trabajará a futuro, ubicando las características socioespaciales más relevantes de la comunidad, es decir, aquellos espacios que cumplen con funciones específicas y reconocidas por los habitantes de la zona, como la iglesia, un parque, una plaza, espacios recreativos o deportivos. Así mismo, permite realizar un registro de la geografía comunitaria, contar y ubicar manzanas, viviendas, calles, para futuros muestreos, de igual manera, efectuar el contacto inicial con personas que por su

posición de alto contacto social pueden llevarnos a los informantes privilegiados, este es el caso de los comerciantes en pequeño, poseedores de “misceláneas”, “tiendas de abarrotes”, “tortillerías”, puestos en “tianguis” o mercados.

Otra utilidad que puede ser obtenida del vagabundeo o el desplazamiento libre por las calles de la comunidad es el inicio de los marcajes informativos sobre costumbres detectadas en la comunidad, comportamientos sincrónicos, como los tiempos escolares, laborales, de comidas, entre otros, que en su momento nos podrán aportar datos para tener una visión más acabada de la vida en la comunidad.

Todo lo que se pueda obtener mediante esta técnica deberá ser registrado en una libreta, de manera manual o de forma electrónica en dispositivos para tal efecto, sin embargo, es necesario realizar esta actividad que es el inicio de un diario o bitácora de campo.

Diario o bitácora de campo

Constituye una herramienta de registro cotidiano de las actividades y observaciones relativas al objeto de estudio que efectúa el investigador, se sugiere se realice en dos apartados, el de notas *in situ* y el de rememoración al final del día, en este último se intentará integrar todas las observaciones efectuadas con ayuda de las notas *in situ*.

El diario de campo es de gran ayuda, ya que permite orientar al investigador en las acciones a continuar, reflexionando sobre lo hecho y mejorándolo, además es una rica veta informativa sobre aspectos que no estaban contemplados desde el inicio y de depuración de los previstos.

El formato de diario de campo es flexible y acorde a los recursos del investigador, de tal manera, puede ser a lápiz y papel en una libreta, se aconseja que sea de un tamaño manejable, se debe hacer notar que debe ser de fácil acceso y manejo para la toma de notas de forma repentina y rápida; en cuanto a los formatos electrónicos, se deben considerar sus dimensiones, la fidelidad en cuanto al registro, reducción de ruido, y particularmente en la capacidad de almacenamiento o tiempo de registro de información, es inapropiado que en el campo las restricciones surjan por falta de espacio en las memorias electrónicas o de energía eléctrica.

Entrevista no estructurada

Después de la detección de informantes privilegiados efectuado en la fase de vagabundeo o aproximación inicial al campo, es esencial establecer el contacto con ellos mediante sesiones de entrevistas semiestructuradas, a fin de obtener información esencial sobre el desarrollo de la comunidad, inicios, reconocimiento oficial, y génesis de comportamiento social compartido.

De igual forma la entrevista debe recoger indicadores subjetivos y objetivos de la calidad de vida en dos niveles: individual y colectivo.

Los informantes privilegiados por su posición sobresaliente en la comunidad pueden aportar referentes sobre las condiciones de la comunidad y la satisfacción general hacia las mismas, en tanto representante del colectivo la percepción de calidad de vida de éstos no sólo es un indicador individual sino colectivo. Por otra parte, los informantes privilegiados proporcionan datos sobre la evolución de las condiciones de vida en diferentes períodos de la comunidad y la satisfacción hacia ellas de parte de los residentes.

La fase de entrevista no sólo debe considerarse como una sesión de preguntas y respuestas apegadas a un guión o cuestionario, en todo momento el investigador debe ser flexible y encontrarse abierto a la formulación de cuestionamientos a propósito de fotografías o elementos del medio ambiente físico y social, como algunos objetos representativos de fiestas o huellas físicas del entorno, como la falta de pintura o deterioro de la misma en las paredes de una habitación. En resumen, se debe ser sensible a todo indicador de las condiciones de vida de los entrevistados y de la comunidad.

Puede ser útil el uso de la narración autobiográfica del entrevistado en lo referente al periodo de residencia en la comunidad, o el análisis de los álbumes fotográficos familiares, en la búsqueda de logros personales, eventos satisfactorios o indicadores de bonanza o deterioro económico, por referir algunos aspectos de interés.

Registro fotográfico

Los métodos hasta aquí revisados prioritariamente tienen un registro verbal, en palabras escritas o audibles, finalmente es el lenguaje el medio donde se imprimen las representaciones de la realidad, la imagen. Sin embargo, pese a que el lenguaje verbal en su formato escrito u oral es el medio más familiar a la mayoría de los humanos para comunicar ideas, apreciaciones, emociones y demás eventos personales o colectivos, se ve limitado en cuanto a la capacidad descriptiva, en *otras palabras*, y señalado en cursiva, el lenguaje no son los objetos, no es la realidad en sí misma, sino una metáfora más, quizá una protometáfora, pero metáfora al fin de cuentas. Esta condición es la que ha dado tanta materia de discusión a metodólogos cuantitativos y cualitativos sobre la objetividad de las observaciones, que debería ser llamada objetivación de la observación. De tal manera, en tanto que las palabras siempre ofrecen un margen de error entre lo que representan y el objeto representado, se considera necesario utilizar mecanismos que permitan tener mayores referentes de la realidad investigada, a fin de compararla con los

ítems ya establecidos. Estos mecanismos, entonces, cumplirían la valiosa función de servir de referentes, pero también de complementos de los datos ya existentes. De tal manera, deberán estar codificados en un formato distinto al verbal, en este sentido el visual es prometedor.

Un medio de registro visual es la fotografía, a través de ella se puede capturar la imagen de escenas comunitarias, que pueden enriquecer nuestros ítems sobre las costumbres o la calidad de vida. Sin embargo, y esto es válido para cualquier medio de registro o medición, también tendrá su margen de error que dependerá de si lo fotografiado corresponde a lo que se pretende valorar, si es representativo y finalmente cómo será interpretado o codificado a partir de las categorías de análisis respectivas. Es usual que una fotografía o imagen, muestre escenas que pueden ser interpretadas en múltiples direcciones, por lo tanto, los juicios que se deriven de ellas deben encontrarse enmarcados dentro de las definiciones de las categorías de análisis, en nuestro caso, de lo relativo a la calidad de vida y sus indicadores. Por otra parte, tiene igual importancia o más lo que se encuentra contenido en la fotografía; lo que es fotografiado debe representar lo más preciso posible, la totalidad del objeto observado, y este objeto debe corresponder a los factores que se estudia, es decir, la fotografía debe cumplir con las condiciones de validez de contenido y de constructo, necesarias a todo dispositivo de registro o medida científico.

Figura 6



Una fotografía tomada sin considerar los anteriores aspectos puede ser igual de engañosa o tendenciosa que el juicio de un sujeto sobre una parcela de la realidad tomada por la realidad misma, pudiendo aplicarse la típica ilustración de los dichos de los ciegos tocando sólo una parte de un elefante. De manera gráfica lo anterior se puede observar en las Figuras 6 y 7, ambas son fotografías del mismo espacio público, una jardinera y un sendero dentro de un parque, la primera de ellas tiene en el centro un objeto sobre el cual se encuentra escrito un texto relativo al lugar, sin

embargo más allá de la legibilidad de las palabras contenidas en él, ese simple hecho de encontrarse en el centro puede sesgar por completo la extracción de información y su futura interpretación el caso de la segunda imagen es distinto, sí bien incluye al mismo objeto, el hecho de que se encuentre en un extremo no produce que la atención del observador se centre en él y permite ver otros detalles del lugar como bancas, jardineras y sendas, lo que aporta mayor información y en consecuencia una imagen con mayor validez o más cercana a lo que es el lugar fotografiado.

Figura 7



Cuestionarios de Calidad de Vida

Dentro de las evaluaciones de la calidad de vida se han considerado diferentes aspectos o indicadores de la misma en el desarrollo de instrumentos objetivos, en la literatura se refieren escalas y cuestionarios que parten de distintas concepciones de la calidad de vida, de tal manera, se encuentran aquellas que tienen su origen en definiciones extensas como la de la Organización Mundial de la Salud (OMS) que se sustenta en la definición de salud, considerada como un estado general de bienestar físico, psicológico y social, así como en aquellos que parten de una base psicosociológica de la calidad de vida, dando un mayor peso a las valoraciones subjetivas de las condiciones de vida, logros alcanzados, expectativas de desarrollo y percepción del entorno social.

El empleo de instrumentos objetivos de recolección de datos, en el sentido clásico de la medición, proporciona información concreta con respecto a las variables estudiadas, esto posibilita el uso de procedimientos de estadística inferencial con el propósito de realizar generalizaciones a la población y predicciones sobre la evolución de las condiciones de vida de la comunidad, así mismo abre la opción de trabajar mediciones temporales bajo la perspectiva de los diseños longitudinales o de cohortes.

Por lo tanto, se recomienda el uso de estos en el estudio de la calidad de vida en contextos comunitarios, en particular dentro de la presente propuesta metodológica.

Técnicas de asociación semántica

Desde que la psicología cognoscitiva develó el papel fundamental que tiene el pensamiento o los procesos que intervienen en la formación de éste en el comportamiento humano, la evaluación de elementos cognoscitivos ha cobrado notable importancia para la elaboración de modelos explicativos del comportamiento, de tal manera, se han diseñado diversas técnicas de medición u observación de las cogniciones, el ejemplo más notable lo constituyen las actitudes, las cuales han recibido una gran atención en psicología social, propiciando la elaboración de formatos escalares como el Likert, Thurstone, u Osgood.

Si bien, los dispositivos objetivos en formato de escala han gozado de una gran aceptación, existen otros que si bien no proporcionan cuantificaciones de un constructo cognoscitivo, si aportan densidad informativa sobre los contenidos del pensamiento o la base informativa de los sistemas de creencias, que entre otras cosas, permite establecer hipótesis sobre la génesis, desarrollo y estructura de los esquemas cognoscitivos individuales y colectivos, los cuales se objetivan a través del lenguaje, de tal manera, el pensamiento se encuentra representado por las palabras, y estas integradas por significados, que cambian de un sujeto a otro, de un grupo a otro y de una época a otra. De lo anterior, se pueden revisar el pensamiento a través de un registro de los significados, dentro de las referidas en la propuesta se encuentran las técnicas de asociación libre y las redes semánticas naturales.

Esta última es descrita por Reyes-Lagunes (1993) y Figueroa (1976), como un método para obtener redes de significados asociados a un constructo, concepto o palabra, que hace referencia a un objeto concreto o simbólico. De tal manera, las redes semánticas son un excelente recurso para conocer las significaciones de la calidad de vida, que bajo un contexto teórico sociocognoscitivo, los significados se encuentran íntimamente ligados a la acción de los sujetos, de tal suerte, lo que se piense de un objeto social, que es social porque es compartido por un grupo o colectivo, puede conducir las acciones de la comunidad con respecto a las decisiones de los integrantes comunitarios para mejorar la calidad de vida, en resumen, los significados de la calidad de vida pueden ofrecernos un indicador de los referentes del bienestar de la comunidad, en particular, cómo se dimensiona el bienestar social y la calidad de vida.

La técnica es descrita de forma extensa en Reyes-Lagunes (1993), lo que se muestra aquí es el procedimiento general de aplicación de las redes semánticas, empero, para pormenores se recomienda consultar la anterior referencia. De forma

inicial se tiene que decidir cuántas palabras, sustantivos o adjetivos se solicitarán a las personas, se puede optar entre cinco a 10 palabras, para una aplicación rápida y evitar la fatiga del respondiente se sugiere pedir cinco palabras. Las respuestas deben ser con respecto a un estímulo verbal, aunque también se pueden utilizar imágenes o sonidos como estímulos, en nuestro caso se puede emplear la frase “calidad de vida”. Posteriormente se debe solicitar a las personas que jerarquicen del uno al cinco sus respuestas en orden de importancia o relevancia semántica, es decir, la palabra o concepto que mejor represente o signifique a la calidad de vida se le asignará un uno y así sucesivamente, hasta llegar al cinco, que corresponderá a la palabra menos representativa.

El análisis de la información obtenida a través de la red semántica consiste en obtener un conjunto de palabras o significados representativos de la calidad de vida, para ello se deben calcular los pesos semánticos (PS) para cada respuesta, lo que se realiza a partir de la suma de las jerarquías inversas (5 cuando se asigna 1, 4 para 2, 2 para 4, 1 para 5) asignadas a cada palabra por parte de los entrevistados. La suma de estos puntajes derivados de la jerarquización original de los sujetos constituye el PS, con el que se ordenarán las palabras de mayor a menor, mediante este procedimiento se determinará el núcleo de la red (NR), que se integrará por las 20 palabras con mayores puntuaciones semánticas. Con lo anterior se podría afirmar que se tendrán los contenidos centrales o al menos los referentes de significado más importantes de la calidad de vida para los integrantes de una comunidad.

Secuencia de aplicación de las técnicas

Teniendo en cuenta que, en cualquier iniciativa de observación de la calidad de vida en comunidades, la aplicación de las técnicas de recolección de datos anteriormente descritas puede ser simultánea y no necesariamente lineal, se ha considerado dividir la fase de trabajo de campo o recogida de información en tres diferentes momentos, caracterizados por actividades y aplicación de técnicas específicas:

- a) *Momento inicial*: vagabundeo, diario o bitácora de campo, entrevista de detección de informantes privilegiados, registro fotográfico.
- b) *Segundo momento (informantes privilegiados)*: diario de campo, entrevista semi estructurada, registro fotográfico, análisis fotográfico, narrativas biográficas.
- c) *Tercer momento (muestras probabilísticas de habitantes de la comunidad)*: diario de campo, registro fotográfico, cuestionarios objetivos de calidad de vida y redes semánticas naturales.

A manera de ilustración de cómo ocurren las anteriores fases se presenta el siguiente caso de investigación empírica no publicado:

En un estudio comunitario se obtuvo información precisa sobre indicadores materiales y cognoscitivos de la calidad de vida en habitantes de dos comunidades urbano marginales de la ciudad de Tampico, Tamaulipas, información que permitiera en el futuro desarrollar programas de intervención para la promoción de conductas asociadas al incremento de la calidad de vida de las comunidades observadas. Se utilizó un diseño mixto combinando la etnografía con un diseño transversal, empleando técnicas de muestreo probabilístico para la selección de 180 viviendas de una población de 338, y no probabilístico en la detección de informantes privilegiados mediante la técnica de “bola de nieve”. *En el primer momento* de la fase de campo, durante tres semanas aproximadamente, se realizaron caminatas por las dos comunidades con el propósito de conocer las condiciones del medio material y social, ubicar espacios comunitarios y detectar a través de entrevistas no estructuradas con algunos residentes a informantes privilegiados, en esta fase se tomaron fotografías, registros de audio y anotaciones diarias en la bitácora de campo. En el *segundo momento*, se contactó y entrevistó a los informantes privilegiados referidos por los residentes inicialmente. En esta fase se utilizó la entrevista no estructurada, el diario de campo y registros fotográficos, además se solicitó a los entrevistados que relataran su vida en la comunidad (biografías). En el *tercer momento*, al conocer previamente el área física de la comunidad compuesta por calles, casas, espacios públicos, entre otros, se realizó el muestreo probabilístico de viviendas, en las cuales se aplicó un cuestionario objetivo mixto sobre la percepción de la calidad de vida, así mismo se continuó empleando la bitácora y los registros fotográficos. Los resultados obtenidos sugieren que existe en algunos rubros de la calidad de vida, en particular en los materiales, una diferencia entre la evaluación física y la percepción de los mismos por parte de los entrevistados, así se puede observar que mientras se carece de un servicio de alcantarillado eficiente que evite las inundaciones en los meses de verano o de un espacio funcional de recreo comunitario, los habitantes de ambas colonias refieren estar satisfechos en términos generales de las condiciones materiales. Los elementos estimados negativamente fueron los sanitarios, seguridad pública y la percepción de distancia física con los centros laborales, este último factor asociado a la baja calidad del transporte público. Sin embargo, el no reconocimiento por parte de los encuestados, de las condiciones que facilitan estos problemas comunitarios o el enmascaramiento de las mismas con una percepción de calidad de vida favorable, ha impactado en la baja movilidad de la comunidad para la mejora de sus actuales condiciones de vida, movilidad que en sus inicios le dio origen.

Conclusiones

La propuesta metodológica para el estudio de la calidad de vida en comunidades, presentada en este capítulo, pretende integrar procesos y herramientas de las tradiciones metodológicas cuantitativa y cualitativa, considerando que la calidad de vida por definición es una variable compleja y por lo tanto es necesario más de una fuente de recolección de información o datos para tener una medida válida y confiable de la misma.

Cómo en todo estudio científico, esta propuesta parte de un planteamiento general sobre la calidad de vida en una comunidad, que involucra la elección de la misma y los propósitos específicos, pasando por la fase de campo, análisis de datos, obtención y discusión de resultados, para llegar a un escenario donde se puede concluir el proceso investigativo o bien hacer replanteamientos e iniciar con nuevas observaciones.

En términos operativos la propuesta metodológica tiene su ocurrencia en el trabajo de campo, en tres momentos, donde el uso de las herramientas descritas a lo largo del capítulo no es lineal, puede ser simultánea y constante, ya que se considera que la riqueza de información sobre la calidad de vida es la que puede dar una mejor medida de la misma, de tal manera, la secuencias de las etapas, herramientas o técnicas descritas se pueden resumir en la siguiente Tabla 1.

Tabla 1. Propuesta metodológica para la evaluación de la calidad de vida en síntesis

| Etapa | Procesos | Herramientas o técnicas | Productos |
|--|--|---|--|
| Planteamiento | Establecimiento de preguntas de investigación sobre la calidad de vida, objetivos, selección de la comunidad o comunidades, planeación del trabajo de campo y análisis de datos. | Revisión de literatura, detección de revisiones sistemáticas y metaanálisis sobre calidad de vida. Revisión de diversas fuentes documentales | Proyecto de investigación sobre calidad de vida. |
| Trabajo de campo Primer momento o inicial | Contacto inicial con el campo o espacios físico-sociales de la comunidad. Registro de las calles, avenidas, | Vagabundeo Entrevistas semiestructuradas. Toma de fotografías. | Obtención de información sobre las prácticas y lugares relevantes de la comunidad. |

| Etapa | Procesos | Herramientas o técnicas | Productos |
|--|---|---|---|
| Trabajo de campo Primer momento o inicial | estimación de viviendas y espacios públicos o relevantes. Entrevistas a residentes y detección de informantes privilegiados. Registros fotográficos, de audio y visuales. | Diario o bitácora de campo. Empleo de dispositivos de registro de audio. Marcaje a través de software para georreferencia. | Detección de informantes privilegiados. Fotografías, archivos de audio y visuales. Información inicial sobre la calidad de vida de la comunidad. Transcripción de entrevistas. |
| Segundo momento | Entrevista de informantes privilegiados. Obtención de narrativas biográficas o relatos de vida. Registros audio-visuales. | Diario de campo. Entrevista semi estructurada. Registro fotográfico. Análisis fotográfico. Video análisis. Narrativas biográficas. Uso de software para georreferencia. | Información sobre la percepción de la calidad de vida en informantes privilegiados. Descripción histórica de la calidad de vida comunitaria a través de narrativas biográficas. Fotografías, archivos de audio y visuales. Mapa de la calidad de vida. Trascripción de entrevistas. |
| Tercer momento | Muestreo probabilístico por racimos o estratificado de viviendas. Aplicación de instrumentos objetivos de calidad de vida y de asociación. | Diario de campo, registro fotográfico, cuestionarios objetivos de calidad de vida y redes semánticas naturales. | Índice e indicadores numéricos de calidad de vida comunitaria. Modelo de la calidad de vida. Significados de la calidad de vida. Fotografías. Matrices de datos obtenidos por cuestionarios y redes semánticas. |
| Obtención y discusión de resultados | Análisis de datos e información. | Software de análisis | Informe de resultados. |

| Etapa | Procesos | Herramientas o técnicas | Productos |
|-------------------------------------|--|---------------------------|---|
| Obtención y discusión de resultados | Discusión de resultados considerando los planteamientos iniciales | cualitativo y estadístico | Informe de resultados |
| Replanteamiento | Evaluación de los alcances y limitaciones del trabajo. Desarrollo de nuevas líneas de trabajo sobre la calidad de vida. | Análisis de resultados | Informe final o elaboración de un nuevo proyecto. |

Fuente: Elaboración propia.

Referencias

- Ardila, R. (2003). "Calidad de vida: una definición integradora". *Revista Latinoamericana de Psicología*. Vol.35 (2), pp. 161-164. Recuperado el 31 de mayo de 2020, de <https://www.redalyc.org/pdf/805/80535203.pdf>
- Arita W., B. Y. (2005). "La capacidad y el bienestar subjetivo como dimensiones de estudio de la calidad de vida". *Revista Colombiana de Psicología*. 14, pp. 73-79. Recuperado el 31 de mayo de 2020, en <https://www.redalyc.org/pdf/804/80401407.pdf>
- Bautista, N. P. (2015). *Proceso de la investigación cualitativa. Epistemología, metodología, aplicaciones*. México: Manual Moderno.
- Clark-Carter, D. (2002). *Investigación cuantitativa en psicología*. México: Oxford.
- Creswell, J. W. & Plano C., V. L. (2011). *Designing and conducting Mixed Methods research*. Thousand Oaks, CA: Sage Publication.
- Denzin, N. K. y Lincoln, Y. S. [Coords.] (2011). *Manual de Investigación Cualitativa*. [Vol. I]. Barcelona, España: Editorial Gedisa.
- Fernández-Ballesteros, R. (1998). "Evaluación psicológica: algunas claves para el futuro". *Papeles del Psicólogo*, 70, 3-12. Recuperado de <http://www.papelesdelpsicologo.es/resumen?pii=780>
- Figuroa, J. G. (1976). *Estudios de redes semánticas y algunos procesos básicos*. D.F., México: Universidad Nacional Autónoma de México.
- Flick, U. (2018). *Introducción a la investigación cualitativa* [4ª. Edición]. Madrid, España: Morata.
- Fuentelsaz G., C. (2004). "Cálculo del tamaño de la muestra". *Matronas Profesión*. Vol. 5 (18), pp. 5-13. <https://dialnet.unirioja.es/servlet/articulo?codigo=6125940>
- Hernández, S. R. (2018). *Metodología de la investigación: Las rutas cuantitativas, cualitativas y mixtas*. México: McGraw Hill.

- Hogg, M. A. & Vaughan, G. M. (2014). *Psicología Social*. México: Médica panamericana.
- Huchim, A. D. y Reyes, C. R. (2013). “La investigación biográfico-narrativa para el estudio de los docentes”. *Actualidades investigativas de la educación*, 13(3), 1-27. Recuperado de <https://www.redalyc.org/pdf/447/44729878019.pdf>
- Indices of Item Congruence for Test Development Validity Assessments. *SAS Conference Proceedings: SAS Users Group International 27*. 14-17 April 2002, Orlando, Florida, USA, 1-5.
- Kerlinger, F. N. y Lee, H. B. (2002). *Investigación del Comportamiento*. México, D.F.: McGraw-Gill.
- Knoblauch, H., Schnettler, B., Raab, J. y Soeffner, H-G. (2006). “Video Analysis: Methodology and Methods. Qualitative audiovisual Data Analysis in Sociology”. Recuperado de: http://icar.univ-lyon2.fr/ecole_thematique/TRANAL_I/documents/video/mond06_video.pdf
- Nadelsticher M., A. (1983). *Técnicas para la construcción de cuestionarios de actitudes y opción múltiple*. México: Instituto Nacional de Ciencias Penales.
- Naupas, P. H., Valdivia, D. M. Palacios, V. J. y Romero, D. H. (2018). *Metodología de la investigación: cuantitativa-cualitativa*. México: Ediciones de la U.
- Osterlind, S. J. (1997). *Constructing Test Items: Multiple-Choice, Constructed-Response, Performance and Other Formats*. New York: Kluwer Academic Publishers.
- Prieto, E. G. (1982). “Los indicadores en la medición de niveles de bienestar social”. *Revista de Política Social*, 133, 109-135. Recuperado de <https://dialnet.unirioja.es/revista/11241/A/1982>
- Reyes-Lagunes, I. (1993). “Las Redes semánticas naturales, su concepción y su utilización en la construcción de instrumentos”. *Revista de Psicología Social y Personalidad*, 9(1), 81-97.
- Rodríguez, G. G., Gil, F. J. y García, J. E. (1996). *Metodología de la investigación cualitativa*. Granada, España: Aljibe.
- Rovinelli, R. J. y Hambleton, R. K. (1977). “On the use of content specialists in the assessment of criterion-referenced test item validity”. *Dutch Journal of Educational Research*, 2, 49-60.
- Taylor, S.J. y Bogdan, R. (2013). *Introducción a los métodos cualitativos de investigación*. Barcelona: Paidós.
- Turner, R. C. y Carlson, L. (2003). “Indexes of Item-Objective Congruence for Multidimensional Items”. *International Journal of Testing*, 3 (2), 163-171.
- Turner, R. C., Mulvenon, S. W., Thomas, S. P. y Balkin, R. S. (2002). Computing

Capítulo 4

La evaluación de la calidad científica en ciencias del comportamiento: una aproximación a las escalas y cuestionarios de calidad¹

Carro-Pérez, Ennio Héctor

Resumen

La evaluación de la calidad científica de los estudios primarios e intervenciones es una actividad fundamental para conocer el nivel de certidumbre o sesgo de los hallazgos y sus efectos, sin ella no contaríamos con información suficiente para conocer si los esfuerzos destinados en una determinada línea de trabajo se realizan de manera adecuada y son fructíferos en tanto contribuciones al conocimiento de un fenómeno. De tal manera, en algunas disciplinas científicas, como las médicas, el desarrollo de procedimientos y herramientas en formato de cuestionarios o escalas para evaluar la calidad de los estudios ha sido abundante, en otras ha sido escaso, esto último es notorio en aquellas que tienen que ver con el estudio del comportamiento humano, y específicamente en el contexto mexicano. El presente trabajo presenta una revisión preliminar de los protocolos y herramientas de evaluación desarrollados en las ciencias de la salud y de la psicología, se describe cada uno de ellos de manera general y se indican las condiciones de su uso.

Palabras clave: calidad científica, escalas, protocolos, ciencias del comportamiento, ciencias de la salud.

Introducción

El presente capítulo tiene como propósito establecer la relevancia de los procesos de evaluación de la calidad científica, a través de la descripción básica de algunos de los principales protocolos e instrumentos de medida para la obtención de indicadores e índices de calidad, aplicables a las ciencias del comportamiento y de la salud.

Para introducir el tema podemos partir de la consideración de que la medición de la calidad científica hace referencia a un conjunto de conceptos y de prácticas

¹ Parte del material de este capítulo también se encuentra en la Tesis Doctoral no publicada del autor: *Avances metodológicos en la evaluación de la calidad de la producción científica (investigaciones y tesis de pregrado y postgrado) en psicología desarrolladas en la región centro y noreste de México*. Texto completo disponible para consulta en el Depósito de Investigación de la Biblioteca de la Universidad de Sevilla, en el siguiente enlace: <https://idus.us.es/handle/11441/11078/browse?type=author&value=Carro+P%C3%A9rez%2C+Ennio+H%C3%A9ctor>

de evaluación respecto al cómo se realizan o ejecutan las diversas iniciativas de investigación e intervención dentro de un campo específico de la ciencia, la calidad en este sentido no sólo implica llevar a cabo una serie de pasos o acciones, sino también la intención y la manera en que son ejecutadas estas acciones, es decir, comprende la revisión de los factores y criterios con los que se toma una decisión respecto a qué métodos o técnicas se emplean para el control de sesgos. En este sentido, si la evaluación de la calidad científica se concretará sólo a verificar la ocurrencia de actividades se encontraría incompleta ya que se limitaría a ser un proceso descriptivo y superficial sobre el cómo se hace la ciencia, sin embargo hasta el momento la mayor parte de la evaluación y escalas desarrolladas para este efecto han caído en lo anterior, dejando al margen la observación de las intenciones o criterios de decisión de los investigadores, lo que se evidencia incluso en los procesos de evaluación de las contribuciones postuladas a las revistas científicas (Rozemblum, Unzurrunzaga, Banzato, y Pucacco, 2015).

Para precisar, la calidad científica involucra el cumplimiento del diseño metodológico en cuanto a sus características y la precisión con las que fueron realizadas. Verhagen (Cascaes, Valdivia, da Rosa, Barbosa y da Silva, 2013) refiere a la calidad científica como calidad metodológica del estudio, la cual tiene que ver con la validez interna, la validez externa, la transparencia en el planteamiento de los objetivos, el tamaño de la muestra para un determinado efecto y la presentación de resultados. En resumen, la calidad científica es una sumatoria de condiciones necesarias para la obtención de un resultado con un alto grado de certidumbre y probabilidad.

Así, evaluar y regular la calidad de los estudios primarios o empíricos es una pieza clave en los procesos de construcción de la ciencia y de la toma de decisiones en el terreno aplicado para la solución de problemas. Muestra de esta importancia es el desarrollo de dispositivos de medida para estimar la calidad científica o normar el contenido de estudios primarios e incluso de trabajos de revisión y metaanálisis, algunas escalas o iniciativas como la declaración CONSORT (Consolidated Standards of Reporting Randomized Trials), TREND (Transparent Reporting of Evaluations with Nonrandomized Designs), STROBE (Strengthening the Reporting of Observational Studies in Epidemiology), SRQR (Standards for Reporting Qualitative Research) entre otras, son muestra de ello.

Protocolos y mediciones de la calidad científica

La evaluación de la calidad de la producción científica en cualquier disciplina es de gran importancia, una vez que se conoce a detalle los procedimientos de realización de un trabajo y los elementos constitutivos de su diseño se puede tener un grado

de certidumbre sobre la validez y confianza de todo el proceso de indagación empírica y de sus resultados, lo anterior si consideramos como hipótesis lo señalado por Sanduvete (2008) sobre la existencia de una correlación negativa entre los sesgos de los estudios y la calidad de los mismos. De tal manera, sesgos como los referidos por Jovell y Navarro-Rubio (1995), en estudios con alta calidad estarían controlados o inexistentes, sin embargo, sesgos como el del editor, en el caso de una revista científica, podrían escapar a una evaluación centrada en los productos de investigación.

Lo señalado nos debe alertar sobre la necesidad de acabar con mitos como el de la imparcialidad en la investigación científica, si bien existe en teoría un método relativamente pulcro, al ejecutarlo se pueden cometer omisiones u “olvidos” que los autores podrían no estar dispuestos a informar. De igual manera, considerar como principio de corrección que la ciencia está mediada por intereses de grupo, cuyas prácticas se encuentran por arriba de la calidad de los trabajos, puede traer más beneficios que perjuicios a la producción de conocimiento científico (Sanduvete, 2008). Un par de acciones que considera esta posibilidad, aunque no la combaten del todo, es la incorporación dentro de las políticas editoriales de las revistas de una declaratoria sobre conflicto de intereses por parte de los autores (Luna, Jolly, Bevilacqua y Jäger, 2011), así como el requerimiento hecho a revistas, por parte de índices o bases de datos científicas, de no exceder cuotas de publicación para investigadores locales o pertenecientes a la institución editora.²

El interés por la evaluación de los procesos de investigación ha conducido en el área de la salud a la medicina basada en evidencias (MBE) (Delgado y Palma, 2006), y en menor proporción, en el estudio del comportamiento humano, a la necesidad de una psicología de este tipo (Daset y Cracco, 2013).

En este sentido, se ha demandado un esfuerzo a los metodólogos de estas áreas para producir técnicas o escalas de calidad como el cuestionario de Chalmers desarrollado en 1981 (Delgado y Palma, 2006) o la escala de Jadad (ver Tabla 1) en 1996 (Cascaes, Valdivia, da Rosa, Barbosa y da Silva, 2013), la declaración QUOROM (Quality of Reporting of Meta-Analyses) (Moher, Cook, Eastwood,

² Para referencias sobre los indicadores de calidad exigidos a nivel nacional a publicaciones periódicas se puede consultar los criterios de evaluación del Manual del Sistema de Clasificación de Revistas Mexicanas de Ciencia y Tecnología [Edición 2019], del Consejo Nacional de Ciencia y Tecnología (Conacyt) de México, disponibles en <http://www.revistascytconacyt.mx/manual-sistema-crmcyt2019.pdf>, o bien si se requiere una muestra a nivel internacional, se pueden revisar los criterios de inclusión de índices de revistas científicas como la Red de Revistas Científicas de América Latina y el Caribe (Redalyc), disponibles en <https://www.redalyc.org/postulacion.oa>.

Olkin, Rennie, Stroup et al., 2000), la iniciativa CONSORT (Moher, et al., 2000; Delgado y Palma, 2006; Sanduverte, 2008), la TREND, STRICTA (Standards for Reporting Interventions in Controlled Trials of Acupuncture) (MacPherson, White, Cummings, Jobst, Rose y Niemtow, 2002), la declaración STARD (Standards for Reporting of Diagnostic Accuracy) para la evaluación de las pruebas diagnósticas, y la declaración STROBE para los estudios de observación (Delgado y Palma, 2006), entre otras. Si bien, estos trabajos han provenido mayormente desde la medicina, el origen de estos esfuerzos se puede ubicar en la psicología cuando Glass (1976) refiere un tipo de estudio que hoy puede ser utilizado en la evaluación de la calidad científica de estudios primarios, el metaanálisis.

Tabla 1. Escala de Jadad (García y Aboitiz, 2012: 89)³

| Ítem | Pregunta | Respuesta |
|------|---|-------------------|
| 1 | ¿El estudio fue descrito como aleatorio? | Sí/No |
| 2 | ¿Se describe el método para generar la secuencia de aleatorización y este método es adecuado? | Sí//No Inadecuado |
| 3 | ¿El estudio se describe como doble ciego? | Sí/No |
| 4 | ¿Se describe el método de cegamiento y este método es adecuado? | Sí/No Inadecuado |
| 5 | ¿Existió una descripción de las pérdidas y las retiradas? | Sí/No |

Considerando la importancia de estas herramientas para la evaluación científica, se describirán algunas de las más reconocidas en la literatura (Cascaes, Valdivia, da Rosa, Barbosa y da Silva, 2013) y en la red *Equator*⁴ (Enhancing the QUALity and Transparency Of health Research), que dentro de sus cualidades más relevantes destacan la amplitud de campos científicos donde pueden ser empleadas y el formato de lista de verificación que posibilita la aplicación tipo escalamiento con dos opciones de respuesta “Sí/No” o “Cumple/No cumple”. Estas respuestas pueden ser codificadas con 1 y 0 y ser sumadas, con lo que se contaría con un indicador cuantitativo de la calidad del estudio.

Uno de estos desarrollos es la escala CONSORT, integrada por 22 ítems en cinco dimensiones, confeccionados para evaluar los ensayos clínicos aleatorios

³ Tomado de García C., D. y Aboitiz C., J. (2012). “Efectividad del entrenamiento de los músculos del suelo pélvico en la incontinencia urinaria: revisión sistemática”. *Fisioterapia*. 34 (2), pp. 87-95.

⁴ Sitio de internet de la red para la “Mejora de la Calidad y la Transparencia de la Investigación en Salud”: <https://www.equator-network.org/>

(Delgado y Palma, 2006; Turpin, 2005; Moher, Schulz, Altman y Grupo CONSORT, 2002), a continuación se reproduce el protocolo presentado por Moher et al. (2002:25), y Turpin (2005:682) (ver Tabla 2).

Tabla 2. Declaración CONSORT. Lista de comprobación de puntos a incluir en un informe de ensayo aleatorizado

| Dimensiones y temas | Ítem | Descriptor | Informado en página |
|--|------|--|---------------------|
| Título y resumen | 1 | Identificación del tipo de estudio en el título El resumen, incluye diseño, método, resultados, conclusiones, como se asignan los participantes a las intervenciones (p.e. asignación al azar o no aleatorizada) | |
| Introducción antecedentes | 2 | Antecedentes científicos, explicación y razonamiento | |
| Métodos participantes | 3 | Criterios de elección de los participantes así como los dispositivos, servicios donde los datos fueron recolectados | |
| Intervenciones | 4 | Precisar detalles de las intervenciones o tratamientos en cada grupo, así como cuándo y cómo fueron administradas | |
| Objetivos | 5 | Especificar los objetivos y las hipótesis | |
| Resultados | 6 | Definir claramente las medidas primarias y secundarias de los resultados y, cuando sea aplicable, cualquier método utilizado para incrementar la calidad de las mediciones (p.e. múltiples observaciones o entrenamiento previo de los observadores) | |
| Tamaño de la muestra | 7 | Cómo fue determinado el tamaño de la muestra y cuándo es aplicable, la explicación de cualquier análisis intermedio y las reglas de finalización | |
| Aleatorización: generación de la secuencia | 8 | Método usado para generar la secuencia aleatoria de asignación, incluyendo detalles de cualquier restricción | |
| Aleatorización: distribución a ciegas | 9 | Método usado para implementar la secuencia aleatoria de la asignación (p.e. contenedores numerados o guía telefónica), clarificando si la secuencia fue ocultada hasta que las intervenciones fueron asignadas | |
| Aleatorización: implementación | 10 | Quién generó la secuencia de asignación, quién enroló a los participantes y quien los asignó a sus grupos | |

| Dimensiones y temas | Ítem | Descriptor | Informado en página |
|---|------|---|---------------------|
| Ciego (enmascaramiento) | 11 | Si los participantes, los que administraron la intervención, y los que evaluaron los resultados fueron ciegos a la asignación de grupos. De haber sido así, como se evalúo el éxito del proceso de cegado (enmascaramiento) | |
| Métodos estadísticos | 12 | Métodos estadísticos utilizados para comparar los grupos en sus resultados primarios; métodos de análisis adicional, tales como análisis de subgrupos o análisis ajustados | |
| Resultados Flujo de participantes | 13 | Flujo de participantes a través de cada fase o estadio (uso de diagrama de flujo). Específicamente para cada grupo informar el número de participantes asignados aleatoriamente, recibiendo el tratamiento pretendido, complementando el protocolo de estudio y analizado para los resultados primarios. Describir las desviaciones del protocolo de estudio diseñado y las razones | |
| Reclutamiento | 14 | Fechas definiendo los periodos de reclutamiento y seguimiento | |
| Datos de base | 15 | Datos demográficos de base y características clínicas de cada grupo | |
| Números de análisis | 16 | Número de participantes (denominador) en cada grupo incluido en cada análisis y si el análisis fue por “intención de tratar”. Establecer los resultados en números absolutos cuando sea factible | |
| Resultados y estimación | 17 | Para cada resultado primario y secundario, un resumen de resultados de cada grupo, y el efecto estimado de la muestra y su precisión (p.e. el intervalo de confianza de 95 %) | |
| Análisis auxiliar | 18 | Agregar multiplicidad informando cualquier otro análisis realizado, incluyendo análisis de subgrupos y análisis ajustados, incluyendo aquellos previamente especificados y aquellos aleatorios | |
| Eventos adversos | 19 | Todos los eventos adversos importantes o efectos colaterales en cada grupo de intervención | |
| Comentarios/discusión interpretación/ limitaciones* | 20 | Limitaciones del ensayo. Interpretación de los resultados tomando en cuenta las hipótesis del estudio, las fuentes de sesgo potencial o de imprecisión, y peligros asociados con la multiplicidad de análisis y resultados | |

| Dimensiones y temas | Ítem | Descriptor | Informado en página |
|-------------------------------|------|--|---------------------|
| Generalización | 21 | Generalización (validez externa) de los hallazgos del estudio | |
| Evidencia global | 22 | Interpretación general de los resultados en el contexto de la evidencia actual | |
| Otra información* Registro | 23 | Número de registro y nombre del registro de prueba | |
| Protocolo | 24 | Dónde se puede acceder al protocolo de prueba completo, si está disponible | |
| Financiamiento | 25 | Fuentes de financiación y otro tipo de apoyo (como suministro de medicamentos), función de los financiadores | |

*Elementos disponibles en la Lista CONSORT 2010

Fuente: Lista CONSORT 2010, que puede consultarse en <https://www.equator-network.org/reporting-guidelines/consort/>

Cómo en otros protocolos de calidad, en CONSORT se debe registrar si el autor informa o no el tópico respectivo y la página en que lo hace, así mismo se pide al autor de la revisión o metaanálisis presente un diagrama de flujo del proceso seguido (Turpin, 2005:683),

Otra escala de calidad, surgida en la medicina, es la declaración STROBE, útil en la evaluación de los estudios observacionales en Epidemiología, ésta tiene su paralelo en la declaración MOOSE (Meta-analysis of Observational Studies in Epidemiology) para la valoración de metaanálisis. La STROBE se encuentra compuesta por 22 ítems agrupados en seis dimensiones muy parecidas a las que conforman la CONSORT (ver Tabla 3) (Elm, Altman, Egger, Pocock, Gotsche y Vandembroucke, 2007:249).

Tabla 3. Declaración STROBE.⁵ Lista de puntos que deben ser incluidos en los reportes de estudios observacionales*

| Dimensiones y temas | Ítem | Recomendaciones |
|---------------------|------|---|
| Título y resumen | 1 | a) Indicar el diseño del estudio con términos comúnmente usados en el título o el resumen |

⁵ Para mayor información consultar el trabajo de: Elm, E. et al., (2007). "The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for reporting observational studies". *Preventive Medicine*, 45, 247-251.

| Dimensiones y temas | Ítem | Recomendaciones |
|---|------|---|
| Título y resumen | 1 | b) Proveer en el resumen información y una síntesis equilibrada de que fue realizado y que fue encontrado Introducción |
| Antecedentes/ razones o justificación | 2 | Explicar los antecedentes científicos y razones por las que la investigación ha sido reportada |
| Objetivos | 3 | Especificar los objetivos planteados, incluyendo una especificación previa de las hipótesis |
| Métodos | | |
| Diseño del estudio | 4 | Presentar elementos clave sobre el diseño del estudio al principio del trabajo |
| Lugares o escenarios | 5 | Describir los escenarios o lugares, fechas relevantes, incluyendo periodo de reclutamiento, exposición, seguimiento, y recolección de datos |
| Participantes | 6 | a) <i>Estudio Cohorte</i> : proporcionar los criterios de elegibilidad, y las fuentes y métodos de selección de los participantes. Descripción de los métodos de seguimiento Estudio de caso control: proporcionar los criterios de elegibilidad, y las fuentes y métodos para la determinación del caso y selección de control. Dar las razones para la selección del caso y el control. Estudio transeccional o transversal: proporcionar los criterios de elegibilidad, y las fuentes y métodos de selección de los participantes b) <i>Estudio Cohorte</i> : para estudios combinados o apareados, proporcionar los criterios de la combinación o apareamiento y número de expuestos y no expuestos. Estudio de caso control: para estudios combinados o apareados, proporcionar los criterios de la combinación o apareamiento y el número de controles por caso |
| Variables | 7 | Definición clara de todos los resultados, exposiciones o tratamientos, predictores, equivocaciones o errores potenciales, y efecto de modificaciones. Si son aplicables, dar criterios de diagnóstico |
| Origen de los datos/ mediciones | 8 | Por cada variable de interés, proporcionar la fuente de los datos y detalle de los métodos de medición. Describir comparativamente los métodos de medición si hay más de un grupo |
| Sesgos | 9 | Describir los esfuerzos para identificar y abordar las potenciales fuentes de sesgo |
| Tamaño del estudio | 10 | Explicar cómo el tamaño del estudio fue conseguido |

| Dimensiones y temas | Ítem | Recomendaciones |
|-------------------------|------|--|
| Variables cuantitativas | 11 | Explicar cómo las variables cuantitativas fueron manipuladas en los análisis. Si aplica, describir cuáles agrupaciones fueron elegidas, y por qué |
| Métodos estadísticos | 12 | <p>a) Describir todos los métodos estadísticos, incluidos los usados para el control de las equivocaciones</p> <p>b) Describir cualquier método empleado para examinar los subgrupos y las interacciones</p> <p>c) Explicar cómo fueron tratados los datos perdidos</p> <p>d) <i>Estudio Cohorte</i>: si aplica, explicar cómo se trataron las pérdidas en el seguimiento</p> <p><i>Estudio de caso control</i>: si aplica, explicar cómo las combinaciones de casos y controles fueron tratados</p> <p><i>Estudio transeccional o transversal</i>: si aplica, describir los métodos de análisis para disponer de las estrategias de muestreo</p> <p>e) Describir los análisis de sensibilidad</p> |
| Resultados | | |
| Participantes | 13 | <p>a) Reportar el número de individuos para cada fase del estudio</p> <p>b) Proporcionar las razones para la no participación en cada fase</p> <p>c) Considerar el uso de un diagrama de flujo</p> |
| Datos descriptivos | 14 | <p>a) Proporcionar las características de los participantes del estudio (p.e. demográficos, clínicos, sociales) e información sobre las exposiciones y confusiones potenciales</p> <p>b) Indicar el número de participantes con datos perdidos en cada variable de interés</p> <p>c) <i>Estudio Cohorte</i>: síntesis cuantitativa del periodo de seguimiento (p.e. promedio y cantidad total)</p> |
| Datos arrojados | 15 | <p><i>Estudio Cohorte</i>: reporte de números de eventos resultados o síntesis de las mediciones por tiempo</p> <p><i>Estudio de caso control</i>: reporte de números en cada categoría de exposición o tratamiento, o la síntesis promedio de la exposición</p> <p><i>Estudio transeccional o transversal</i>: reporte de números de eventos resultados o síntesis de las mediciones</p> |
| Resultados principales | 16 | <p>a) Proporcionar estimaciones del ajuste y, si aplica, estimación del ajuste del error y su precisión (p.e. intervalos de confianza de 95 %). Clarificar qué confusiones o equivocaciones fueron ajustadas y porque estas fueron incluidas.</p> <p>b) Reportar los límites de la categoría cuando variables continuas fueron categorizadas.</p> <p>c) Si es relevante, considerar la transformación de las estimaciones de riesgo relativo a riesgo absoluto por un periodo significativo de tiempo.</p> |

| Dimensiones y temas | Ítem | Recomendaciones |
|---------------------|------|--|
| Otros análisis | 17 | Reportar otros análisis realizados (p.e. análisis de subgrupos e interacciones, y análisis de sensibilidad) |
| Discusión | | |
| Resultados clave | 18 | Resumen de los resultados clave con referencia a los objetivos del estudio |
| Limitaciones | 19 | Discusión de las limitaciones del estudio, recuperando las fuentes de sesgo potencial o imprecisión. Discusión sobre la dirección y magnitud de cualquier sesgo potencial |
| Interpretación | 20 | Dar una interpretación global cautelosa de los resultados considerados objetivos, limitaciones, diversidad de análisis, resultados para estudios similares, y otra evidencia relevante |
| Generalización | 21 | Discusión sobre la generalización de los resultados del estudio (validez externa) |
| Otra información | | |
| Financiamiento | 22 | Proporcionar las fuentes de financiamiento y el papel de los patrocinadores en el presente estudio y, si aplica, el estudio original en el que se basó el presente artículo |

*STROBE tiene versiones específicas para estudios Cohort, de Caso-control, y Transversales, disponibles en la Red Equator (<https://www.equator-network.org/reporting-guidelines/strobe/>).

Como se puede apreciar, el protocolo de la declaración STROBE y el CONSORT, no sólo son semejantes en la cantidad de ítems y dimensiones que los integran, sino también en el contenido de las mismas, si acaso la STROBE por estar orientada a estudios observacionales es más específica en cuanto a los tipos de estudio que puede optar un trabajo observacional, por ello la distinción entre los estudios tipo Cohorte, de caso controlado y transversal.

Dentro de las escalas referidas por Sanduvete (2008) se encuentra la TREND, utilizada para valorar los estudios con diseños no aleatorios, la estructura y conformación de la misma es similar a las anteriormente revisadas, nuevamente los puntos de distinción son las características de los tipos de estudio o diseños. (ver Tabla 4).

Tabla 4. Escala para estudios con diseños no aleatorios TREND (Des Jarlais, D., Lyles, C., Crepaz, N. y TREND Group, 2004:363-365).

| Dimensiones y temas | Ítem | Descriptor |
|-------------------------|------|--|
| Título y resumen | 1 | Información sobre cómo las unidades fueron ubicadas para la intervención. Uso de la estructura recomendada para el resumen. Información sobre la población objetivo o muestra del estudio |
| Introducción | | |
| Antecedentes | 2 | Antecedentes científicos y explicación de razones (justificación) Teorías utilizadas en el diseño de las intervenciones conductuales |
| Métodos | | |
| Participantes | 3 | Criterios de elegibilidad de los participantes, incluyendo los criterios para diferentes niveles en el reclutamiento/diseño de muestreo Método de reclutamiento, incluyendo el método de muestreo, si un muestreo sistemático fue efectuado Lugares de reclutamiento Lugares donde los datos fueron recolectados |
| Intervenciones | 4 | Detalles de la intervención para cada condición del estudio, cómo, cuándo y dónde se administró, reportando específicamente: - Contenido: ¿qué fue administrado? - Método de suministro: ¿cómo fue dado el contenido? - Unidades de suministro: ¿cómo fueron agrupados los sujetos durante la administración? - Lugar: ¿dónde fue administrada la intervención? - Cantidad de exposición y duración: ¿cantidad de sesiones empleadas para la administración del tratamiento? ¿duración de las sesiones para la administración del tratamiento? - Tiempo de extensión: ¿qué tiempo se requirió para la toma del tratamiento o intervención para cada una de las unidades - Actividades para aumentar el acuerdo o consentimiento (p.e. incentivos) |
| Objetivos | 5 | Especificar objetivos e hipótesis |
| Obtención de resultados | 6 | Definir con claridad las mediciones primarias y secundarias Métodos usados para la recolección de los datos y cualquier método utilizados para proporcionar calidad a las mediciones Información sobre la validación de los instrumentos |
| Tamaño de la muestra | 7 | Cómo se determinó el tamaño de la muestra, donde aplique, explicación de cualquier análisis interino y detención de reglas |

| Dimensiones y temas | Ítem | Descriptor |
|-------------------------------------|------|--|
| Método de asignación | 8 | Unidad de asignación a cada una de las condiciones del estudio (individuo, grupo, comunidad) Método utilizado para la asignación de las unidades a las condiciones del estudio, incluyendo detalles de cualquier restricción Reporte de los mecanismos empleados para disminuir el potencial de sesgo inducido por la no aleatorización (p.e. apareamiento) |
| Ciego (enmascaramiento) | 9 | Si o no los participantes, los que administraron la intervención, los que evaluaron los resultados fueron cegados en cada condición del estudio; además informar como el enmascaramiento fue acordado y como fue evaluado |
| Unidad de análisis | 10 | Descripción de la más pequeña unidad que fuera analizada para la evaluación del efecto de la intervención Si la unidad de análisis difirió de la unidad de asignación, qué método analítico fue usado para contarla |
| Métodos estadísticos | 11 | Métodos estadísticos usados para comparar los resultados de los grupos de estudios primarios, incluyendo métodos complejos de correlación de datos Métodos estadísticos empleados para análisis adicionales, tales como análisis de los subgrupos y análisis de ajuste Métodos para tratar datos perdidos Uso de programas o software estadístico |
| Resultados | | |
| Flujo de los participantes | 12 | Flujo de los participantes en cada etapa del estudio (se recomienda el uso de un diagrama) <i>Enlistamiento o selección:</i> número de participantes examinados para la selección, criterios de inclusión y exclusión <i>Asignación:</i> número de participantes asignados a cada condición <i>Colocación y exposición de la intervención:</i> número de participantes asignados a cada condición del estudio, y número de participantes que fueron expuestos a la intervención o tratamiento <i>Seguimiento:</i> número de participantes que concluyeron el tratamiento y aquellos que no lo concluyeron por cada condición del estudio <i>Análisis:</i> número de participantes incluidos y excluidos de los análisis principales, por condición del estudio Descripción de las desviaciones del protocolo del estudio como fue diseñado, exponiendo las razones |
| Reclutamiento | 13 | Reporte de las fechas, de los periodos de reclutamiento y seguimiento. |
| Información de partida (Línea base) | 14 | Información demográfica, características clínicas de los participantes en cada condición del estudio. |

| Dimensiones y temas | Ítem | Descriptor |
|-------------------------------------|------|--|
| Información de partida (Línea base) | 14 | Características iniciales relevantes de cada condición del estudio para especificar la prevención de la enfermedad Línea base para las comparaciones de las pérdidas en el seguimiento, de forma global y para cada condición del estudio Comparación entre la población estudiada como línea base y la población de interés |
| Equivalencia de Línea base | 15 | Datos sobre la igualación de los grupos de estudio, línea base y métodos estadísticos empleados para controlar las diferencias iniciales |
| Números analizados | 16 | Número de participantes incluidos en cada análisis de cada condición, particularmente cuando los participantes cambiaron en los diferentes resultados, reporte de los resultados en números absolutos cuando sea posible Indicar si los análisis fueron con la “intención de tratar” o si no |
| Resultados y estimaciones | 17 | Por cada resultado primario y secundario, una síntesis de resultados por cada condición del estudio, y la estimación del tamaño de efecto e intervalo de confianza para indicar la precisión Inclusión de resultados negativos y positivos Inclusión de resultados para pruebas preestablecidas de causalidad cuando el tratamiento fue efectuado |
| Análisis previos | 18 | Resumen de otros análisis realizados, incluyendo subgrupos o análisis de restricción, indicando cuales fueron preestablecidos o exploratorios |
| Eventos adversos | 19 | Síntesis de todos los eventos adversos importantes o efectos no esperados en cada condición |
| Discusión | | |
| Interpretación | 20 | Interpretación de los resultados, tomando en cuenta las hipótesis del estudio, fuentes potenciales de sesgo, imprecisión de las mediciones, análisis multiplicativos, y otras limitaciones del estudio Discusión de los resultados tomando en cuenta los mecanismos por los cuales las intervenciones fueron realizadas, o explicaciones alternativas. Discusión de los éxitos y fracasos en la implementación de la intervención, y de la precisión de la implementación |
| Generalización | 21 | Generalización de los resultados (validez externa), tomando en cuenta la población del estudio, las características de la intervención, extensión del seguimiento, incentivos, niveles de acuerdo, y elementos contextuales |
| Evidencia global | 22 | Interpretación general de los resultados en el contexto de evidencia y teoría actual |

Otra escala es la STARD (Sanduverte, 2008), utilizada para evaluar trabajos de diagnóstico médico, la cual está integrada por 30 ítems en su versión 2015⁶ (ver Tabla 5). Autores como Lumbreras, Jarrín y Hernández (2006), la han empleado en la valoración metodológica de estudios sobre pruebas de genética.

Tabla 5. Escala STARD para la evaluación de la calidad de estudios de diagnóstico médico⁷

| Dimensiones y temas | Ítem | Descriptor |
|---------------------|------|---|
| Título o resumen | | |
| Resumen | 1 | Identificación como un estudio de precisión diagnóstica utilizando al menos una medida de precisión (como sensibilidad, especificidad, valores predictivos u otros) |
| Introducción | 2 | Resumen estructurado del diseño del estudio en métodos, resultados y conclusiones (para una guía específica ver la STARD para resúmenes) |
| | 3 | Antecedentes científicos y clínicos, incluido el uso previsto y el papel clínico de la prueba índice o de referencia |
| | 4 | Objetivos del estudio e hipótesis |
| Métodos | | |
| Diseño del estudio | 5 | La recopilación de datos se planificó antes de la prueba de índice y el estándar de referencia (estudio prospectivo) o después (estudio retrospectivo) |
| Participantes | 6 | Criterios de elección |
| | 7 | ¿Sobre qué base se identificaron los participantes potencialmente elegibles? (como síntomas, resultados de pruebas anteriores, inclusión en el registro) |
| | 8 | Dónde y cuándo se identificaron los participantes potencialmente elegibles (entorno, ubicación y fechas) |
| | 9 | Si los participantes formaron una serie consecutiva, aleatoria o conveniente |
| Métodos de prueba | 10 | Prueba de índice, con suficiente detalle para permitir la replicación |

⁶ La versión 2015 de la STARD reproducida en este capítulo, fue tomada y traducida del sitio de la Red Equator (<https://www.equator-network.org/reporting-guidelines/stard/>), el día 5 de marzo de 2020.

⁷ Para una mayor información sobre la STARD, consultar el trabajo de: Bossuyt, P., Reitsma, Jh., Bruns, D., Gatsonis, C., Glasziou, P., Irving, L, et al., (2003). "The STARD Statement for Reporting Studies of Diagnostic Accuracy: Explanation and Elaboration". *Annals of Internal Medicine*, 138 (1), W1-W12.

| Dimensiones y temas | Ítem | Descriptor |
|-------------------------|------|---|
| Métodos de prueba | 10 | Estándar de referencia, con suficiente detalle para permitir la replicación |
| | 11 | Justificación para elegir el estándar de referencia (si existen alternativas) |
| | 12 | Definición y justificación de los cortes de positividad de la prueba o categorías de resultados de la prueba de índice, distinguiendo lo preespecificado del exploratorio |
| | 13 | Si la información clínica y los resultados estándar de referencia estaban disponibles a los intérpretes / lectores de la prueba de índice Si la información clínica y los resultados de las pruebas de índice estaban disponibles a los evaluadores del estándar de referencia |
| Análisis | 14 | Métodos para estimar o comparar medidas de precisión diagnóstica |
| | 15 | Cómo se manejó la prueba de índice indeterminada o los resultados estándar de referencia |
| | 16 | Cómo se manejaron los datos faltantes en la prueba de índice y el estándar de referencia |
| | 17 | Cualquier análisis de variabilidad en la precisión diagnóstica, distinguiendo los pre especificados de los exploratorios |
| | 18 | Tamaño de muestra previsto y cómo se determinó |
| | | Resultados |
| Participantes | 19 | Flujo de participantes, usando un diagrama |
| | 20 | Características demográficas y clínicas basales de los participantes |
| | 21 | Distribución de la gravedad de la enfermedad en aquellos con la condición objetivo Distribución de diagnósticos alternativos en aquellos sin la condición objetivo |
| | 22 | Intervalo de tiempo y cualquier intervención clínica entre la prueba índice y el estándar de referencia |
| Resultados de la prueba | 23 | Tabulación cruzada de los resultados de la prueba de índice (o su distribución) por los resultados del estándar de referencia |
| | 24 | Estimaciones de la exactitud diagnóstica y su precisión (como intervalos de confianza del 95 %) |
| | 25 | Cualquier evento adverso al realizar la prueba de índice o el estándar de referencia |

| Dimensiones y temas | Ítem | Descriptor |
|---------------------|------|--|
| | | Discusión |
| | 26 | Limitaciones del estudio, incluidas las fuentes de sesgo potencial, incertidumbre estadística y generalización |
| | 27 | Implicaciones para la práctica, incluido el uso previsto y el papel clínico de la prueba índice |
| | | Otra información |
| | 28 | Número de registro y nombre del registro |
| | 29 | Donde se puede acceder al protocolo completo del estudio |
| | 30 | Fuentes de financiamiento y otro tipo de apoyo; papel de los patrocinadores |

La escala STARD como las anteriores sólo difiere en el grado de especificidad, debido en su mayor parte al tipo de estudio que pretende evaluar.

A las anteriores herramientas de calidad se pueden agregar la lista CARE (CAse REport), la SRQR (Standards for Reporting Qualitative Research), y la COREQ (Consolidated Criteria for Reporting Qualitative Research), estas dos últimas para la valoración de estudios cualitativos.

La CARE, integrada por 13 ítems, ha sido diseñada para evaluar el reporte de estudios de caso en el campo médico, sin embargo con algunos ajustes al contenido de sus ítems puede aplicarse a cualquier área de las ciencias sociales o del comportamiento. (ver Tabla 6).

Tabla 6. Lista CARE⁸ para la redacción de un reporte de estudio de caso

| Sección del trabajo/Tema | Ítem | Recomendaciones y Descriptores |
|--------------------------|------|---|
| Título | 1 | Las palabras “informe de caso” o “reporte de caso” deben aparecer en el título junto con lo más interesante del mismo |
| Palabras clave | 2 | Los aspectos relevantes de este caso en 2 - 5 palabras clave |
| Resumen | 3 | a) <i>Introducción</i> : ¿Qué es único en este caso? ¿Qué aporta de nuevo a la literatura médica? |

⁸ Tomado de Gagniera, J.J., Kienlec, G., Altman, D.G., Moher, D., Sox, H., Riley, D. and the CARE Group (2014). “The CARE guidelines: consensus-based clinical case report guideline development”. *Journal of Clinical Epidemiology*. 67. pp. 46-51. Recuperado el 8 de septiembre de 2015, en <http://dx.doi.org/10.1016/j.jclinepi.2013.08.003>

| Sección del trabajo/Tema | Ítem | Recomendaciones y Descriptores |
|--------------------------|------|---|
| Resumen | 3 | <ul style="list-style-type: none"> b) Los principales síntomas del paciente y los hallazgos clínicos importantes c) Los principales diagnósticos, intervenciones terapéuticas y resultados d) <i>Conclusión</i>: ¿Cuáles son las principales lecciones que se pueden extraer de este caso? |
| Introducción | 4 | Breve resumen de los antecedentes de este caso haciendo referencia a la literatura médica pertinente |
| Información del paciente | | <ul style="list-style-type: none"> a) Información demográfica (como edad, sexo, origen étnico, profesión) b) Principales síntomas de paciente (sus principales molestias) c) Historial médico, familiar y psicosocial que incluya la dieta, el estilo de vida y la información genética pertinente d) Enfermedades concomitantes pertinentes, incluyendo intervenciones anteriores y sus resultados |
| Hallazgos clínicos | 6 | Describir los hallazgos pertinentes de la exploración física |
| Línea de tiempo | 7 | Fechas y horas importantes para este caso |
| Evaluación diagnóstica | 8 | <ul style="list-style-type: none"> a) Métodos diagnósticos b) Problemas para el diagnóstico c) Razonamiento diagnóstico, incluidos otros posibles diagnósticos tenidos en cuenta d) Características de pronóstico, cuando aplique |
| Intervención terapéutica | 9 | <ul style="list-style-type: none"> a) Tipos de intervención (como farmacológica, quirúrgica, preventiva, autocuidados) b) Administración de la intervención (como dosis, concentración, duración) c) Cambios justificados en la intervención |
| Seguimiento y resultados | 10 | <ul style="list-style-type: none"> a) Resultados evaluados por el médico y por el paciente b) Resultados importantes de la prueba de seguimiento c) Adhesión y tolerabilidad a la intervención (¿cómo fue evaluada?) d) Acontecimientos adversos e imprevistos |
| Discusión | 11 | <ul style="list-style-type: none"> a) Puntos fuertes y limitaciones en el manejo de este caso b) Discusión de la literatura médica pertinente c) Justificación de las conclusiones (incluida la evaluación de las posibles causas) d) Principales lecciones que se pueden extraer de este informe de caso |

| Sección del trabajo/Tema | Ítem | Recomendaciones y Descriptores |
|--------------------------|------|--|
| Perspectiva del paciente | 12 | ¿Comunicó el paciente su perspectiva o experiencia? (Incluir siempre que sea posible) |
| Consentimiento Informado | 13 | ¿Dio su consentimiento informado el paciente? Proporcionarle si es solicitado |

Las siguientes dos normas o protocolos de evaluación son exclusivas para los estudios cualitativos, si bien las anteriores listas de verificación y escalas pueden adaptarse a los estudios de este tipo, inicialmente no contemplan aspectos específicos de esta metodología, por lo que pueden ser evaluados de manera incorrecta al requerir datos que sólo tienen sentido en la lógica experimental, como la selección aleatoria de los casos o participantes. La SRQR, una de las normas referidas, describe a través de 21 ítems agrupados en seis dimensiones, un conjunto de aspectos indispensables para elaborar un reporte o informe de investigación cualitativa (ver Tabla 7).

Tabla 7. Lista SRQR⁹ sobre los puntos a incluir en un informe de investigación cualitativa

| Dimensión | Ítem | Descriptores y Recomendaciones |
|--------------------------------------|------|--|
| Título y resumen | | |
| Título | 1 | Descripción breve de la naturaleza y el tema del estudio. Identificar el estudio como cualitativo o indicando el enfoque (por ejemplo, etnografía, teoría fundamentada) o métodos de recolección de datos (por ejemplo, la entrevista, el enfoque grupo) |
| Resumen | 2 | Resumen de los elementos clave del estudio utilizando el formato de la publicación; incluye típicamente fundamentos, propósito, métodos, resultados y conclusiones |
| Introducción | | |
| Formulación del problema | 3 | Descripción e importancia del problema/fenómeno estudiado; revisión de la teoría relevante y trabajo empírico; planteamiento del problema |
| Objetivo o pregunta de investigación | 4 | Referir el propósito de los objetivos o preguntas de estudio generales y específicas |

⁹Tomada de O'Brien, B., Harris, I.B., Beckman, Th. J., Reed, D. A. y Cook, D.A. (2014). "Standards for Reporting Qualitative Research: A Synthesis of Recommendations". *Academic Medicine*. 89 (9). pp.1245-1251. Recuperado el día 9 de septiembre de 2015, de <http://links.lww.com/ACADMED/A218>. (doi: 10.1097/ACM.0000000000000388).

| Dimensión | Ítem | Descriptores y Recomendaciones |
|--|------|--|
| Métodos | | |
| Enfoque cualitativo y paradigma de investigación | 5 | Indicar el enfoque cualitativo (etnografía, teoría fundamentada, estudio de caso, fenomenología, investigación narrativa, u otro) y la base teórica; identificar el paradigma de investigación (postpositivista, constructivista/interpretativo); justificación de la elección del enfoque teórico |
| Características del investigador y reflexividad | 6 | Características de los investigadores que pueden influir en la investigación, incluyendo atributos personales, cualificaciones/experiencia, relación con los participantes, suposiciones y/o presupuestos; interacción real o potencial entre las características de los investigadores y la investigación, preguntas, enfoque, métodos, resultados, y/o la transferibilidad |
| Contexto | 7 | Ajuste/sitio y factores contextuales más destacados; justificación del contexto. |
| Estrategia de muestreo | 8 | ¿Cómo y por qué la elección de los participantes, documentos, o eventos?; criterios para decidir cuándo el muestreo era suficiente (por ejemplo, la saturación de muestreo); justificación de los procesos |
| Cuestiones éticas relacionadas con sujetos humanos | 9 | Documentos sobre la aprobación de un comité de revisión ética, sobre el consentimiento de los participantes, seguridad y confidencialidad, o explicación para la falta de los mismos |
| Métodos de recolección de datos | 10 | Tipos de datos recolectados; detalles de los procedimientos de recolección, incluyendo (según el caso) fechas de recogida y análisis de datos, repetición de procedimientos, métodos de triangulación de las fuentes, y modificación de los procedimientos en respuesta a la evolución de los resultados del estudio; justificación |
| Instrumentos y tecnologías de recolección de datos | 11 | Descripción de los instrumentos (por ejemplo, guías de entrevistas, cuestionarios) y dispositivos de recolección (grabadoras de audio, cámaras fotográficas, entre otros); y, si es el caso, cómo el instrumento ha cambiado durante el curso del estudio |
| Unidades de estudio | 12 | Número y características relevantes de los participantes, documentos, o eventos incluidos en el estudio; nivel de participación (podría ser reportado en los resultados) |
| Procesamiento de datos | 13 | Métodos para el tratamiento de datos antes de y durante el análisis, incluyendo transcripción, entrada de datos, gestión de datos, confiabilidad y validez de los datos, codificación de datos, e identificación de extractos |

| Dimensión | Ítem | Descriptorios y Recomendaciones |
|---|------|---|
| Análisis de datos | 14 | Proceso y criterios teóricos por el cual las inferencias o categorías fueron identificadas y desarrolladas, incluyendo los investigadores involucrados en el análisis |
| Técnicas para mejorar la confiabilidad | 15 | Técnicas para mejorar la confiabilidad y la credibilidad del análisis de datos (por ejemplo, la triangulación) |
| Resultados | | |
| Síntesis e interpretación | 16 | Principales resultados (por ejemplo, interpretaciones, inferencias, y categorías); desarrollo de una teoría o modelo, entre otros |
| Relaciones con los datos empíricos | 17 | Evidencias (por ejemplo, citas, notas de campo, fragmentos de texto, fotografías) para fundamentar conclusiones analíticas |
| Discusión | | |
| Integración con el trabajo previo, implicaciones, transferibilidad, y la contribución para el campo | 18 | Breve resumen de las principales conclusiones; explicación de cómo los hallazgos y las conclusiones se conectan o apoyan un problema; discusión de ámbito de aplicación/generalización; identificación de la contribución en la disciplina o el campo |
| Limitaciones | 19 | Confiabilidad y limitaciones de los hallazgos |
| Otro | | |
| Conflicto de intereses | 20 | Las posibles fuentes de influencia o influencia percibida en la conducción del estudio y las conclusiones; cómo fueron manejados |
| Financiamiento | 21 | Fuentes de financiación y otros apoyos; papel de los donantes en los datos recolectados, interpretación y presentación de informes |

La lista COREQ al igual que la SRQR, establece un grupo de directrices o requisitos que debe cumplir un adecuado informe de investigación cualitativa, a diferencia de esta última es más específica en cuanto al tipo de estudio (ver Tabla 8).

Tabla 8. Lista COREQ¹⁰ sobre criterios unificados para la presentación de informes de investigación cualitativa, con entrevistas y grupos focales

| Dimensión | Ítem | Preguntas guía o descriptor |
|--|------|--|
| Equipo de investigación y reflexividad | | |
| Características personales Entrevistador / facilitador | 1 | ¿Qué autor o investigador llevó a cabo el grupo focal o la entrevista? |
| Credenciales, certificación, grados académicos | 2 | ¿Cuáles fueron los grados, certificaciones, o credenciales del investigador? |
| Profesión | 3 | ¿Cuál era su ocupación o profesión en el momento del estudio? |
| Género | 4 | ¿Hombre o mujer? |
| Experiencia y entrenamiento | 5 | ¿Qué experiencia o formación tenía el investigador? |
| Relación con los participantes Relación establecida | 6 | ¿La relación se estableció antes del inicio del estudio? |
| Conocimiento del participante sobre el entrevistador | 7 | ¿Qué conocían sobre el investigador? |
| Características del entrevistador. | 8 | ¿Qué características fueron reportadas sobre el entrevistador o facilitador? |
| Diseño del estudio | | |
| Marco teórico orientación metodológica y teoría | 9 | ¿Qué orientación metodológica sustenta el estudio? (teoría fundamentada, análisis del discurso, etnografía, entre otras) |
| Selección de participantes muestreo | 10 | ¿Cómo fueron seleccionados los participantes? |
| Método de abordaje | 11 | ¿Cómo fueron abordados los participantes? |
| Tamaño de la muestra | 12 | ¿Cantidad de participantes en el estudio? |
| No participantes | 13 | ¿Cuántas personas abandonaron o se negaron a participar?, Razones para ello |
| Ajuste o configuración configuración de la recopilación de datos | 14 | ¿Dónde se recogieron los datos? |

¹⁰ Tomado de Tong, A., Sainsbury, P. y Craig, J. (2007). Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care*. 19 (6), pp. 349-357.

| Dimensión | Ítem | Preguntas guía o descriptor |
|--|------|---|
| Presencia de no participación | 15 | ¿Nadie más estaba presente, además de los participantes y los investigadores? |
| Descripción de la muestra | 16 | ¿Características importantes de la muestra ? |
| Recolección de datos | 17 | ¿Se formularon preguntas, instrucciones, guías proporcionadas por los autores? ¿Fue una prueba piloto? |
| Guía de entrevista | | |
| Repetición de entrevistas | 18 | ¿Fue necesaria la repetición de entrevistas?, ¿Cuántas? |
| Grabación audiovisual | 19 | ¿La investigación utilizó grabaciones de audio y video para recolectar datos? |
| Notas de campo | 20 | ¿Las notas de campo fueron hechas durante y/o después de la entrevista o grupo focal? |
| Duración | 21 | ¿Cuál fue la duración del grupo focal o de las entrevistas? |
| Saturación de datos | 22 | ¿Se discutió la saturación de datos? |
| Transcripciones devueltas | 23 | ¿Fueron devueltas a los participantes transcripciones para comentarlas o corregirlas? |
| Análisis y resultados | | |
| Análisis de datos | 24 | ¿Cuántos codificadores de datos existieron en el estudio? |
| Número de codificadores de datos | | |
| Descripción del árbol de códigos | 25 | ¿Los autores proporcionan una descripción del árbol de codificación? |
| Derivación de temas o categorías | 26 | ¿Fueron los temas o categorías identificadas por adelantado o emergieron de los datos? |
| Software | 27 | En su caso, ¿qué software se utilizó para el tratamiento de los datos? |
| Verificación del participante | 28 | ¿Los participantes retroalimentaron los resultados? |
| Informe, citas o argumentos presentados | 29 | ¿Se presentaron citas de los participantes para ilustrar los resultados? ¿de qué participante (número)? |
| Datos y resultados consistentes | 30 | ¿Hubo coherencia entre los datos presentados y los resultados? |
| Claridad de los temas principales | 31 | ¿Los temas principales fueron presentados con claridad en los resultados? |
| Claridad de los temas de menor importancia | 32 | ¿Hay una descripción de los diversos casos o discusión de temas de menor importancia? |

Hasta el momento se han revisado escalas y directrices que originalmente surgieron en el campo de la medicina, sin embargo han existido esfuerzos desde otras áreas del

conocimiento como la psicología donde se han producido trabajos con el propósito de establecer lineamientos para la redacción de informes, una muestra son las publicaciones de Virués-Ortega y Moreno-Rodríguez (2008) sobre directrices para el reporte de casos clínicos en psicología clínica conductual (DICC's PCC),¹¹ de Leech y Onwuegbuzie (2010) sobre lineamientos para la realización y presentación de informes de investigación mixta en el área de la asesoría o consejería psicológica, o el de Ramos-Álvarez y Catena (2004), donde se exponen un conjunto de ítems dicotómicos que permiten evaluar mediante un puntaje la calidad del informe de estudios experimentales en ciencias del comportamiento. A continuación se reproducen los lineamientos propuestos por Virués-Ortega y Moreno-Rodríguez (2008) (ver Tabla 9), los cuales también se encuentran disponibles para el área en el sitio *Equator* y se organizan según la estructura tradicional de un informe de investigación para su publicación, mientras que el trabajo de Leech y Onwuegbuzie (2010), hace referencia a contenidos generales del trabajo, ya contemplados en otras herramientas.

Tabla 9. Lineamientos para los informes de casos clínicos en Psicología Clínica Conductual (DICC'sPCC)¹²

| Sección o dimensión | Contenido |
|-----------------------------|---|
| Título (20 palabras máx.) | Descripción del caso y la intervención |
| Resumen (250 palabras máx.) | Datos personales Objetivo del estudio / intervención Métodos de evaluación Tipo de tratamiento Diseño Análisis de los datos Resultados principales Seguimiento |
| Palabras clave | Añadir "experimento de caso único" |
| Introducción | Objetivos y justificación Antecedentes empíricos relevantes |

¹¹ Abreviación de los autores del presente capítulo

¹² Tomado de Virués-Ortega, J y Moreno-Rodríguez, R (2008). "Guidelines for clinical case reports in behavioral clinical Psychology". *International Journal of Clinical and Health Psychology*, Vol. 8. N° 3, pp. 173-189. Recuperado el 9 de septiembre de 2015, de www.equator-network.org

| Sección o dimensión | Contenido |
|---|--|
| Identificación del paciente y motivo de consulta o remisión | <p><i>Información descriptiva:</i> edad, sexo, educación, profesión, situación familiar</p> <p><i>Antecedentes clínicos:</i> diagnósticos y tratamientos anteriores y actuales</p> <p>Motivo de referencia</p> <p>Consentimiento informado y confidencialidad</p> <p>Tratamientos anteriores (psicológico y farmacológico)</p> |
| Estrategias de evaluación | <p>Parámetros de comportamiento y razonamiento</p> <p>Confiabilidad entre evaluadores</p> <p>La literatura empírica y razón de ser de los instrumentos</p> <p>Uso de medidas de funcionamiento general</p> <p>Evaluación de los efectos secundarios</p> |
| Formulación de casos clínicos | <p>Descripción del problema</p> <p>Circunstancias de la adquisición del problema</p> <p>Mantenimiento de las variables</p> <p>Recapitulación de la información pertinente para describir y explicar el problema del paciente</p> <p>Hipótesis</p> |
| Tratamiento | |
| Elección del tratamiento | Método para la selección del tratamiento |
| Tratamiento | Número, periodicidad, duración y contenido de las sesiones clínicas |
| Implementación | <p>Cronograma</p> <p>Descripción sesión por sesión</p> <p>Transcripción de sesiones</p> <p>Medidas de fidelidad</p> |
| Terapeuta, factores relacionados | <p>Métodos para mejorar la adherencia al tratamiento</p> <p>Uno o múltiples terapeutas</p> <p>Información sobre la experiencia del terapeuta</p> <p>Medidas de adherencia</p> <p>Métodos para minimizar el sesgo en la toma de decisiones clínicas</p> |
| Diseño del estudio | <p>Mediciones de series de tiempo, o medidas repetidas</p> <p>Diagrama de Diseño (AB, ABA, ABAB, etc.)</p> <p>Sujeto, contexto, terapeuta y relación con las variables de control del sesgo</p> |
| Análisis de datos | Justificación de análisis visual o estadístico |
| Efectividad y eficiencia de la intervención | <p>Salidas desde la línea de base</p> <p>Tamaño del efecto de Intervención</p> |

| Sección o dimensión | Contenido |
|---|---|
| Efectividad y eficiencia de la intervención | Significación clínica de la intervención Prueba pre y postest Período de seguimiento |
| Discusión | Evaluación de los resultados en el contexto de la formulación clínica, diseño de la intervención y literatura empírica relevante Aspectos más destacados del estudio Deficiencias del estudio Temas para investigaciones futuras |

Todas las escalas, directrices o iniciativas descritas previamente están dirigidas a lo que Glass (1976) denominó estudios primarios, sin embargo existe otro tipo de estudios que han tomado relevancia en la medida en que la producción científica ha ido en aumento, las revisiones sistemáticas y metaanálisis, de tal manera con el propósito de contribuir a la visibilidad de este tipo de estudios y a su evaluación en las ciencias del comportamiento se reproduce de Urrutía y Bonfill (2010, pp. 509-510), la directriz PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), que a nuestro juicio mejora lo ya hecho en protocolos como QUOROM (ver Tabla 10).

Tabla 10. Protocolo PRISMA,¹³ sobre los aspectos a incluir en una revisión sistemática o metaanálisis

| Dimensión/tema | Ítem | Contenido o descriptor |
|----------------------|------|---|
| | | Título y resumen |
| Título | 1 | Identificar la publicación como revisión sistemática, metaanálisis o ambos |
| Resumen estructurado | 2 | Facilitar un resumen estructurado que incluya, según corresponda: antecedentes; objetivos; fuente de los datos; criterios de elegibilidad de los estudios, participantes e intervenciones; evaluación de los estudios y métodos de síntesis; resultados; limitaciones; conclusiones e implicaciones de los hallazgos principales; número de registro de la revisión sistemática |

¹³ Para mayor información respecto a PRISMA, puede consultar desde www.equator-network.org, el original de 2009: Moher, D, Liberati, A, Tetzlaff, J, Altman, DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med* 6(7): e1000097. doi:10.1371/journal.pmed.1000097.

| Dimensión/tema | Ítem | Contenido o descriptor |
|--|------|--|
| Introducción | | |
| Justificación | 3 | Describir la justificación de la revisión en el contexto de lo que ya se conoce sobre el tema |
| Objetivos | 4 | Plantear de forma explícita las preguntas que se desea contestar en relación con los participantes, las intervenciones, las comparaciones, los resultados y el diseño de los estudios (PICOS) ¹⁴ |
| Métodos | | |
| Protocolo y registro | 5 | Indicar si existe un protocolo de revisión al que se pueda acceder (por ej., dirección web) y, si está disponible, la información sobre el registro, incluyendo su número de registro |
| Criterios de elegibilidad | 6 | Especificar las características de los estudios (por ej., PICOS, duración del seguimiento) y de las características (por ej., años abarcados, idiomas o estatus de publicación) utilizadas como criterios de elegibilidad y su justificación |
| Fuentes de información | 7 | Describir todas las fuentes de información (por ej., bases de datos y períodos de búsqueda, contacto con los autores para identificar estudios adicionales, etc.) En la búsqueda y la fecha de la última búsqueda realizada |
| Búsqueda | 8 | Presentar la estrategia completa de búsqueda electrónica en, al menos, una base de datos, incluyendo los límites utilizados, de tal forma que pueda ser reproducible |
| Selección de los estudios | 9 | Especificar el proceso de selección de los estudios (por ej., el cribado y la elegibilidad incluidos en la revisión sistemática y, cuando sea pertinente, incluidos en el metaanálisis) |
| Proceso de extracción de datos | 10 | Describir los métodos para la extracción de datos de las publicaciones (por ej., formularios pilotado, por duplicado y de forma independiente) y cualquier proceso para obtener y confirmar datos por parte de los investigadores |
| Lista de datos | 11 | Listar y definir todas las variables para las que se buscaron datos (por ej., PICOS, fuente de financiación) y cualquier asunción y simplificación que se hayan hecho. |
| Riesgo de sesgo en los estudios individuales | 12 | Describir los métodos utilizados para evaluar el riesgo de sesgo en los estudios individuales (especificar si se realizó |

¹⁴ Acrónimo en inglés, formado por P de participantes, I de intervenciones, C, de comparaciones, O de resultados, y S de diseño del estudio (Urrutia y Bonfill, 2010).

| Dimensión/tema | Ítem | Contenido o descriptor |
|--|------|---|
| Riesgo de sesgo en los estudios individuales | 12 | al nivel de los estudios o de los resultados) y cómo esta información se ha utilizado en la síntesis de datos |
| Medidas de resumen | 13 | Especificar las principales medidas de resumen (por ej., razón de riesgos o diferencia de medias) |
| Síntesis de resultados | 14 | Describir los métodos para manejar los datos y combinar resultados de los estudios, cuando esto es posible, incluyendo medidas de consistencia (por ej., ítem 2) para cada meta análisis |
| Riesgo de sesgo entre los estudios | 15 | Especificar cualquier evaluación del riesgo de sesgo que pueda afectar la evidencia acumulativa (por ej., sesgo de publicación o comunicación selectiva) |
| Análisis adicionales | 16 | Describir los métodos adicionales de análisis (por ej., análisis de sensibilidad o de subgrupos, meta regresión), en el caso de que se hiciera, indicar cuáles fueron pre especificados |
| Resultados | | |
| Selección de estudios | 17 | Facilitar el número de estudios cribados, evaluados para su elegibilidad e incluidos en la revisión, y detallar las razones para su exclusión en cada etapa, idealmente mediante un diagrama de flujo |
| Características de los estudios | 18 | Para cada estudio presentar las características para las que se extrajeron los datos (por ej., tamaño, PICOS y duración del seguimiento) y proporcionar las citas bibliográficas. |
| Riesgo de sesgo en los estudios | 19 | Presentar datos sobre el riesgo de sesgo en cada estudio y, si está disponible, cualquier evaluación del sesgo en los resultados (ver ítem 12) |
| Resultados de los estudios individuales | 20 | Para cada resultado considerado en cada estudio (beneficios o daños), presentar: a) el dato resumen para cada grupo de intervención y b) la estimación del efecto con su intervalo de confianza, idealmente de forma gráfica mediante un diagrama de bosque (forest plot) |
| Síntesis de los resultados | 21 | Presentar los resultados de todos los meta análisis realizados, incluyendo los intervalos de confianza y las medidas de consistencia |
| Riesgo de sesgo entre los estudios | 22 | Presentar los resultados de cualquier evaluación del riesgo de sesgo entre los estudios (ver ítem 15) |
| Análisis adicionales | 23 | Facilitar los resultados de cualquier análisis adicional, en el caso de que se hayan realizado (por ej., análisis de sensibilidad o de subgrupos, meta regresión [ver ítem 16]) |

| Dimensión/tema | Ítem | Contenido o descriptor |
|-------------------------|------|--|
| Discusión | | |
| Resumen de la evidencia | 24 | Resumir los hallazgos principales, incluyendo la fortaleza de las evidencias para cada resultado principal; considerar su relevancia para grupos clave (por ej., proveedores de cuidados, usuarios y decisores en salud) |
| Limitaciones | 25 | Discutir las limitaciones de los estudios y de los resultados (por ej., riesgo de sesgo) y de la revisión (por ej., obtención incompleta de los estudios identificados o comunicación selectiva) |
| Conclusiones | 26 | Proporcionar una interpretación general de los resultados en el contexto de otras evidencias, así como las implicaciones para la futura investigación |
| Fondos | | |
| Financiamiento | 27 | Describir las fuentes de financiamiento de la revisión sistemática y otro tipo de apoyos (por ej., aporte de los datos), así como el rol de los financiadores en la revisión sistemática |

PRISMA, al igual que otras directrices o escalas presentadas requiere el acompañamiento de un diagrama de flujo que puede ser consultado en Urrutúa y Bonfill (2010) o el sitio *Equator*.

Lo hasta aquí expuesto intenta ser una aproximación inicial a los procesos de medición de la calidad científica, considerando que esta tarea es una fuente de información a través de la cual se puede perfeccionar el trabajo del investigador, ampliando y en su caso corrigiendo los hallazgos o resultados.

Conclusiones

La evaluación de la calidad científica debe ser una actividad prioritaria para el mejoramiento de los procesos de investigación, las aportaciones de esta actividad recaen básicamente en las diferentes etapas o elementos integrantes del método, desde el planteamiento de preguntas de investigación, medidas, levantamiento y análisis de datos, hasta la elaboración del informe de resultados.

En este sentido, el desarrollo de protocolos, directrices o lineamientos, ha sido una tarea atendida de manera desigual según el campo de estudio, de tal manera en la Medicina han surgido mayores iniciativas que en aquellas ciencias encargados de estudiar el comportamiento humano, sin embargo, gran parte de estos desarrollados pueden ser aplicados independientemente de la disciplina o área del conocimiento, al final la metodología, salvo ciertas excepciones, es inherente

a toda disciplina científica, por lo que un estudio experimental debe cumplir con ciertas condiciones básicas independientemente de los campos disciplinares, como la existencia de un tratamiento o la condición control.

Por ello las herramientas aquí expuestas deben ser vistas bajo esa condición, la de su amplia aplicabilidad, quedando como criterio de elección las características de cada una o el tipo de estudio al que van dirigidas más que su origen, teniendo en cuenta esto, en la tabla once se sintetizan algunas de las características más notables por protocolo, escala o directriz, para que el lector tenga la posibilidad de realizar comparaciones que guíen una futura elección.

Tabla 11. Resumen de las herramientas para evaluar la calidad científica

| Nombre | Objetivo | N.D. | Dimensiones/secciones | N.I. | Tipo de estudios |
|-----------------|------------------------------|------|---|------|---|
| Escala de Jadad | Calidad general | 1 | Método | 5 | Experimentales y no experimentales |
| CONSORT | Calidad general y específica | 6 | Título y resumen; Introducción; Métodos; Resultados; Discusión; Otra información | 25 | Ensayos clínicos aleatorios, estudios experimentales |
| STROBE | Calidad general y específica | 6 | Título y resumen; Introducción; Métodos; Resultados; Discusión; Otra información | 22 | Estudios observacionales en general, epidemiológicos. Transversales, longitudinales, caso-control |
| TREND | Calidad general y específica | 5 | Título y resumen; Introducción; Métodos; Resultados; Discusión | 22 | Estudios no aleatorios con intervención o tratamiento |
| STARD | Calidad general y específica | 7 | Título o resumen; Resumen; Introducción; Métodos; Resultados; Discusión; Otra información | 30 | Estudios o reportes sobre diagnóstico médico o en salud |
| CARE | Calidad general y específica | 13 | Título; Palabras clave; Resumen; Introducción; Información del paciente; Hallazgos clínicos; Línea de tiempo; Evaluación diagnóstica; Intervención terapéutica; | 30 | Estudios o reportes de caso no cualitativos |

| Nombre | Objetivo | N.D. | Dimensiones/secciones | N.I. | Tipo de estudios |
|-----------|------------------------------|------|---|------|--|
| CARE | Calidad general y específica | 13 | Seguimiento y resultados; Discusión; Perspectiva del paciente; Consentimiento informado. | 30 | Estudios o reportes de caso no cualitativos |
| SRQR | Calidad general y específica | 6 | Título y resumen; Introducción; Métodos; Resultados; Discusión; Otra información | 21 | Estudios cualitativos (etnografía, teoría fundamentada, estudio de caso, fenomenología, investigación narrativa, u otro) |
| COREQ | Calidad general y específica | 3 | Equipo de investigación y reflexividad; Diseño del estudio; Análisis y resultados | 32 | Estudios cualitativos con entrevistas y grupo focal |
| DICC'sPCC | Calidad general y específica | 12 | Título; Resumen; Palabras clave; Introducción; Identificación del paciente y motivos de remisión; Estrategias de evaluación; Formulación de casos clínicos; Tratamiento; Diseño del estudio; Análisis de datos; Efectividad y eficiencia de la intervención; Discusión. | 50 | Estudios de caso clínico con enfoque o tratamiento conductual |
| PRISMA | Calidad general y específica | 7 | Título; Resumen; Introducción; Métodos; Resultados; Discusión; Fondos | 27 | Revisiones sistemáticas y metaanálisis |

N.D. = Número de dimensiones N.I. = Número de ítems

Las iniciativas revisadas no agotan el amplio campo de la evaluación científica, prueba de ello, son la diversidad de escalas, directrices, lineamientos, o listas de verificación que surgen desde distintas disciplinas o metodologías, por ejemplo la

directriz ARRIVE¹⁵ (Animal Research: Reporting of In Vivo Experiments) en la investigación experimental con sujetos animales (Kilkenny, Browne, Cuthill, Emerson y Altman, 2010), la lista de verificación SPIRIT (Standard Protocol Items for Clinical Trials) (Chan, Tetzlaff, Altman, Laupacis, Gøtzsche, Krleža-Jerić, Hróbjartsson et al., 2013) en la regulación del protocolo de ensayos clínicos, GRADE (Grading of Recommendations, Assessment, Development and Evaluation) (Alonso-Coello, Rigau, Sanabria, Plazac, Miravittlesd y Martínez, 2013) o SQUIRE (Standars for QUALity Improvement Reporting Excellence) (Davidoff, Batalden, Stevens, Ogrinc, y Mooney, 2008) en la evaluación de las intervenciones con el propósito de mejorar la calidad de la atención en salud.¹⁶ De tal manera, aún falta trabajo por hacer, particularmente en el mejoramiento de los estudios en comportamiento.

Referencias

- Alonso-Coello, P., Rigau, D., Sanabria, A. J., Plazac, V., Miravittlesd, M. y Martínez, L. (2013). “Calidad y fuerza: el sistema GRADE para la formulación de recomendaciones en las guías de práctica clínica”. *Archivos de Bronconeumología*. 49 (6), pp. 261-267.
- Asociación de Psicología Americana (2002). *Manual de estilo de publicaciones de la American Psychological Association*. México: Manual Moderno.
- Bono, R. y Arnau, J. (1995). “Consideraciones generales en torno a los estudios de potencia”. *Anales de Psicología*. 11 (2), pp. 193-202.
- Bossuyt, P., Reitsma, Jh., Bruns, D., Gatsonis, C., Glasziou, P., Irving, L., et al. (2003). “The STARD Statement for Reporting Studies of Diagnostic Accuracy: Explanation and Elaboration”. *Annals of Internal Medicine*, 138 (1), W1-W12
- Cascaes da S., F., Valdivia A., B. A., da Rosa I., R., Barbosa G. F., P. J. y da Silva, R. (2013). “Escala y listas de evaluación de la calidad de estudios científicos”. *Revista Cubana de Información en Ciencias de la Salud*. 24 (3), pp. 295-312.
- Chan, A., Tetzlaff, J. M., Altman, D. G., Laupacis, A., Gøtzsche, P. C., Krleža-Jerić, K., Hróbjartsson, A., Mann, H., Dickersin, K., Berlin, J. A., Doré, C. J., Parulekar, W. R., Summerskill, W. S. M., Groves, T., Schulz, K. F., Sox, H. C., Rockhold, F. W., Rennie, D. y Moher, D. (2013). “SPIRIT 2013 Statement: Defining Standard Protocol Items for Clinical Trials”. *Annals of Internal Medicine*. 158 (3), pp. 200-207.

¹⁵ Para mayor información sobre ARRIVE, SPIRIT, GRADE o SQUIRE puede consultar en www.equator-network.org

¹⁶ Un estudio que presenta una revisión más extensa que la efectuada en el presente trabajo es el realizado por Cascaes y cols. (2013), donde se revisan 25 dispositivos de evaluación, por lo que se recomienda al lector interesado la consulta de la misma y el sitio *Equator* (www.equator-network.org).

- Davidoff, F., Batalden, P., Stevens, D., Ogrinc, G., y Mooney, S. (2008). "Publication guidelines for quality improvement in health care: evolution of the SQUIRE project". *Qual Saf Health Care*. 17 (Sup. 1), pp. i3-i9. Recuperado el 9 de septiembre de 2015, doi:10.1136/qshc.2008.029066
- De Jesus Loureiro, L. M. y Henriques Gameiro, M. G. (2011). "Interpretação crítica dos resultados estatísticos: para lá da significância estatística". *Revista de Enfermagem Referência*. III (3), pp. 151-162. Recuperado de <http://www.scielo.mec.pt/pdf/ref/vserIIIIn3/serIIIIn3a16.pdf>
- Delgado R., M. y Palma P., S. (2006). "Aportaciones de la revisión sistemática y del metaanálisis a la salud pública". *Revista Española de Salud Pública*. 80 (5), pp. 483-489.
- Des Jarlais, D., Lyles, C., Crepaz, N. y TREND Group (2004). "Improving the Reporting Quality of Nonrandomized Evaluations of Behavioral and Public Health Interventions: The TREND Statement". *American Journal of Public Health*. 94 (3), 361-366.
- Daset, L. R. y Cracco, C. (2013). "Psicología Basada en la Evidencia: algunas cuestiones básicas y una aproximación a través de una revisión bibliográfica sistemática". *Ciencias Psicológicas*. VII (2), pp. 209-220.
- Elm, E., Altman, D., Egger, M., Pocock, S., Gøtzsche, P., y Vandenbroucke, J. (2007). "The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for reporting observational studies". *Preventive Medicine*, 45, 247-251.
- Gagniera, J.J., Kienlec, G., Altman, D.G., Moher, D., Sox, H., Riley, D. and the CARE Group (2014). "The CARE guidelines: consensus-based clinical case report guideline development". *Journal of Clinical Epidemiology*. 67. pp. 46-51. doi: 10.1016/j.jclinepi.2013.08.003
- García C., D. y Aboitiz C., J. (2012). "Efectividad del entrenamiento de los músculos del suelo pélvico en la incontinencia urinaria: revisión sistemática". *Fisioterapia*. 34 (2), pp. 87-95.
- Glass, G. V. (1976). "Primary, secondary, and meta-analysis of research". *Educational Researcher*. 5 (10), pp. 3-8.
- Jovell, A. J. y Navarro – Rubio, M. D. (1995). "Evaluación de la evidencia científica". *Medicina Clínica*. 105 (19). pp. 740-743.
- Kilkenny, C., Browne, W. J., Cuthill, I. C., Emerson, M., y Altman, D. G. (2010). "Improving Bioscience Research Reporting: The ARRIVE Guidelines for Reporting Animal Research". *PLoS Biol*, 8 (6). e1000412. doi: 10.1371/journal.pbio.1000412
- Kline, R. B. (2004). *Beyond significance testing: Returning data analysis methods in behavioral research*. (pp.247-271). Washington, D.C., US: American Psychological Association.

- Leech, N. L. y Onwuegbuzie, A. J. (Winter, 2010). "Guidelines for Conducting and Reporting Mixed Research in the Field of Counseling and Beyond". *Journal of Counseling and Development*. 88, pp. 61-70.
- Lumbreras, B., Jarrín, I. y Hernández, I. (2006). "Valoración de la metodología de la investigación en pruebas genética, molecular y proteómica". *Gaceta Sanitaria*. 20 (5), 368-374.
- Luna, C. M., Jolly, E. C., Bevilacqua, C., y Jäger, N. (2011). "Conflicto de Intereses Editoriales". *Rev Am Med Resp*. 11 (3), pp. 134-140.
- MacPherson, H., White, A., Cummings, M., Jobst, K., Rose, K. y Niemtow, R. (2002). "Standards for Reporting Interventions in Controlled Trials of Acupuncture: The STRICTA Recommendations". *Acupuncture in Medicine*. 20 (1), pp. 22-25. Descargado de <http://aim.bmj.com/>, en Septiembre 10 de 2015.
- Moher, D., Cook, D. J., Eastwood, S., Olkin, I., Rennie, D., Stroup, D. et al. (2000). "Mejora de la calidad de los informes de los metaanálisis de ensayos clínicos controlados: el acuerdo QUOROM". *Revista Española de Salud Pública*. 74 (2), pp. 107-118.
- Moher, D., Schulz, K., y Altman, D. y Grupo CONSORT. (2002). "La declaración CONSORT: recomendaciones revisadas para mejorar la calidad de los informes de ensayos aleatorizados de grupos paralelos." *Revista de Sanidad Militar*. 56 (1), 23-28.
- Montero, I. y León, O. G. (2005). "Sistema de clasificación del método en los informes de investigación en Psicología". *International Journal of Clinical and Health Psychology*. Vol. 5 (1), pp. 115-127.
- O'Brien, B., Harris, I. B., Beckman, Th. J., Reed, D. A. y Cook, D. A. (2014). "Standards for Reporting Qualitative Research: A Synthesis of Recommendations". *Academic Medicine*. 89 (9), pp. 1245-1251. doi: 10.1097/ACM.0000000000000388
- Ramos-Álvarez, M. M. y Catena, A. (2004). "Normas para la elaboración y revisión de artículos originales experimentales en Ciencias del Comportamiento". *International Journal of Clinical and Health Psychology*. 4 (1), pp. 173-189.
- Rozemblum, C.; Unzurrunzaga, C.; Banzato, G. y Pucacco, C. (2015). "Calidad editorial y calidad científica en los parámetros para inclusión de revistas científicas en bases de datos en Acceso Abierto y comerciales". *Palabra Clave*, Vol. 4 (2), pp. 64-80. Recuperado de <https://www.palabraclave.fahce.unlp.edu.ar/article/view/PCv4n2a0>
- Sanduvete, S. (2008). *Innovaciones metodológicas en la evaluación de la formación continua*. Tesis de Doctorado en Aprendizaje y Cognición. Departamento de Psicología Experimental, Universidad de Sevilla.
- Tong, A., Sainsbury, P. y Craig, J. (2007). "Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups". *International Journal for Quality in Health Care*. 19 (6), pp. 349-357.

- Turpin, D. (2005). "CONSORT and QUOROM guidelines for reporting randomized clinical trials and systematic reviews". *American Journal of Orthodontics and Dentofacial Orthopedics*. 128 (6), 681-685.
- Urrutia, G. y Bonfill, X. (2010). Declaración PRISMA: una propuesta para mejorar la publicación de revisiones sistemáticas y metaanálisis. *Medicina Clínica*. 135 (11), pp. 507-511. Recuperado de <http://www.elsevier.es/122/02/2011>
- _____(2013). "La declaración PRISMA: un paso adelante en la mejora de las publicaciones de la Revista Española de Salud Pública". *Revista Española de Salud Pública*. 87 (2), pp. 99-102.
- Virués-Ortega, J y Moreno-Rodríguez, R (2008). "Guidelines for clinical case reports in behavioral clinical Psychology". *International Journal of Clinical and Health Psychology*. Vol. 8. N° 3, pp. 173-189. Recuperado el 9 de septiembre de 2015, de www.equator-network.org

Capítulo 5

Revisiones sistemáticas: su definición, directrices y dos casos prácticos en el campo de las ciencias sociales

*Mendoza de los Santos, Oscar Eliezer
Castelán-Félix, Jocelyn Pamela*

Resumen

Las revisiones sistemáticas son una herramienta fundamental para la identificación, evaluación, descripción y síntesis de los hallazgos obtenidos en distintas investigaciones empíricas sobre un objeto de estudio determinado. Si bien han sido ampliamente utilizadas en el campo de las ciencias de la salud, su uso en las ciencias sociales es cada vez más frecuente, razón por la cual resulta importante contar con una definición clara de las mismas, así como directrices para su ejecución en dicho campo. El objetivo del presente capítulo es caracterizar la revisión sistemática como un tipo de investigación cuyo objetivo sea la descripción y/o síntesis del conocimiento producto de la investigación científica primaria. Para ello se ubica el papel de las revisiones sistemáticas en el plano de la investigación científica y se compara con las revisiones tradicionales de la literatura, presentando una clasificación de las revisiones sistemáticas, para proveer a los interesados una visión amplia del tema, pero a la vez clara, de qué puede considerarse una revisión de este tipo y en qué variedad de situaciones de investigación pueden aplicarse. Se exponen también una serie de directrices generales para realizar este tipo de revisiones, y herramientas para evaluar su calidad. Finalmente, por medio de dos casos de investigaciones en las cuales se hizo uso de los métodos propios de una revisión sistemática se ilustra la ejecución de este tipo de revisiones. Se concluye que, si bien en el presente capítulo se han expuesto diversos ejemplos desde las ciencias sociales, exponer casos propios de otras disciplinas ayudará a esclarecer el uso de las revisiones sistemáticas en contextos diversos. Se sugiere, además, elaborar guías de revisiones sistemáticas en campos particulares, asumiendo que en cada uno de estos se encuentran retos específicos para la elaboración de revisiones.

Palabras clave: revisiones sistemáticas, investigación primaria, investigación secundaria, síntesis de la evidencia científica, ciencias sociales.

Introducción

Las revisiones sistemáticas se han constituido como una herramienta fundamental para el estudio de un cuerpo de conocimiento cada vez más amplio, coadyuvando a la consolidación de prácticas científicas y profesionales fundadas en evidencias

sólidas, tal como ha ocurrido con las ciencias de la salud. De acuerdo a Petticrew y Roberts (2006), los primeros intentos de revisiones sistemáticas pueden rastrearse en trabajos pertenecientes al área de psicología, como por ejemplo la revisión realizada por Hebert Nichols sobre experimentos ejecutados en torno a la Ley de Weber, a finales del siglo de XIX, así como la revisión de Lazarus, a mediados del siglo XX, en relación a los efectos del estrés sobre el desempeño de los individuos. No deben olvidarse, además, los esfuerzos de Glass por evaluar la efectividad de la psicoterapia a través del meta-análisis, término que acuñara para referir al análisis estadístico de otros análisis con la finalidad de integrar los resultados de los mismos (Glass, 1976; Smith y Glass, 1977; Smith, Glass y Miller, 1980).

El presente capítulo tiene por objetivo general la caracterización de la revisión sistemática como un tipo de investigación cuyo objetivo sea la descripción y/o síntesis del conocimiento producto de la investigación científica primaria. Al revisar el papel de las revisiones sistemáticas en el plano de la investigación científica comparadas con las revisiones tradicionales de la literatura al clasificarlas. Acto seguido, se presentará una clasificación de las revisiones sistemáticas, con la finalidad de proveer a los interesados en el tema una visión amplia, pero a la vez clara, de qué puede considerarse una revisión de este tipo y en qué variedad de situaciones de investigación pueden aplicarse. Posteriormente, se expondrá también una serie de directrices generales para realizar este tipo de revisiones, y herramientas para evaluar su calidad. Finalmente, se expondrán dos casos de investigaciones en las cuales se hizo uso de los métodos propios de una revisión sistemática. El primero de dichos casos corresponde a un análisis comparativo de la regulación jurídica en materia de medios de comunicación y el ejercicio de la psicología en México. Por otra parte, en el segundo caso se aborda el análisis de la calidad científica de las investigaciones realizadas sobre medios de comunicación.

El papel de las Revisiones Sistemáticas en la ciencia

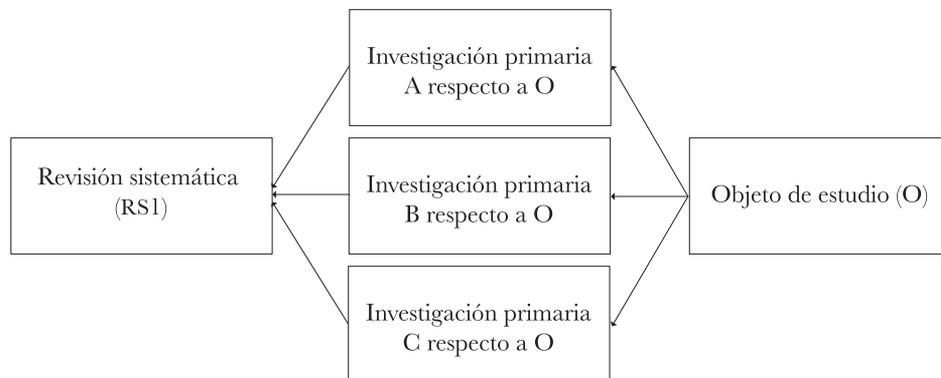
Generalmente, se entiende por revisión sistemática (RS) un tipo de investigación que tiene como fuente de datos la literatura científica existente sobre un determinado tema y cuyos objetivos son identificar, evaluar, describir y, en muchos de los casos, sintetizar, por medio de métodos estrictamente especificados e imparcialmente aplicados, los resultados de las investigaciones expuestas en la literatura en cuestión (Letelier, Manríquez y Rada, 2005; Sampaio y Mancini, 2007; Ferreira, Urrútia y Alonso, 2011; Rodríguez, Zafra, Quintero, 2015; Moreno, Muñoz, Cuellar, Domancic y Villanueva, 2018).

Es frecuente que las RS sean incluidas dentro del rótulo general de investigación secundaria, término utilizado para distinguir a este tipo de

investigaciones de aquellas otras en las que los datos son obtenidos del estudio directo del fenómeno o problema de interés, es decir las investigaciones primarias o empíricas. (Glass, 1976; Torgerson, 2003). Así, el término investigación secundaria no debería tomarse en un sentido despectivo, sino más bien, como un rótulo que permite diferenciar entre formas de investigación legítimas que se aproximan de manera distinta a un mismo objeto de estudio; en este sentido, la Figura 1 ilustra la forma en que una revisión sistemática provee información sobre un determinado objeto de estudio, a través de la recopilación, análisis y síntesis de los hallazgos obtenidos en distintas investigaciones primarias respecto a dicho objeto.

Ahora bien, las RS ocupan actualmente el interés de una gran cantidad de investigadores en áreas como la medicina, en la cual su uso se ha extendido ampliamente (Salvador, Marco y Arquero, 2018), pero también en la política y las ciencias sociales en general, por su potencial para proveer respuestas a preguntas sobre problemáticas que requieren una solución práctica, por lo que, también, ha crecido el interés por desarrollar y refinar métodos para la construcción de RS de mayor calidad.

Figura 1. Esquema de las investigaciones de acuerdo a su acercamiento a un determinado objeto de estudio. Mientras que las investigaciones primarias (A, B y C) se aproximan directamente al Objeto de Estudio (O) a través de la recolección y análisis de datos empíricos, la Revisión Sistemática (RS1) estudia los resultados obtenidos en A, B y C, y, por medio de la caracterización y/o síntesis de estos, logra un conocimiento integral en relación a O.



Fuente: Elaboración propia.

Decir, sin embargo, que durante mucho tiempo las revisiones de la literatura científica han sido llevadas a cabo de formas distintas a lo que actualmente se considera una revisión sistemática (la importancia de una correcta revisión de

la literatura se justifica en Moncada-Hernández, 2014, por otro lado, para una detallada clasificación de los distintos tipos de revisiones de literatura véase Grant y Booth, 2009). En este último sentido, Gough, Oliver y Thomas (2017) señalan que las revisiones tradicionales de la literatura científica se han caracterizado por la exposición crítica de los hallazgos de una serie de trabajos de investigación, pero sin tomar en consideración criterios específicos al momento de elegir cuales de estos trabajos serán incluidos o no en la revisión (criterios de inclusión y exclusión). Frecuentemente, en este tipo de revisiones se expone información sobre los aspectos de interés para el revisor siguiendo un formato narrativo, por lo que suele denominárseles genéricamente como revisiones narrativas (Letelier, Manríquez y Rada, 2005).

Al depender, en gran medida, de la opinión de los autores que llevan a cabo la revisión, las revisiones narrativas presentan una serie de desventajas enunciadas a continuación (Torgerson, 2003; Urrea y Barría, 2010):

- Al no ser contemplados criterios de selección explícitos, pueden no incluirse artículos de potencial relevancia para el tema en cuestión.
- Una muestra de artículos sesgada, puede llevar a interpretaciones erróneas de la evidencia analizada.
- La misma carencia de criterios de selección explícitos, supone dificultades para aquellos investigadores que deseen replicar con precisión la revisión.

Precisamente, el interés por llevar a cabo revisiones sistematizadas del conocimiento responde a las dificultades y desventajas propias de la realización de revisiones narrativas (subjetivas) de la literatura científica, pues el rápido incremento de las publicaciones de trabajos empíricos en los múltiples sectores de la ciencia, desde el siglo XX, volvió imperioso el generar alternativas que permitieran a los profesionales llegar a conclusiones fiables respecto a las evidencias de las investigaciones, optimizando así la toma de decisiones en determinados rubros (Sánchez-Meca, 2010).

Así, al contar con mayor fuerza de evidencia en relación a una determinada cuestión, las RS se han posicionado en la punta de la pirámide de jerarquía de la evidencia científica y han tomado un papel protagónico en lo que concierne a la llamada Práctica Basada en Evidencias (PBE), cuyo principal interés es proporcionar información confiable sobre la efectividad de distintos tratamientos o intervenciones, que pudiera servir para la toma de decisiones clínicas o sociales, generalmente sintetizando la información de múltiples ensayos aleatorizados (Letelier y Moore, 2003; Solomon, Cavanaugh y Draine 2009; Vidal, Oramas y Borroto, 2015). No obstante lo anterior, es importante destacar dos aspectos de suma importancia:

Primero, si bien en las RS se prefiere la inclusión de intervenciones aleatorizadas (por representar éstas el ideal científico en cuanto a establecer relaciones causales se refiere), en ciencias sociales este tipo de estudios no siempre es factible, sea por limitaciones éticas, prácticas o por la misma naturaleza del objeto de estudio, por lo que su presencia puede ser, en muchas ocasiones, escasa (Solomon, Cavanaugh y Draine, 2009). A este respecto hay que señalar que, si bien las intervenciones aleatorizadas, conocidas también como ensayos o experimentos aleatorios, constituyen la regla de oro en la investigación científica, no son la única opción para el estudio de los fenómenos psicosociales; se ha señalado frecuentemente en la literatura que otros diseños, como los cuasi-experimentos, los experimentos de campo y la investigación no experimental, son herramientas de utilidad cuando se vuelve inviable la consecución de experimentos aleatorizados (Katz, 1953; French, 1953; Campbell y Laurence, 1968; Campbell y Stanley, 1973; Kerlinger y Lee, 2002). Por ello, actualmente, ha incrementado el interés por llevar a cabo RS que contemplen no solamente diseños experimentales aleatorizados, sino también otro tipo de diseños, incluyendo investigación cualitativa (Gough y Thomas, 2017; Gough, 2015; Pearson, White, Bath-Hextall, Salmond, Apostolo y Kirkpatrick, 2015).

Como segundo punto, es de suma importancia destacar que las RS, per se, no aseguran la síntesis de la mejor evidencia posible; la calidad de las revisiones se encuentra supeditada a la calidad de los estudios que son incluidos en ella (Pértega y Pita, 2005), por lo que decir que las RS proporcionan el menor sesgo en los resultados entorno a una pregunta de investigación (Letelier y Moore, 2003) puede no ser necesariamente cierto. En este sentido, es imperioso observar la estrecha relación entre evaluación de la calidad científica y la elaboración de revisiones sistemáticas, siendo la primera un prerrequisito indispensable para la consecución de la segunda, que, de pasarse por alto, aumentaría el riesgo de sesgo en las conclusiones obtenidas. Sobre este punto se volverá cuando al exponerse las directrices básicas para el desarrollo de RS.

Clasificación de las Revisiones Sistemáticas

Ahora bien, al igual que ocurre con la investigación primaria, es de suma importancia distinguir entre las distintas variedades de RS. En relación con este punto, es posible encontrar algunos esfuerzos de clasificación de estos estudios. Así por ejemplo, Letelier, Manríquez y Rada (2005) distinguen dos modalidades de estos estudios: aquellas RS que presentan la información de forma puramente descriptiva (RS cualitativas) y aquellas otras que, por medio de técnicas estadísticas, integran los resultados cuantitativos de distintas investigaciones sobre un tema común (RS cuantitativas).

Dicha clasificación, aunque relevante por distinguir entre descripción y síntesis de estudios científicos, resulta limitada por dos razones:

- 1) Es errado referirse a las RS descriptivas como cualitativas, pues dichas descripciones pueden darse en términos cuantitativos.
- 2) Si bien frecuentemente se opta por síntesis estadísticas de artículos cuantitativos, no es ésta la única forma de integrar los resultados de investigaciones, existiendo también integraciones de estudios cualitativos.

Gough y Thomas (2017) ofrecen una amplia y profunda discusión sobre las revisiones sistemáticas, considerando tanto aspectos estructurales como epistemológicos, logrando así una clasificación de dichos diseños. Estos autores sugieren que existen múltiples formas en que las revisiones pueden ser llevadas a cabo, manteniendo, sin embargo, el criterio de sistematicidad que caracteriza a este tipo de investigaciones. El valor de esta clasificación radica en su capacidad para aprehender de forma crítica y flexible las variadas formas de las RS, proporcionando un amplio abanico de posibilidades a los interesados en diseñar y ejecutar revisiones, así como un marco conceptual para la identificación de las mismas.

La propuesta de Gough y Thomas (2017) parte de la delimitación de una serie de ejes sobre los cuales las RS pueden ser entendidas, a saber: 1) las diferencias en extensión, detalle, aproximación y epistemología; 2) la estructura de las revisiones sistemáticas y, por último: 3) la conceptualización de las revisiones.

El primero de los ejes hace referencia al hecho de que ninguna RS se lleva a cabo de forma ajena a alguna determinada perspectiva que guía los objetivos de investigación, misma que influye en la cantidad de trabajo respecto a un problema de investigación que la revisión pretende cubrir; así como su alcance y nivel de detalle, la cantidad de tiempo invertido y los tipos de trabajos que se considerarán en la revisión (Gough y Thomas, 2017). Debe observarse que tal situación no es exclusiva de las RS y permea a toda la investigación, tanto en ciencias naturales, como en las ciencias sociales. Los científicos conceptualizan sus objetos de estudio y dirigen sus investigaciones a través del filtro de sus creencias y valores ontológicos y epistemológicos (Mouton y Marais, 1990); por lo que, estrictamente hablando, las RS no son actividades libres de cierto componente subjetivo, aun cuando se trate de una alternativa sólidamente fundada para el estudio estricto de la literatura científica (Torgerson, 2003).

Dicho lo anterior, es importante mencionar que las RS pueden variar en amplitud (la extensión de trabajos abarcados) y profundidad (el grado de detalle con que se analizan y sintetizan los trabajos contemplados), aun cuando frecuentemente ha sido señalado que requieren partir de preguntas y objetivos particulares y

altamente específicos (Petticrew y Roberts, 2006); dicha especificidad puede variar en función de los intereses y posibilidades de los investigadores, siendo factible llevar a cabo RS de considerable amplitud y nivel de detalle, pero siempre tomando en cuenta que un objetivo amplio conllevará una mayor carga de trabajo en cada etapa de la revisión (Gough y Thomas, 2017; Sobrido y Rumbo-Prieto, 2018).

En lo que respecta a su estructura (segundo eje), las RS pueden clasificarse de forma general, en aquellas en las que se busca determinar qué se ha investigado en un determinado campo de estudio; las que tienen por objetivo sintetizar la información de las investigaciones detectadas; y aquellas otras en las que se pretende revisar revisiones previas (Gough y Thomas, 2017; Gough, Oliver y Thomas, 2017). Dentro de las RS en las que sólo se pretende detectar la investigación que se ha producido en un determinado campo de estudio, se encuentran los mapeos sistemáticos (MS). Este tipo de revisiones tiene la finalidad de describir el cuerpo de investigaciones realizadas en algún área de conocimiento, partiendo de una pregunta de investigación considerablemente amplia, de tal suerte que se obtenga una visión de las temáticas, así como perspectivas teóricas y metodológicas, presentes en el ámbito de conocimiento estudiado. Pese a la amplitud que supone la empresa de realizar MS, estos siguen los mismos métodos rigurosos, objetivos y transparentes que caracterizan a toda RS de calidad (James, Randall y Haddaway, 2016).

Generalmente, los MS son el paso previo a las síntesis sistemáticas de la literatura científica (SS) que tienen el objetivo de evaluar un conjunto de investigaciones y ordenar sus resultados, sea para probar teorías o hipótesis (por ejemplo, respecto a la efectividad de un tratamiento), o bien para construirlas (Oliver y Sutcliffe, 2017). Cabe destacar, que las formas para lograr una síntesis de información científica son variadas y su elección dependerá de las características de los trabajos a integrar. Cuenta con una amplia presencia en este rubro el meta-análisis, término acuñado por Glass (1976) para referirse a un método de cariz eminentemente cuantitativo, cuyo principal objetivo es lograr estimar el *efecto real* de un determinado tratamiento (variable causal), a través de la integración estadística de múltiples estudios donde tal efecto es estudiado. En el caso de los estudios cualitativos la meta-etnografía, desarrollada por Noblit y Dwight (1988) ha resultado ser una valiosa aproximación para la síntesis de investigaciones etnográficas, las cuales indagan sobre creencias y percepciones de individuos (France, Wells, Lang y Williams, 2016).

Finalmente, existen RS cuyo objetivo principal consiste en caracterizar y/o sintetizar revisiones previamente realizadas. En este sentido, dos categorías básicas son las Revisiones de Revisiones (RR) y las Meta-Epidemiologías (ME), las cuales, valiéndose de la información proporcionada por otras RS, se enfocan en construir mapeos y/o síntesis de RS realizadas previamente. En la Tabla 1 se ofrece una

descripción breve de las RS según su estructura y se ofrecen ejemplos recientes (sin el afán de ser exhaustivo) desarrollados en las ciencias del comportamiento.

Respecto al tercer eje, cabe decir que las revisiones pueden conceptualizarse de dos formas básicas, según si tienen por objetivo la verificación de hipótesis o teorías, en cuyo caso se conocen como RS por agregación. Este tipo de revisiones suelen incluir estudios más homogéneos conceptual y metodológicamente. Ahora bien, si la RS tiene por objetivo contribuir a generar nuevas teorías o explorar la relevancia de alguna teoría en un determinado campo de estudio, se tratará de una RS por configuración, la cual, a diferencia del tipo anterior, suele incluir estudios más heterogéneos.

Tabla 1. Clasificación de las revisiones sistemáticas de acuerdo a su estructura y objetivos, y ejemplos de dichos estudios

| Revisiones según su estructura | Descripción | Ejemplos de estudios en psicología y ciencias del comportamiento |
|----------------------------------|--|--|
| Mapeos sistemáticos | Son revisiones descriptivas que aportan información tal como: temas de interés, aproximaciones conceptuales y metodológicas, objetivos, etcétera, que se presentan en las investigaciones de dicho campo | Morgan, Morgan, Johansson y Ruud (2016) [<i>mapeo sobre los efectos de las TIC en el aprendizaje</i>]; Berg, Munthe-Kass y Ross (2016) [<i>mapeo sobre estudios relativos a la homonegatividad internalizada</i>] |
| Revisiones de estudios primarios | Síntesis | Silbaugh, et al., (2016) [<i>síntesis sistemática sobre intervenciones conductuales en el ámbito alimenticio</i>]; Milton y Mullan (2014) [<i>síntesis de estudios al respecto de la comunicación de diagnósticos de padecimientos mentales</i>] |
| | Métodos mixtos y revisiones mixtas del conocimiento | Se llevan a cabo cuando los estudios a ser revisados cuentan con distintas metodologías y datos de diferentes tipos Guillaumie, Boiral y Champagne (2016) [<i>revisión que analiza tanto intervenciones aleatorizadas como no aleatorizadas respecto a la efectividad del mindfulness</i>]; Sheals, Tombor, McNeill y Shahab (2016) [<i>revisión que incluye estudios cuantitativos y cualitativos sobre la relación entre actitudes al tabaquismo en profesionales de la salud y el abandono del tabaco por parte de pacientes mentales</i>] |

| Revisiones según su estructura | Descripción | Ejemplos de estudios en psicología y ciencias del comportamiento |
|--------------------------------|--------------------------|--|
| Meta-revisiones | Revisiones de revisiones | Ward, White y Druss (2015) [<i>revisión de las revisiones al respecto de las intervenciones en el estilo de vida de pacientes con factores de riesgo cardiovascular</i>]; Shepherd, Laurens, Matheson, Carr y Green (2012) [<i>meta-revisión sobre alteraciones cerebrales estructurales en pacientes con esquizofrenia</i>] |
| | Meta-epidemiología | En este caso, una serie de revisiones son analizadas con la finalidad de dar respuesta a una pregunta sobre las características esenciales de un determinado campo de investigación, como podrían serlo las implicaciones de tipo metodológico |

Fuente: Elaboración propia con información de Gough y Thomas (2017) y diversos estudios utilizados para ejemplificar cada tipo de RS.

Como podrá notarse, la clasificación previamente expuesta es lo bastante flexible como para dar cabida a distintos métodos, estrictos y objetivos, de revisión de la literatura que con justeza pueden considerarse como RS.

Directrices generales para el desarrollo de Revisiones Sistemáticas

Como ya se ha expuesto, las RS representan una vía para la ejecución de revisiones de la literatura con un considerable control de sesgos en los distintos procesos que componen dicha labor. Existe alguna directrices para el desarrollo de RS, entre las que destacan aquellas propuestas por la Colaboración Cochrane (Higgins y Green, 2011); así como protocolos cuya finalidad es facilitar la comprensión y evaluación de los métodos utilizados para realizar revisiones, tales como los lineamientos *Meta-analysis of observational studies in epidemiology* [MOOSE] (Stroup et al., 2000) y los *Preferred Reporting Items for Systematic review and Meta-Analysis Protocols* [PRISMA-P] (Moher et al., 2015).

Pese a esto, existe una escasez de directrices para el desarrollo de RS, hecho sugerido por la poca frecuencia con la que se reporta el uso de protocolos para la ejecución de RS (Moher et al., 2015).

Como ya se ha expresado en la sección previa, existen diversos tipos de RS (cada uno con sus particularidades metodológicas), la presente sección tiene como

finalidad presentar una serie de lineamientos generales para el desarrollo de RS, entendidas éstas en un amplio sentido. Para ello, es necesario reconocer que toda RS se compone, fundamentalmente, de las siguientes fases: 1) Estructuración de la pregunta que guía la revisión, 2) Identificación de los estudios relevantes, 3) Evaluación de la calidad de dichos estudios, 4) Análisis y/o síntesis de datos y 5) Presentación de resultados (Khan, Kunz, Kleijnen y Antes, 2003; Russell et al., 2009; Perestelo-Pérez, 2013).

Respecto al primer punto, es importante destacar que en el ámbito de la salud y el desarrollo e implementación de políticas públicas, es frecuente la presencia de RS del tipo *What Works* que parten de una pregunta respecto a la efectividad de algún determinado tratamiento o intervención (Petticrew y Roberts, 2006). Para el desarrollo adecuado de este tipo de preguntas resulta relevante considerar los componentes delimitados por el estándar PICO (Methley, Campbell, Chew-Graham, McNally y Cheraghi-Sohi, 2014; García-Perdomo, 2015) enlistados a continuación:

- *Population* (Población). Hace referencia a la población diana del estudio y sus características específicas.
- *Intervention* (Intervención). Se refiere al tratamiento (variable causal) cuya efectividad se está poniendo a prueba.
- *Comparison* (Comparación). Grupo de comparación que se encuentra en una condición experimental distinta, como podría ser un grupo control o un grupo expuesto a un tratamiento alternativo.
- *Outcome* (Resultado). Variable de respuesta, sobre la que se espera observar el efecto del tratamiento suministrado.

Si bien PICO es ampliamente utilizado en el contexto clínico, y como base en la construcción de preguntas para RS de estudios cuantitativos, es importante destacar que, recientemente, se han llevado a cabo esfuerzos por detectar elementos clave de preguntas de investigación para revisiones de estudios cualitativos. En este sentido, el esquema SPIDER proporciona lineamientos para el desarrollo de preguntas en RS que involucran investigaciones tanto cualitativas como mixtas (Cooke, Smith y Booth, 2012), al reconocerse las limitaciones inherentes al esquema PICO con respecto al uso de investigaciones cualitativas.

Desde el esquema SPIDER se contemplan los siguientes aspectos (Cooke, Smith y Booth, 2012):

- *Sample* (Muestra): se refiere a un conjunto pequeño de participantes, puesto que en la investigación cualitativa es más frecuente el trabajo con muestras pequeñas.

- *Phenomenon of Interest* (Fenómeno de interés): hace referencia al cómo y por qué de los comportamientos y experiencias individuales. En este sentido la implementación de una intervención o tratamiento no es siempre el eje central de una investigación cualitativa
- *Design* (Diseño): este punto refiere a los métodos de investigación utilizados en el estudio y al marco teórico que rige dicha investigación
- *Evaluation* (Evaluación): al igual que el *Outcome*, se refiere a las variables de respuesta observadas, las cuales sin embargo, en el caso de las investigaciones cualitativas pueden tratarse de constructos subjetivos y no observables.
- *Research Type* (Tipo de investigación): se refiere a tres paradigmas generales en los cuales se inscribe una investigación, a saber: cualitativo, cuantitativo o mixto.

Sea cual sea la estrategia utilizada para elaborar la pregunta de investigación, ésta última debe ser lo suficientemente clara y libre de ambigüedades, delimitando con suma especificidad las variables de interés involucradas. Este último hecho, claro está, no es exclusivo de las RS y es uno de los aspectos fundamentales de cualquier tipo de investigación científica.

La identificación de los estudios relevantes para una RS es una tarea que, aunque puede parecer sencilla vista superficialmente, es en realidad una labor que requiere suma atención al detalle. De forma muy general, puede decirse que la proliferación de artículos científicos ha vuelto necesario el perfeccionamiento de los repositorios que dan cabida a estos, y los investigadores que desean realizar RS de calidad deben estar al tanto de este hecho. En este sentido, existen diversas aproximaciones para la estructuración de estrategias de búsqueda e identificación de estudios, llegando a constituirse un genuino campo de desarrollos metodológicos en torno a las RS (véase, por ejemplo, Hausner, Guddat, Hermanns, Lampert y Waffenschmidt, 2015).

Una de las formas más comunes de proceder en la identificación de estudios es a través de la búsqueda en bases de datos, seleccionando un conjunto de palabras clave (o definidores) que cubran de la manera más comprehensiva posible la pregunta de investigación; dichos términos son introducidos en el buscador de la base correspondiente (v.g. *Web of Science*, *Scopus*, *DOAJ*, *Medline*, etcétera) y relacionados a través de un conjunto de operadores lógicos (véase Tabla 2); además, considerando los criterios de inclusión como el tipo de publicación (artículo completo, libro); fecha de publicación; lenguaje, así como características de la población/muestra de interés, es posible realizar búsquedas filtradas, más detalladas y precisas.

Tabla 2. Lista de operadores lógicos utilizados en algunos de los sistemas de búsqueda de repositorios científicos

| Tipo de operador | Operador | Función |
|--------------------------|----------|---|
| Operadores booleanos | AND | Expone los resultados que contengan todos los términos introducidos en el buscador |
| | OR | Expone los resultados que contengan al menos un término de los utilizados en la búsqueda |
| | NOT | Arroja los resultados que contengan un término específico A pero no contengan un término B |
| Operadores de proximidad | NEAR/x | Muestra resultados que contengan los términos de búsqueda, siempre que se encuentren a no más de “x” palabras de distancia el uno del otro, independientemente de su orden de aparición |
| | SAME | Muestra los resultados que contengan dos términos en un mismo campo de búsqueda (título, resumen, etcétera) |

Fuente: Elaboración propia.

A pesar de lo común de la aproximación previamente expuesta, es importante considerar que algunas de sus principales limitaciones radican en su dependencia en criterios subjetivos del investigador y de su limitado conocimiento de todos los posibles términos relacionados a un tema o fenómeno particular (Hausner, Guddat, Hermanns, Lampert y Waffenschmidt, 2015). Por otra parte, la búsqueda en grandes repositorios científicos no contempla aquella literatura gris (v.g. tesis de pre y posgrado) de la cual puede obtenerse información valiosa sobre el tema en cuestión, siendo, por lo tanto, necesario que el investigador considere el uso de diversas aproximaciones alternativas para la detección de este tipo de productos científicos, como podría ser la búsqueda en bibliotecas universitarias o la comunicación personal con investigadores de un determinado campo (Sánchez-Meca, 2010; Relevo, 2012).

Una vez que los artículos han sido descargados (o al menos sus títulos y resúmenes) de sus respectivos repositorios, es posible llevar a cabo un cribado o depuración del conjunto, cuyo objetivo es excluir de la muestra aquellos artículos que, a pesar de haber aparecido en las búsquedas realizadas en los repositorios, no cumplan con los criterios de selección previamente determinados. En este punto es importante no dejar de lado el papel que juega el recurso humano. El uso de distintos codificadores puede ser necesario ya sea por querer reducir el sesgo individual en la

elección de artículos, o bien por tratarse de RS de gran envergadura, en donde quizá es necesario el uso de varios sujetos para agilizar la búsqueda e identificación de artículos. La adecuada experiencia en el campo de estudio y/o correcta capacitación de los codificadores implicados en la fase de cribado es de suma importancia, para reducir la probabilidad de sesgo en la elección de los artículos. Resulta útil el uso de índices de acuerdo, como el índice Kappa, para determinar de manera objetiva y precisa el grado de consistencia entre codificadores.

Una vez identificados los artículos de interés, es importante llevar a cabo un filtro más, cuya finalidad es evaluar la calidad de los artículos detectados. Esta tarea es imperiosa, pues la validez de las conclusiones alcanzadas a través de la RS dependerá del grado de sesgo de las investigaciones incluidas en ella (Pértega y Pita, 2005). Para la consecución de ésta labor, los investigadores disponen de una variedad de herramientas destinadas a la evaluación de la calidad científica de diversos estudios, habiendo *checklist* y escalas de evaluación de la calidad científica tanto para estudios completamente aleatorizados, como para investigación no experimental y cualitativa (una exposición detallada y actual de éstas herramientas se encuentra en Carro, 2016). Es importante señalar que, a pesar del uso cada vez más extendido de dichas escalas, la mayoría de éstas han sido desarrolladas en contextos clínicos y de salud y, además, algunas de adolecen de limitaciones derivadas de sus características en términos de validez y confiabilidad (véase Cascaes, Valdivia, da Rosa, Barbosa y da Silva, 2013, para la exposición detallada de 14 escalas y 11 listas utilizadas frecuentemente en el área de salud).

El análisis y síntesis de datos en una RS puede llevarse a cabo de distintas maneras, dependiendo de los objetivos y el tipo de estudios revisados. Cabe decir que no en todas las RS se pretenden realizar una síntesis de resultados, siendo posible ejecutar únicamente una descripción de las principales características de los estudios en cuestión (por ejemplo: características metodológicas y conceptuales). Dentro de las técnicas más utilizadas se encuentra el análisis de contenido que resulta ser una herramienta útil para la generación de teorías y conocimiento a partir del análisis riguroso de datos textuales, siendo de principal relevancia en la integración de estudios cualitativos (Fingfeld-Connett, 2014). El uso de estadística, resulta de principal relevancia en la integración de datos cuantitativos, como puede notarse en los meta-análisis.

Con la finalidad de ilustrar la ejecución de las RS, se ofrecen dos ejemplos reales de este tipo de investigaciones. Particularmente, se describirán los procedimientos realizados en las mismas, contribuyendo así a esclarecer el aspecto práctico de las RS.

Caso 1: Análisis de la regulación jurídica de medios de comunicación y del ejercicio de la psicología en México: un estudio comparativo de la producción científica

El presente caso corresponde al trabajo realizado por Castelán (2017) que tuvo como objetivo analizar la producción científica de la regulación jurídica en medios de comunicación y de la práctica de la psicología en México. Para lograrlo, la autora planteó los siguientes objetivos específicos: describir la práctica de la regulación en los medios de comunicación y en el ejercicio de la psicología, establecer la proporción de estudios primarios o empíricos y de aquellos de carácter teórico o tipo ensayo dentro del total de la producción científica, comparar el tipo de diseño en la producción científica así como evaluar la calidad científica de los estudios primarios a través de la Escala de Calidad de Estudios Primarios (ECEP), determinar las principales categorías sobre el tipo de estudio, diseño, temáticas particulares en los estudios observados en la producción científica, determinar las principales categorías sobre el tipo de estudio, diseño, temáticas particulares en los estudios observados en función de la institución productora, y año de publicación, comparar la calidad científica de los estudios primarios observados, en función a la institución productora, tipo de estudio y diseño, año de publicación y temática particular y comparar los estudios primarios tipo teórico o ensayo mediante la temática, objetivo y las conclusiones.

Para llevar a cabo esto las unidades de análisis utilizadas fueron artículos de estudios primarios concerniente a la regulación jurídica en medios de comunicación y del ejercicio de la psicología en México. Estos artículos fueron obtenidos de la base de datos de la Red de Revistas Científicas de América Latina y el Caribe, España y Portugal (Redalyc) mediante la extracción de los artículos de las 19 revistas en ciencias de la comunicación y las 99 revistas de psicología con las que cuenta.

Se tomaron como características de inclusión que los estudios primarios tuvieran la temática de la regulación jurídica de medios de comunicación y de la regulación jurídica de la práctica en psicología, que las publicaciones fueran del año 2000 al 2016 y que las regulaciones mencionadas en los trabajos fueran hacia México. Se excluyó todo aquel estudio que no cumpliera con las características mencionadas antes.

Se revisaron 43 407 artículos, de los cuales 35 181 fueron del área de psicología y 8226 del área ciencias de la comunicación. De este total sólo se tomó una muestra de 10 artículos del área de ciencias de comunicación ya que únicamente estos cumplieron con las características de inclusión planteadas. Por su parte en el área de psicología ningún estudio cumplió con los criterios solicitados.

El diseño utilizado fue no experimental de corte transversal, desde un enfoque cualitativo y con un alcance descriptivo. Fueron dos las herramientas que se utilizaron para el estudio. La primera fue:

- Análisis de contenido: esta técnica consiste en la recolección e interpretación de textos en diferentes fuentes de información, la técnica utiliza procedimiento sistemático ya que analiza cada una de las partes del documento.

- Escala de Calidad de Estudios Primarios (ECEP), de Chacón-Moscoso, Sánchez-Meca, Sanduete y Alarcón (citados en Carro, 2013) la cual evalúa la calidad científica, cuenta con 34 ítems, los cuales están divididos en tres secciones la primera son las características Metodológicas que están del ítem 1 al ítem 23, después está la segunda sección que son las características sustantivas que va del ítem 24 al ítem 33 y por último la tercera sección que son las características extrínsecas en el cual sólo se encuentran los ítem 34 al ítem 37. El instrumento maneja puntuaciones del 0 al 1, siendo el 0 lo más bajo y el 1 lo más alto, sólo se suman los primeros 19 ítems es decir que la puntuación más alta es de 19. Además de utilizarse un formato complementario que incluye un ítem para describir la regulación jurídica en el marco, y una segunda evaluación que cuenta con 4 ítems los cuales son proposiciones que explora los objetivos, la temática y las conclusiones.

Al contar con todos los artículos de las 19 revistas de ciencias de la comunicación y 99 de psicología, se realizó la búsqueda exhaustiva de los artículos que cumplieran con las características de inclusión mencionadas con anterioridad. El procedimiento que se siguió fue descargar las revistas dividir las por área, después por nombre, volumen y número al tener todo organizado se revisaron título, resumen y palabras claves de cada artículo seleccionando aquellos que incluyeran las palabras: regulación, regulación jurídica, medios de comunicación, ley, leyes, autorregulación, reglamento, práctica, psicología, reglamentación, práctica profesional, profesión, medios de comunicación masiva.

Al término de la búsqueda se obtuvieron 10 artículos en el área de ciencias de la comunicación; en el área de psicología no se encontraron artículos que cumplieran con dichas características. Los trabajos obtenidos se recopilaron en una matriz de Microsoft Excel donde se sistematizó la información en función de: contenido o temática, área, práctica de la regulación, año de publicación, características de la revistas, cumplimiento de la regulación, tipo de estudio y diseño de investigación.

Con los artículos detectados se realizó el análisis de contenido mediante la descripción del tópico principal a partir de la lectura de los textos completos. Se realizó una tabla con los porcentajes de artículos obtenidos en cada una de las 19 revistas del área de ciencias de la comunicación y las 99 revistas del área de psicología.

Y por último para realizar el análisis de la calidad científica se utilizó la Escala de Calidad de Estudios Primarios (ECEP), aplicándose a un total de 10 artículos de ciencias de la comunicación.

Caso 2: Análisis de la calidad científica de estudios primarios sobre comunicación mediática en México

El análisis de la calidad científica, si bien constituye una fase importante dentro de cualquier RS, también puede ser, por sí solo, un objetivo de investigación. El estudio de la calidad de los productos científicos en un determinado campo permite la retroalimentación de la comunidad científica de la cual surgen dichos productos y posibilita, también, el desarrollo de políticas de apoyo e incentivación a la ciencia, con base en criterios precisos de evaluación.

La investigación que se presenta en este caso, desarrollada por Mendoza (2017), tuvo por objetivos determinar la calidad científica de los estudios primarios realizados en materia de comunicación mediática en el contexto mexicano, además de determinar diferencias en la calidad de los estudios primarios analizados de acuerdo al año de publicación y tipo de estudio.

Para la consecución de dichos objetivos se realizó un mapeo sistemático. Se consideraron 69 revistas de Red de Revistas Científicas de América Latina y el Caribe, España y Portugal (Redalyc), abarcándose el periodo comprendido entre el año 2000 y 2016. Se utilizó una estrategia de búsqueda exhaustiva, evitando el uso de palabras clave en buscador; optando por la identificación manual en cada una de las revistas. En dicha labor participaron cuatro codificadores (A, B, C y D), siendo uno de ellos (A) quien fungió como capacitador de los tres restantes, debido a su experiencia en el área de investigación. Cabe mencionar que, mientras que el Codificador A conocía los objetivos e hipótesis de la presente investigación, los otros tres codificadores no, reduciendo así posibilidades sesgo. Los criterios de inclusión y exclusión fueron los siguientes:

- a. Artículos de investigación empírica (estudios primarios), teórica o investigación secundaria (revisiones de literatura, meta-análisis).
- b. Fueran de producción mexicana; para esto, se consideró aquellos artículos en los que la institución de adscripción de, al menos, uno de los autores, fuera mexicana.
- c. Que la fecha del artículo estuviera en el intervalo entre los años 2000 y 2016.
- d. Que no se tratara de editoriales, cuentos, poemas, entrevistas o reseñas literarias y, en general, cualquier producto literario que no fuera resultado de un proceso de investigación empírica y/o teórica, o secundaria (cabe resaltar que, los ensayos de opinión fueron también incluidos, considerándose como tentativas de trabajo teórico, siempre que contaran con referencias literarias).

e. Para ser incluido, el artículo debía adecuarse a la clasificación de estudios de medios que a continuación se sintetiza:

i. Debían incluirse aquellos trabajos que contemplaran los siguientes medios de comunicación: televisión, radio, prensa (revistas, periódicos, libros), medios telemáticos (internet, telefonía móvil) y, en general, cualquier sistema o herramienta que sirviera como mediación para la consecución de un proceso de comunicación.

ii. Debían incluirse trabajos en donde los medios de comunicación fueran analizados ya sea como una variable dependiente de otros factores (socio-culturales, económicos y/o políticos), o bien, como una variable independiente, generadora de efectos. Asimismo, habrían de incluirse aquellos trabajos donde los medios no fueran parte de una relación de dependencia, como en el caso de estudios de tipo descriptivo sobre los medios.

f. Como criterio de exclusión se consideró que los artículos de contenido eminentemente técnico/ingenieril, no serían incluidos en el conjunto de unidades de análisis.

La primera parte de la recolección, llevada a cabo por el codificador A, fue realizada entre el 27 de mayo del 2016 y el 30 de septiembre de 2016, a lo largo de las 15 revistas indexadas en la sección de ciencias de la comunicación de la base de datos Redalyc. En esta primera inspección, se obtuvieron 275 artículos.

La segunda parte de la recolección fue llevada a cabo por los codificadores B, C y D, quienes tuvieron la tarea de revisar 54 revistas de la categoría multidisciplinar (ciencias sociales) de la base de datos Redalyc; éstas se repartieron equitativamente entre los tres. La cantidad de artículos obtenidos en esta recolección fue de 46.

De dicha recolección se obtuvo una muestra aleatoria de 25 artículos que sí cumplieran con los criterios de inclusión, mientras que, otros 25 que no cumplieran con dichos criterios también fueron seleccionados, siendo así un total de 50 unidades, otorgadas a los cuatro codificadores. Cada uno de ellos tuvo que decidir cuales artículos de esa muestra cumplieran con los criterios de inclusión; el registro de dichas decisiones se llevó a cabo con respuesta dicotómica “SI/NO”.

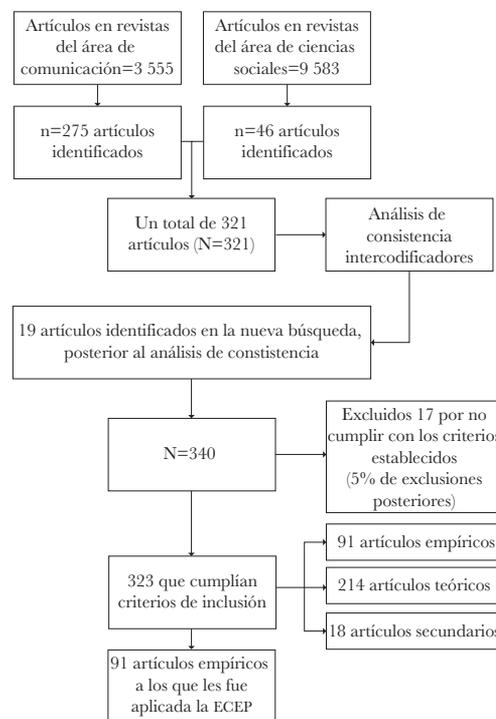
De esta forma procedió a evaluarse el grado de acuerdo entre cada uno de los codificadores a través del índice Kappa, el cual es un indicador apropiado para variables de tipo dicotómico, en situaciones de comparación que involucran como máximo a dos codificadores (Siegel y Castellan, 1995).

En general, los resultados fueron positivos, indicando grados de congruencia aceptable entre los codificadores; sin embargo, para resarcir las discrepancias observadas, se volvió a realizar una capacitación de los codificadores B, C y D, y

se llevó a cabo una tercera etapa de recolección solamente en aquellos años de publicación en los cuales fueron encontrados los artículos de la revisión realizada por los codificadores B, C y D. Esta nueva revisión arrojó 19 nuevos artículos, por lo que el conjunto de artículos recolectados por los codificadores B, C y D pasó de 46 a 65.

Finalmente 17 artículos fueron excluidos tras la recolección, debido a que se trataba de textos de carácter técnico y/o ingenieril. Así, de 13 138 artículos revisados, 323 constituyeron las unidades de análisis utilizadas en la investigación (véase Figura 2).

Figura 2. Diagrama de flujo del proceso de recolección e identificación de los artículos sobre comunicación mediática. Fuente: Mendoza (2017).



Para medir la calidad científica de los artículos identificados, se procedió a hacer uso de los 19 ítems que miden la calidad metodológica, correspondientes a la Escala de Calidad de Estudios Primarios (ECEP), misma que ha sido descrita en el caso anterior. Para la codificación de los estudios se utilizó una hoja de datos en Microsoft Excel (véase Tabla 3), en la cual, además de incluirse los ítems de la ECEP, se incluyeron cuatro ítems más para recolectar información sobre el tipo de

análisis de datos utilizado en el artículo evaluado, la institución de adscripción del autor, el título de la revista, año, volumen y medio de comunicación abordado en el texto.

Tabla 3. Formato de registro utilizado para la recolección de los datos

| Código | Ítem 0 | ... | Ítem 19 | Índice de calidad | TAD* | IAA** | Revista | Año | Volumen y número | MCA*** |
|--|--------|-----|---------|-------------------|------|--|---------|-----|------------------|--------|
| 1 | | | | | | | | | | |
| . | | | | | | | | | | |
| . | | | | | | | | | | |
| . | | | | | | | | | | |
| 323 | | | | | | | | | | |
| TAD*: Tipo de análisis de datos | | | | | | MCA***: Medio de comunicación abordado | | | | |
| IAA**: Institución de adscripción del autor | | | | | | | | | | |
| Nota: los puntos suspensivos (...) se utilizan para ahorrar espacio, pues se trata de una serie consecutiva. | | | | | | | | | | |

Fuente: (Mendoza, 2017).

Tras la captura de todos los datos en Excel se traspasaron al programa SPSS (versión 23) para los análisis pertinentes. Se utilizó estadística descriptiva e inferencial. Se llevaron a cabo análisis para estudiar la cantidad y tipo de producción científica por revistas, institución de procedencia y año, haciendo uso de tablas de contingencias. Posteriormente se presentaron porcentajes para cada uno de los ítems de la ECEP. Finalmente, se realizaron análisis para estudiar la calidad científica de acuerdo al tipo de estudio de los artículos, y al año de publicación, los análisis inferenciales para la verificación de las hipótesis planteadas en esta investigación consistieron en el uso de pruebas de distribución libre para variables nominales, así como pruebas no paramétricas para diferencias entre grupos.

Conclusiones

A lo largo de este texto, se han definido las RS como un tipo de investigación cuyo objetivo es la descripción y/o síntesis de la literatura científica de un determinado campo. Se ha puesto hincapié en su status como un tipo de investigación legítima, que puede aportar información veraz sobre determinado objeto de estudio. Asimismo, se han comparado sus características con otras formas de revisión de la literatura, como las revisiones narrativas, recalándose las ventajas que ofrecen las

RS con respecto a estas últimas y se han expuesto dos casos prácticos con la finalidad de ilustrar adecuadamente el proceso de ejecución de este tipo de revisiones. El presente texto contribuye a tener una noción general, pero clara, respecto al uso y construcción de una revisión sistemática, contemplándola como una forma de investigación legítima, capaz de contribuir al cuerpo de conocimiento científico en un área determinada.

El énfasis que se ha puesto en proporcionar ejemplos y directrices claras para la ejecución de este tipo de investigaciones, permitirá que este texto sirva como una guía para estudiantes, docentes y, en general, aquellos interesados en realizar RS. Es importante señalar que, en un era donde la cantidad de estudios científicos es cada vez mayor y el acceso a estos se facilita gracias a los medios de comunicación digitales y las redes de información, el diseño y uso de RS juega, y jugará aún más, un papel fundamental en la identificación, evaluación e integración del conocimiento científico, siendo, por lo tanto, necesario que los investigadores reciban una mayor formación en la elaboración, y mejora, de este tipo de investigaciones con el fin de obtener la mejor evidencia posible en el estudio del fenómenos sociales y del comportamiento humano. En este orden de ideas, resultará útil llevar a cabo guías de revisiones sistemáticas en campos particulares, asumiendo que en cada uno de estos se encuentran retos específicos para la elaboración de revisiones. Si bien aquí se han presentado diversos ejemplos desde las ciencias del comportamiento, exponer casos propios de otras disciplinas, ayudará a esclarecer el uso de RS en contextos diversos. En todo caso, los autores del presente texto consideran que el objetivo de ofrecer una caracterización y guía general para el desarrollo de revisiones sistemáticas se ha cumplido.

Referencias

- Berg, R. Munthe-Kaas, H. & Ross, M. (2016). "Internalized Homonegativity: A Systematic Mapping Review of Empirical Research", *Journal of Homosexuality*, 63(4), 541-558. doi: 10.1080/00918369.2015.1083788
- Campbell, D. & Laurence, H. (1968). "The Connecticut Crackdown on speeding: time series data in quasi-experimental analysis". *Law and Society Review*, 3(1), 33-53. Recuperado de: http://www.jstor.org/stable/3052794?origin=JSTORpdf&seq=1#page_scan_tab_contents
- Campbell, D. y Stanley, J. C. (1973). *Diseños experimentales y cuasiexperimentales en la investigación social*. Argentina: Amorrortu Editores.
- Carro, E. (2016). "Avances metodológicos en la evaluación de la calidad de la producción científica (investigadores y tesis de pregrado y postgrado) en psicología desarrolladas

- en la región centro y noreste de México” (Tesis doctoral). Universidad de Sevilla, España. Recuperado de: <https://idus.us.es/xmlui/handle/11441/41092>
- Carro, E. (2013). “Características básicas y utilidad del metaanálisis en las ciencias del comportamiento”. *Revista de Psicología y Ciencias del Comportamiento de la U.A.C.¿.S.*, 4(2), 47-66. Recuperado de: <http://www.revistapcc.uat.edu.mx/index.php/RPC/article/view/56/68>
- Cascaes, F., Valdivia, B. A., da Rosa, R., Babosa, P.J. y da Silva, R. (2013). “Escalas y listas de evaluación de la calidad de estudios científicos”. *Revista Cubana de Información en Ciencias de la Salud*, 24(3), 295-312. Recuperado de: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2307-21132013000300007
- Castelán, J. (2017). *“Análisis de la regulación jurídica de medios de comunicación y del ejercicio de la psicología en México: un estudio comparativo de la producción científica”* (tesis de licenciatura). Universidad Autónoma de Tamaulipas, México.
- Cooke, A., Smith, D. & Booth, A. (2012). “Beyond PICO: The SPIDER tool for qualitative evidence synthesis”. *Qualitative Health Research*, 22(10), 1435-1443. doi: 10.1177/1049732312452938
- Dechartres, A., Ravaud P., Atal I., Riveros, C. & Boutron ,I. (2016). “Association between trial registration and treatment effect estimates: a meta-epidemiological study”. *BMC Medicine*, 14(1), 14:100. doi: 10.1186/s12916-016-0639-x.
- Ferreira, I., Urrútia, G. y Alonso-Coello, P. (2011). “Systematic reviews and meta-analysis: scientific rationale and interpretation”. *Revista Española de Cardiología*, 64(8), 688-696. doi: 10.1016/j.recesp.2011.03.029
- Fingfeld-Connett, D. (2014). “Use of content analysis to conduct knowledge-building and theory-generating qualitative systematic reviews”. *Qualitative Research*, 14(3), 341-352. doi: 10.1177/1468794113481790
- France, E., Wells, M., Lang, H. & Williams, B. (2016). “Why, when and how to update a meta-ethnography qualitative synthesis”. *Systematic Reviews*, 5:44. doi: 10.1186/s13643-016-0218-4
- French, J.R. (1953). “Experimentos en contextos de campo”. En L. Festinger y D. Katz (Comps.) *Los métodos de investigación en las ciencias sociales* (pp. 104-136). D.F, México: Paidós.
- García-Perdomo, H. (2015). “Conceptos fundamentales de las revisiones sistemáticas/metaanálisis”, *Revista Urología Colombiana*, 24(1), 28-34. doi:10.1016/j.uroco.2015.03.005
- Glass, G. (1976). “Primary, secondary and meta-analysis research”. *Educational Researcher*, 5(10), 3-8. Recuperado de: <http://nutrigen.ph.ucla.edu/files/view/epi-m258-spring-2012/Glass.pdf>
- Gough, D. (2015). “Qualitative and mixed methods in systematic reviews”, *Systematic Reviews*, 4:181, 1-3. doi: 10.1186/s13643-015-0151-y

- Gough, D., Oliver, S. & Thomas, J. (2017). "Introducing systematic reviews". En D. Gough, S. Oliver & J. Thomas (Comp.), *An introduction to Systematic Reviews* (pp. 1-18). Sage.
- Gough, D. & Thomas, J. (2017). "Commonality and diversity in reviews". En D. Gough, S. Oliver & J. Thomas (Comp.), *An introduction to Systematic Reviews* (pp. 43-70). Sage.
- Grant, M. & Booth, A. (2009). "A typology of reviews: an analysis of 14 review types and associated methodologies", *Health Information and Libraries Journal*, 26(2), 91-108. doi: 10.1111/j.1471-1842.2009.00848.x
- Guillaumie, L. Boiral, O. & Champagne, J. (2016). "A mixed-methods systematic review of the effects of mindfulness on nurses". *Journal of Advanced Nursing*, 73(5), 1017-1034. doi: 10.1111/jan.13176
- Hausner, E., Guddat, C., Hermanns, T., Lampert, U. & Waffenschmidt, S. (2015). "Development of search strategies for systematic reviews: validation showed the noninferiority of the objective approach". *Journal of Clinical Epidemiology*, 68(2), 191-199. doi: 10.1016/j.jclinepi.2014.09.016
- Higgins, J. P. y Green, S. (2011). "Cochrane Handbook for Systematic Reviews of Interventions Versión 5.1.0". Recuperado de: https://es.cochrane.org/sites/es.cochrane.org/files/public/uploads/Manual_Cochrane_510_reduit.pdf
- James, K., Randall, N. & Haddaway, N. (2015). "A methodology for systematic mapping in environmental sciences", *Environmental Evidence*, 5:7, doi: 10.1186/s13750-016-0059-6
- Katz, D. (1953). "Los estudios de campo". En L. Festinger y D. Katz (Comps.) *Los métodos de investigación en las ciencias sociales* (pp. 67-103). D.F., México: Paidós.
- Kerlinger, F. y Lee, H. B. (2002). *Investigación del comportamiento: métodos de investigación en ciencias sociales*. México, D.F.: McGraw Hill.
- Khan, K., Kunz, R., Kleijnen, J. & Antes, G. (2003). "Five steps to conducting a systematic review". *Journal of the Royal Society of Medicine*, (96)3, 118-121. Recuperado de: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC539417/>
- Letelier, L. M., Manríquez, J. J. y Rada, G. (2005). "Revisiones sistemáticas y metaanálisis: ¿son la mejor evidencia?". *Revista Médica Chilena*, 133, 246-249. Recuperado de: http://www.scielo.cl/scielo.php?pid=S003498872005000200015&script=sci_arttext
- Letelier, L. M. y Moore, P. (2003). "La medicina basada en evidencia. Visión después de una década". *Revista Médica de Chile*, 131, 939-946. doi: 10.4067/S0034-98872003000800016
- Mendoza, O. (2017). "Análisis de la calidad científica de estudios primarios sobre comunicación mediática en México: una aproximación metodológica desde las ciencias del comportamiento" (Tesis de licenciatura). Universidad Autónoma de Tamaulipas, México.
- Methley, A., Campbell, S., Chew-Graham, C., McNally, R. & Cheraghi-Sohi, S. (2014). "PICO, PICOS and SPIDER: a comparison study of specificity and sensitivity in three

- search tools for qualitative systematic reviews”. *BMC Health Services Research*, 14:579. doi: 10.1186/s12913-014-0579-0
- Milton, A. & Mullan, B. (2014). “Communication of a mental health diagnosis: A systematic synthesis and narrative review”, *Journal of Mental Health*, 23(5), 261-270. doi: 10.3109/09638237.2014.951474
- Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L. & PRISMA-P Group (2015). “Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) statement” *Systematic Reviews*, 4:1. doi: 10.1186/2046-4053-4-1
- Moncada-Hernández, S. (2014). “Cómo realizar una búsqueda de información eficiente. Foco en estudiantes, profesores e investigadores en el área educativa”, *Investigación en educación médica*, 3(10), 106-115. Recuperado de <http://www.scielo.org.mx/pdf/iem/v3n10/v3n10a7.pdf>
- Moreno, B., Muñoz, M., Cuellar, J., Domancic, S. y Villanueva, J. (2018). “Revisiones Sistemáticas: definición y nociones básicas”, *Revista clínica de periodoncia, implantología y rehabilitación oral*, 11(3), 184-186. doi: 10.4067/S0719-01072018000300184
- Morgan, K., Morgan, M., Johansson, L. & Ruud, E. (2016). *Systematic Mapping of the Effects of Ict on Learning Outcomes*, Oslo, Norway: Knowledge Center for Education.
- Mouton, J. & Marais, H. (1990). *Basic Concepts in the Methodology of Social Sciences*. Pretoria, South Africa: HSRC Publishers.
- Noblit, G. & Dwight, R. (1988). *Meta-Ethnography: Synthesizing Qualitative Studies*. United States of America: Sage Publications Inc.
- Oliver, S., Sutcliffe, K. & Richardson, M. (2017). “Describing and analyzing studies”. En D. Gough, S. Oliver y J. Thomas (Comp.), *An introduction to Systematic Reviews* (pp. 123-144). London: Sage.
- Pearson, A., White, H., Bath-Hextall, F., Salmond, S., Apostolo, J. & Kirkpatrick P. A. (2015). “Mixed-methods approach to systematic reviews”, *International Journal of Evidence-Based Healthcare*, 13(3), 121-131. doi: 10.1097/XEB.0000000000000052
- Perestelo-Pérez, L. (2013). “Standards on how to develop and report systematic reviews in Psychology and Health”. *International Journal of Clinical and Health Psychology*, 13(1), 49-57. doi: 10.1016/S1697-2600(13)70007-3
- Pérttega, S. y Pita, S. (2005). “Revisiones sistemáticas y metaanálisis”. *Cadernos de Atención Primaria*, 12(2), 109-112. Recuperado de: <http://www.fisterra.com/mbe/investiga/metaanalisis/RSyMetaanalisis.asp>
- Petticrew, M. & Roberts, H. (2006). *Systematic Reviews in the Social Sciences: A Practical Guide*. Oxford: Blackwell.
- Relevó, R. (2012). “Effective Search Strategies for Systematic Reviews of Medical Tests”. In: S. Chang, D. Matchar, G. Smetana, C. Umscheid (eds). *Methods Guide for Medical*

- Test Reviews* [Internet]. Rockville, United States of America: Agency for Healthcare Research and Quality. Recuperado de: <https://www.ncbi.nlm.nih.gov/books/NBK98242/>
- Rodríguez, B., Zafra, S., Quintero, S. (2015). “La revisión sistemática de la literatura científica y la necesidad de visualizar los resultados de las investigaciones”, *Revista Logos, Ciencia & Tecnología*, 7(1), 101-103. doi: 10.22335/rfct.v7i1.232
- Russell, R., Chung, M., Balk, E., Atkinson, S., Giovannucci, E., Ip, S., Lau, J. (2009). “Issues and Challenges in Conducting Systematic Reviews to Support Development of Nutrient Reference Values: Workshop Summary”. *Nutrition Research Series*, vol. 2, United States: Agency for Healthcare Research and Quality. Recuperado de: <https://www.ncbi.nlm.nih.gov/books/NBK44081/>
- Salvador, J., Marco, G. y Arquero, R. (2018). “Las revisiones sistemáticas en biblioteconomía y documentación: análisis y evaluación del proceso de búsqueda”, *Revista Española de documentación científica*, 41(2), 1-16. Recuperado de: <http://redc.revistas.csic.es/index.php/redc/article/view/1011/1599>
- Sampaio, R. F. y Mancini, M. C. (2007). “Systematic review studies: a guide for careful synthesis of scientific evidence”. *Revista Brasileira de Fisioterapia*, 11(1), 77-82. doi: 10.1590/S1413-35552007000100013
- Sánchez-Meca, J. (2010). “Cómo realizar una revisión sistemática y un meta-análisis”. *Aula Abierta*, 38(2), 53-64. Recuperado de: <http://www.um.es/metaanalysis/pdf/5030.pdf>
- Sheals, K., Tombor, I., McNeill, A. & Shahab, L. (2016). “A mixed-method systematic review and meta-analysis of mental health professionals’ attitudes toward smoking and smoking cessation among people with mental illnesses”. *Addiction*, 111(9), 1536-53. doi: 10.1111/add.13387
- Shepherd, A., Laurens, K., Matheson, S., Carr, V. & Green, M. (2012). “Systematic meta-review and quality assessment of the structural brain alterations in schizophrenia”. *Neuroscience and Biobehavioral Reviews*, 36(4), 1342-1356. doi: 10.1016/j.neubiorev.2011.12.015
- Silbaugh, B. C., Penrod, B., Whelan, C. M., Hernandez, D. A., Wingate, H. V., Falcomata, T. S., ... Lang, R. (2016). “A systematic synthesis of behavioral interventions for food selectivity in children with autism spectrum disorders”. *Review Journal of Autism and Developmental Disorders*, 3, 345-347. doi:10.1007/s40489-016-0087-8
- Siegel, S. y Castellan, J. (1995). *Estadística no paramétrica: aplicada a las ciencias de la conducta*. (4a ed.). México: Trillas
- Smith, M. & Glass, G. (1977). “Meta-analysis of psychotherapy outcome studies”. *American Psychologist*, 32(9), 752-760. Recuperado de: https://www.researchgate.net/publication/22234005_MetaAnalysis_of_Psychotherapy_Outcome_Studies

- Smith, M., Glass, G. & Miller, T. (1980). *The benefits of psychotherapy*. Baltimore, USA: The John Hopkins University Press.
- Sobrido, M. y Rumbo-Prieto, M. (2018). “La revisión sistemática: pluralidad de enfoques y metodologías”, *Enfermería Clínica*, 28(6), 387-393. doi: 10.1016/j.enfcli.2018.08.008
- Solomon, P., Cavanaugh, M. & Draine, J. (2009). *Randomized Controlled Trials: Design and Implementation for Community-Based Psychosocial Interventions*. New York: Oxford University Press.
- Stroup, D., Berlin, J., Morton, S., Olkin, I., Williamson, G., Rennie, D., Thacker, S. (2000). “Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group”. *Journal of the American Medical Association*, 283(15), 2008-2012. doi: 10.1001/jama.283.15.2008
- Torgerson, C. (2003). *Systematic reviews*. Gran Bretaña: Continuum.
- Urrua, E. y Barría, R. M. (2010). “La revisión sistemática y su relación con la práctica basada en la evidencia en salud”. *Revista Latinoamericana de Enfermagem*, 18(4), 1-8 pantallas. Recuperado de: http://www.scielo.br/pdf/rlae/v18n4/es_23.pdf
- Vidal, M., Oramas J. y Borroto, R. (2015). “Revisiones sistemáticas”, *Educación Médica Superior*, 29(1), 198-207. Recuperado de: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-21412015000100019
- Ward, M., White, D. & Druss, B. (2015). “A meta-review of lifestyle interventions for cardiovascular risk factors in the general medical population: lessons for individuals with serious mental illness”. *The Journal of Clinical Psychiatry*, 76(4), e477-e486. doi: 10.4088/JCP.13r08657

Capítulo 6

Relaciones entre confiabilidad y tamaño de efecto en dos muestras de registros empíricos¹

*Carro-Pérez, Ennio Héctor
Ampudia-Rueda, Amada*

Resumen

La confiabilidad es una cualidad fundamental que permite asegurar la estabilidad en las mediciones de variables, y en consecuencia formular afirmaciones a partir de los hallazgos en un estudio científico, por tal motivo innovar mediante herramientas que abonen a la demostración de los procedimientos actuales o bien mediante alternativas a los existentes es relevante. Por tal motivo el presente trabajo tuvo como propósito demostrar la relación que existe entre la confiabilidad y el tamaño de efecto en dos grupos de puntuaciones empleados tradicionalmente para la determinación de consistencia interna entre codificadores o en instrumentos de medida en procesos test-retest, de formas alternas o paralelas, o de mitades aleatorias o equivalentes. Se analizaron 225 pares de registros temporales, recolectados o codificados por 2 asistentes de investigación en dos sesiones. Los procedimientos para observar la consistencia de los registros fueron el R de Pearson y el Alfa de Cronbach, para el tamaño de efecto se calculó la d de Cohen. Los resultados permiten observar la existencia de valores cercanos a 0.9 de consistencia interna a la par de valores menores a 0.25 de efecto, lo cual se acentuaba en la medida en que los registros con más diferencias entre codificadores se eliminaban. Lo que permite afirmar la probable existencia de una correlación negativa o inversa entre la consistencia o confiabilidad de dos conjuntos de puntuaciones el tamaño de efecto.

Palabras clave: confiabilidad, tamaño de efecto, grupos de puntuaciones, medición, metodología.

Introducción

La confiabilidad y la validez de las medidas en las ciencias son aspectos fundamentales para determinar con un margen de certidumbre el estado o nivel de las variables que intervienen en un fenómeno, lo que se traduce en la posibilidad de contar

¹ El presente trabajo se pudo realizar gracias al apoyo otorgado por la Dirección General de Cooperación e Internacionalización (DGEI) de la UNAM, a la Dirección de Internacionalización y Colaboración Académica, y a la Dirección de Programas de Apoyo de la Universidad Autónoma de Tamaulipas.

con estimaciones y descripciones precisas de la realidad estudiada. De tal manera, tener procedimientos y algoritmos para calcular ambos aspectos es fundamental, especialmente en las ciencias del comportamiento donde las medidas atraviesan con ciertas dificultades para ser desarrolladas e instrumentadas, dificultades propias de las condiciones del ser humano, al ser sujeto y objeto de medida.

Manuales sobre metodología como los de Kerlinger y Lee (2002), Clark-Carter (2002), Thorndike y Hagen (2012), Nunnally (2013), Coolican (2005) y Nadelsticher (1983) coinciden en la existencia de al menos cuatro tipos de validez, la de contenido, constructo, criterio y de discriminación del ítem, mientras que la confiabilidad se ve como la consistencia interna de las puntuaciones de una prueba o instrumento de medida, la cual es calculada a través de procedimientos básicos como el de formas paralelas; contraste de la prueba en dos momentos o test-retest; la división del instrumento en dos partes aleatorias y simétricas de la prueba para ser comparadas; y el uso de coeficientes de confiabilidad.

En condiciones ideales los indicadores de confiabilidad de las medidas en comportamiento humano suelen obtenerse mediante la comparación de dos conjuntos de puntuaciones sobre una variable a fin de establecer un índice de correlación, como el coeficiente R de Pearson, Rho de Spearman, Kappa de Cohen entre otros, que permitan representar en valores de 0 a 1 o en porcentajes la consistencia o estabilidad de la medida, lo cual también puede lograrse a través de un solo conjunto de puntuaciones mediante un índice de confiabilidad como el Alfa de Cronbach, KR-20 o KR-21 (Cohen y Swerdlik, 2001; Nadelsticher, 1983). Las limitaciones de ambos procedimientos surgen, en el caso de los primeros, en el momento de considerar el valor de referencia para establecer la consistencia y/o validez idónea de una variable continua monodimensional como el tiempo en que se emite un comportamiento o bien la emisión de una conducta, y en el caso de los segundos en el simple hecho de que los coeficientes han sido diseñados para obtener la consistencia de variables con $n > 1$ dimensiones.

A pesar de estas limitaciones, los procedimientos son usados ampliamente y los cálculos de los mismos son fáciles y rápidos de obtener al estar integrados en las herramientas de base de software de análisis estadístico. De tal manera, no es deseable considerar que los procedimientos para determinar confiabilidad actualmente disponibles para investigadores y usuarios en el campo de las ciencias del comportamiento, sean los únicos posibles y que no se puedan perfeccionar o desarrollar otros, o bien puedan existir alternativas que permitan obtener indicadores de consistencia interna de las mediciones o de las puntuaciones obtenidas a través de los instrumentos. En este sentido el campo de la metodología en las ciencias del comportamiento, particularmente en el área de la medición y el análisis de datos es

un continente dinámico, en el que investigadores se encuentran comprometidos en el mejoramiento de los procedimientos existentes y en la búsqueda de alternativas a los mismos. El presente capítulo tiene por objetivo mostrar la relación que existe entre el tamaño de efecto y la estabilidad o consistencia de las medidas, con lo que el cálculo del tamaño de efecto puede ser una alternativa para la obtención de indicadores de confiabilidad, de tal manera, esta interacción entre ambos aspectos, confiabilidad y efecto, hace que el tamaño de efecto puedan ser empleado como indicador de consistencia o estabilidad de las puntuaciones independientemente de la demostración de la misma mediante un coeficiente o algoritmo de confiabilidad.

Se parte del supuesto de que la consistencia o confiabilidad de las medidas es obtenida a partir de que tan estables son las puntuaciones recogidas por un instrumento, lo cual se evidencia a través de correlatos positivos altos entre dos grupos de puntuaciones, en este sentido una confiabilidad aceptable estaría indicada por la existencia de correlaciones superiores a 0.7 (Arribas, 2004) entre un conjunto de puntajes A (que podría ser un primer momento de medición, una mitad aleatorizada de un conjunto de puntuaciones o las mediciones de la versión A de un instrumento) con un conjunto de puntajes B (que podría ser un segundo momento de medición, una mitad aleatorizada de un conjunto de puntuaciones o las mediciones de la versión B de un instrumento), en otras palabras lo que se obtiene en un primer momento de medida (A) debe ser similar en un segundo momento (B), (Thorndike y Hagen, 2012), así mismo esta condición de no variabilidad entre A y B puede indicar un nulo o escaso efecto entre estos grupos de puntajes, en sentido contrario, la falta de similitud o igualdad entre A y B , siempre que las condiciones de aplicación o medición sean idénticas en ambos casos, indicaría baja confiabilidad o estabilidad y un alto efecto de A sobre B .

Con base en lo anterior, la hipótesis planteada es que las medidas de confiabilidad como el Alfa de Cronbach o el coeficiente R de Pearson aplicado en formas paralelas, test-retest o mitades similares del instrumento, se asociaría a medidas del efecto como la d de Cohen de manera negativa o inversa, es decir, una alta confiabilidad de un instrumento o de la medida de una variable implicará un tamaño de efecto bajo.

Para probar el supuesto anterior y observar si el cálculo del tamaño de efecto puede ofrecer un indicador de la estabilidad de un conjunto de mediciones, se analizaron 269 registros de tiempo, tomados a automóviles en un tramo de vía urbana de 63.61 metros de longitud. Los cuales fueron empleados para calcular la velocidad de los automóviles e igualar los registros de esta variable por parte de codificadores independientes.

Se considera que los resultados obtenidos pueden ser generalizables a variables cuya medida se puede sintetizar en un solo valor o dimensión como es el caso de la velocidad u otras como los tiempos de reacción de un individuo a un estímulo o las tasas de respuesta o emisión de comportamientos en contextos diversos, en este tenor, los registros de emociones y motivaciones entrarían en este supuesto, inclusive podría ser aplicado a los puntajes globales de escalamientos en formato Likert u Osgood.

Método

Muestra

Para el presente trabajo se consideraron dos muestras no aleatorias de pares de registros de tiempo en segundos y centésimas de segundo (1/100), tomados a automóviles particulares, no comerciales, tipo sedán y camionetas cerradas clasificadas como SUV (Sport Utility Vehicle), al desplazarse dentro de un tramo carretero urbano de una longitud de 63.61 metros. La primera muestra estuvo constituida por 44 pares de registros de tiempo, y la segunda por 225 pares de registros. Para los fines del presente trabajo la falta de aleatorización de los registros no constituye un sesgo, ya que no se pretende realizar una inferencia sobre los mismos respecto al comportamiento de una población de automovilistas, sino exponer y probar la relación existente entre la estabilidad de los datos y el tamaño de efecto de los mismos calculado a través de la d de Cohen (Cárdenas y Arancibia, 2014; Sullivan y Feinn, 2012) y como este puede aportar un indicador de confiabilidad o al menos reforzar los utilizados en investigación.

Instrumentos

Para la medida del tiempo se ocuparon cronómetros marca Casio, modelo HS-3(V), hechos en China, con medidas de hora, minutos, segundos, y 1/100 s, y hojas de registro para la anotación del tiempo obtenido a través del cronómetro, en segundos y 1/100 s. Los registros fueron realizados por dos codificadores de manera simultánea.

Procedimiento

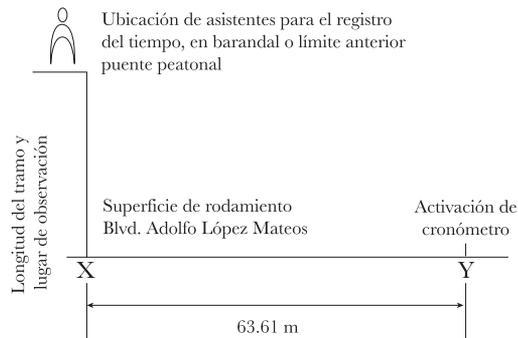
Los registros fueron levantados por tres asistentes de investigación en campo, dos de ellos manipularon cronómetros para obtener la medida de tiempo, en segundos y 1/100 s, que empleaba un automóvil para cubrir una distancia de 63.61 metros, establecida por dos puntos, uno de origen X y uno de término T , marcados en el carril central de una vía carretera o autovía principal de carácter urbano de la zona conurbada del sur de Tamaulipas (ZCST), específicamente en el Boulevard Adolfo

López Mateos (BALM), frente al acceso principal del Campus Sur de la Universidad Autónoma de Tamaulipas, en la ciudad de Tampico.

Para realizar los registros, los asistentes de investigación se ponían de acuerdo respecto al intervalo de selección de los automóviles que transitaran por el carril central del boulevard y pasaran por los puntos de referencia, el intervalo de selección acordado tenía la finalidad de evitar equívocos en los asistentes y midieran el tiempo del mismo automóvil.

Las mediciones de tiempo se obtenían cuando eran activados los cronómetros al momento en que la parte delantera del automóvil pasaba por el punto X , y se desactivaban o paraban cuando el automóvil pasaba por el punto Y , las observaciones se realizaban en un puente peatonal que cruza BALM, desde el cual se tenía la posibilidad de ver sin interrupciones las marcas del tramo definido (ver Figura 1).

Figura 1. Ubicación de asistentes para el registro del tiempo



Los tiempos de ambos cronómetros eran anotados por un tercer asistente en una hoja de registro confeccionada para este fin. De tal manera, se contó con dos conjuntos de registros simultáneos recolectados en dos sesiones o días, en la primera sesión se levantó una muestra de 44 casos y en la segunda de 225. Estos registros se analizaron mediante estadística descriptiva, y los coeficientes de correlación de Pearson (R) y Alfa de Cronbach (α) para determinar la confiabilidad o consistencia de los mismos previa su transformación a unidades de velocidad en km/h. A pesar de las consideraciones planteadas por Pita, Pértegas y Rodríguez (2003) en favor del coeficiente de correlación intraclase (CCI) y en contra del uso de R de Pearson para la determinación de la confiabilidad, en el presente trabajo se consideró que el uso de este último se justificaba a partir de su facilidad de cálculo y el control de la validez de sus resultados al estar emparejado con el Alfa de Cronbach y la d de Cohen.

Después de la obtención de los registros de tiempo, fueron capturados en el programa de análisis estadístico SPSS, donde se calculó la velocidad en km/h promedio (\bar{v}) del automóvil por registro de tiempo y una velocidad diferencial o residual (v_d) como parte de un procedimiento de depuración de las puntuaciones, antes de obtener la consistencia o confiabilidad. La velocidad se determinó a partir de la fórmula clásica que considera una aceleración constante o movimiento uniforme:

$$v = d/t$$

Dónde:

v = velocidad,

d = distancia,

t = tiempo,

De tal manera, la \bar{v} , teniendo en cuenta la existencia de dos registros de tiempo (A y B), se obtuvo a partir de: $\bar{v} = (v_A + v_B)/2$ ó bien, tratándose de más de dos registros: $\bar{v} = \sum v_i / n_v$.

La v_d fue producto de la diferencia entre v_A y v_B y su utilidad consistió por una parte en magnificar las varianzas de los grupos de registros obtenidos por los asistentes, permitiendo apreciar variaciones de velocidad en metros por hora hasta más de 10km/h, que en los registros originales en tiempo representaba una diferencia no mayor a un segundo entre un asistente codificador u otro. Por otra parte, permitió depurar los registros de tiempo y establecer comparaciones entre un antes y un después de la depuración para comprobar si los cálculos de confiabilidad y de efecto cambiaban o se mantenían.

Cómo parte de la comprobación del procedimiento ejecutado se calculó el tamaño de efecto de los registros de tiempo A y B, como conjuntos independientes mediante la ecuación simple de la d de Cohen referida por Clark-Carter (2002) y Sanduvete (2008), de la cual se muestra la siguiente versión adaptada al presente estudio:

$$d = \frac{\bar{x}_1 - \bar{x}_2}{\sigma} = \frac{\bar{x}_A - \bar{x}_B}{\sigma_{A,B}}$$

Donde:

\bar{x}_A = media del tiempo o registro A

\bar{x}_B = media del tiempo o registro B

$\sigma_{A,B}$ = desviación estándar combinada de los dos grupos de registros, A y B

En resumen, los pasos ejecutados fueron:

- 1) Conversión de los registros de tiempo en unidades de velocidad según el estándar de lo observado en términos de comportamiento
- 2) Cálculo de la diferencia de las velocidades obtenidas por registro simultáneo
- 3) Comprobación de la diferencia producida por las diferentes unidades de medida en un algoritmo de consistencia o correlación como R
- 4) Eliminación de los registros con un diferencial o residuo de velocidad mayor a la unidad, en nuestro caso a 1km/hr
- 5) Comprobación de la consistencia de los registros antes y después de la eliminación de los registros a través de R y Alfa (α) de Cronbach
- 6) Cálculo de d , considerando la existencia de una relación inversamente proporcional entre la confiabilidad y la d ,² lo que nos permite obtener un indicador adicional del impacto que tienen las diferencias en los registros respecto a la consistencia de los mismos
- 7) Comparación de la consistencia entre registros de tiempo y velocidad una vez efectuada la depuración de los mismos

Resultados

El análisis de los datos consistió en el cálculo de dos indicadores de consistencia o confiabilidad para medidas escalares, el coeficiente R de Pearson empleado para determinar la estabilidad de dos grupos de puntuaciones, y el Alfa (α) de Cronbach, de igual manera se obtuvo el tamaño de efecto a partir de la d de Cohen. Si bien R es un indicador de asociación o correlación, en condiciones de igualdad de herramientas y unidades de medida idénticas aporta un indicio de que tan coincidentes o diferentes pueden ser dos conjuntos de registros y en consecuencia tasar un nivel de consistencia o confiabilidad.

Los índices de confiabilidad y efecto fueron calculados en las muestras de tiempo convertidos a unidades de velocidad antes y después de la depuración de registros descrita en el procedimiento de este estudio a través de la v_d , con la cual se eliminaron los valores entre codificadores o asistentes de registro que suponían una diferencia entre ellos mayor a 1 km/h, hay que recordar que los registros también tenían el propósito de verificar la consistencia intercodificador, y contar con personal entrenado en futuras mediciones.

En el primer levantamiento de registros temporales ($n=44$), las medias de tiempo entre ambos asistentes o codificadores (A y B) fue muy parecida, con una diferencia no mayor a dos décimas, lo que en velocidad se traduce a más de 2km/h,

²En el presente trabajo se empleó el programa GPower 3.1.2 para el cálculo de d y la obtención del gráfico de distribución respectivo.

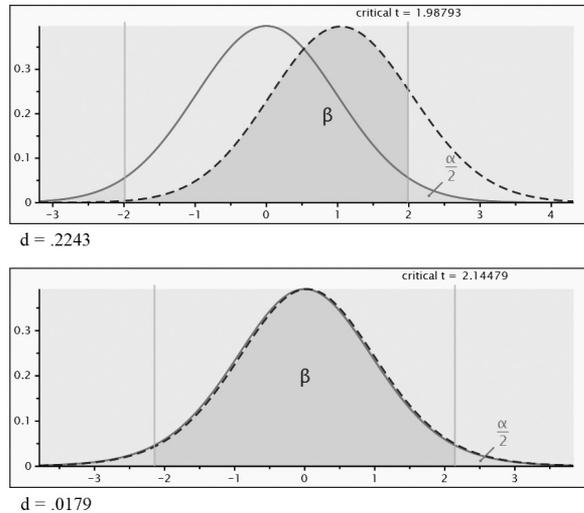
después de la exclusión de 36 registros mayores a 1km/h, la diferencia entre ambos codificadores se redujo a menos de 1km/h, lo que implicó una mayor consistencia o estabilidad de los datos, confirmada en los valores de R , α , y d , antes y después de la depuración o exclusión de registros. En todos los casos la variabilidad intra-conjunto indicada por la desviación estándar o típica (DS) es parecida en el antes y después de la exclusión (ver Tabla 1).

Tabla 1. Valores descriptivos y de consistencia antes-después de la exclusión de los valores con residuales > 1 (km/h)

| Exclusión | Tiempos | Media | DS | V(km/h) | n | R | α | d |
|-----------|---------|--------|---------|---------|----|-------|----------|--------|
| Antes | A | 3.8234 | 0.62393 | 59.89 | 44 | 0.982 | 0.991 | 0.2243 |
| | B | 3.9636 | 0.62589 | 57.77 | 44 | | | |
| Después | A | 4.1625 | 0.75284 | 55.01 | 8 | 0.998 | 0.999 | 0.0179 |
| | B | 4.1763 | 0.78228 | 54.83 | 8 | | | |

Fuente: Elaboración propia.

Figura 2. Curvas poblacional y muestral según tamaño de efecto, antes ($n=44$) y después ($n=8$)



En la tabla anterior se puede apreciar que R y α , se encuentran asociados a valores bajos o cercanos a 0.2 en d (Cárdenas y Arancibia, 2014; Avilés, Morales, Sao y Cañedo, 2004), esto se marca aún más después de excluir los registros con mayor

varianza y dejar los más parecidos entre codificadores, teniendo valores cercanos a uno en los índices de confiabilidad y a cero en el tamaño de efecto. Lo anterior se aprecia de manera gráfica en las curvas de distribución poblacional (en la línea continua) y muestral (en la línea punteada) para prueba t de grupos independientes proporcionadas por el programa *GPower 3.1.2*, las cuales se traslapan en la medida en que el tamaño de efecto disminuye (ver Figura 2).

Si bien, la representación gráfica de las curvas ofrece una aproximación al conocimiento de la probabilidad de cometer el error Tipo I asociado a la significación, y el Tipo II asociado a β y en consecuencia la potencia ($1 - \beta$) para la prueba estadística t (Cárdenas y Arancibia, 2014; Bono y Arnau, 1995), no es de interés inmediato para el presente trabajo, sin embargo a futuro será importante discutir sobre lo esperado en estos casos cuando se intenta establecer similitudes entre conjuntos de datos y no diferencias, donde ya está indicado como deseable una significación igual o mayor a 0.05, una β igual a 0.20 y una potencia de 0.80 (Bono y Arnau, 1995).

Respecto a la segunda sesión donde se realizaron 225 registros de tiempo simultáneos, se pudo observar un comportamiento similar que en el primer levantamiento, así los índices de consistencia mayores a 0.8 se asociaron a un valor de efecto cercano a cero, esta tendencia se acentuó después de la exclusión de valores con márgenes de diferencia intercodificadores mayores a 1 km/h, así con 116 registros simultáneos la confiabilidad se elevó a uno en α y muy cercano a este en R , y la d fue menor de 0.004 (ver Tabla 2).

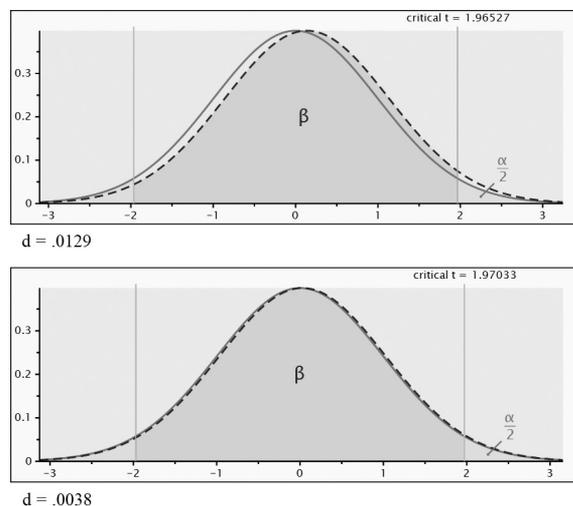
Tabla 2. Valores descriptivos y de consistencia después de la exclusión de los valores con residuales > 1 (km/h)

| Exclusión | Tiempos | Media | DS | V(km/h) | n | R | α | d |
|-----------|---------|--------|---------|---------|-----|-------|----------|--------|
| Antes | A | 4.0220 | 0.84393 | 56.93 | 225 | 0.892 | 0.943 | 0.0129 |
| | B | 4.0331 | 0.87439 | 56.78 | 225 | | | |
| Después | A | 4.2592 | 0.93601 | 53.76 | 116 | 0.999 | 1.000 | 0.0038 |
| | B | 4.2556 | 0.93315 | 53.81 | 116 | | | |

Fuente: Elaboración propia.

Las gráficas producidas por el programa *GPower 3.1.2*, muestran el solapamiento entre las curvas poblacional y muestral para las medias y DS obtenidas en los registros A y B antes y después, el cual es casi total en la medida en que el efecto disminuye (ver Figura 3).

Figura 3. Curvas poblacional y muestral según tamaño de efecto, antes (n=225) y después (n=116).



En ambas muestras de registros a partir de los resultados, se puede establecer que la confiabilidad o estabilidad entre las medidas de tiempo hechas por los asistentes de investigación, se incrementa después de realizar las depuraciones o excluir los registros que implicaban diferencias mayores a 1 km/h entre codificadores. Esto confirma la base en la que se sustenta la confiabilidad, que implica que la invarianza de los datos comparados es igual a una consistencia o estabilidad de uno o 100 por ciento. Así mismo se determinó qué tamaños de efecto bajos (<0.20) (Cárdenas y Arancibia, 2014; Avilés, Morales, Sao y Cañedo, 2004) se asociaron a valores de confiabilidad altos (>0.70) (Arribas, 2004), y entre más bajos fueron los primeros más altos los segundos. Esta tendencia observada, puede permitir plantear que valores de efecto alto pueden estar asociados a valores de confiabilidad bajos, lo que tendría que ser probado en futuros trabajos.

Conclusiones

La confiabilidad de las mediciones, es un aspecto fundamental para tener la certeza de que se cuenta con indicadores de la realidad estables sobre los que se pueden realizar inferencias y predicciones con cierto rigor científico, el carecer de instrumentos o herramientas cuyas mediciones sean confiables añadiría una gran incertidumbre a las observaciones en la ciencia, esto se podría ejemplificar con el hipotético caso de un observador que al querer registrar la temperatura del agua, se ve imposibilitado de obtener una temperatura idéntica o similar cada vez que la

mide, por falta de confiabilidad con el instrumento empleado, a pesar de que todas las demás condiciones ambientales e instrumentales permanecen iguales (idéntica o misma(o) temperatura del ambiente, hora, lugar de observación, termómetro, recipiente donde se encuentra el agua, entre otras); esto también se puede aplicar al estudio del comportamiento humano cuando un investigador intenta realizar un registro o medición de una conducta o comportamiento, empleando como en el ejemplo anterior una herramienta que evalúa características como la intensidad, cantidad o contenido de la conducta, siempre cambiantes a pesar de que las condiciones personales, sociales y ambientales que tiene al individuo que emite o emitió la conducta sean las mismas y no cambien de una medición a otra. En lo cotidiano, esta falta de confiabilidad, desquiciaría a cualquier individuo que intente tener algo de certidumbre o seguridad sobre sus relaciones personales cada vez que preguntara sobre la validez de las mismas a los demás, y recibiera respuestas diferentes de estos últimos independientemente del cuándo, cómo, o circunstancias de la pregunta y respuesta.

De tal manera, el estudio de la confiabilidad y el desarrollo de procedimientos para determinarla es relevante para el campo de la medición y observación de variables, bajo esta consideración el presente trabajo exploró las relaciones entre este aspecto, que operacionalmente se verifica a través de la estabilidad de las mediciones obtenidas sobre una variable, y el tamaño de efecto, empleando para ello procedimientos comunes en la literatura como el coeficiente R de Pearson, el Alfa de Cronbach (Arribas, 2004; Pita, Pértegas y Rodríguez, 2003), y la d de Cohen (Cárdenas y Arancibia, 2014; Bono y Arnau, 1995).

Los resultados obtenidos del análisis a dos grupos de registros de tiempo, tomados en dos diferentes sesiones, muestra por una parte que la similitud, expresada por los indicadores de confiabilidad utilizados, de los registros de tiempo y en consecuencia de velocidad, crece en la medida en que estos mismos datos son depurados eliminando aquellos en donde se observan mayores diferencias entre codificadores, por otra parte, el cálculo del tamaño de efecto en las diversas condiciones de análisis, antes y después de la depuración, indica que el efecto disminuye y la confiabilidad crece, en este sentido se puede afirmar que en situaciones diseñadas para determinar la confiabilidad entre dos conjuntos de mediciones u observaciones, como nuestro caso, la estimación del tamaño de efecto puede servir también como un indicador más de confiabilidad y robustecer el resultado de otros coeficientes u algoritmos empleados o bien constituirse en una alternativa a los mismos.

La aplicación del tamaño de efecto, particularmente la d de Cohen, como indicador de confiabilidad, es aplicable de manera directa a aquellos procedimientos donde se emplean dos conjuntos de observaciones como

el de *test-retest* donde se correlacionan las puntuaciones de un instrumento en dos momentos diferentes, el de formas paralelas, en el cual se correlacionan las puntuaciones de dos versiones de un instrumento, o en aquellos casos donde se obtiene la confiabilidad a través de la correlación de dos partes aleatorias del instrumento producto de su división, de igual manera es probable que los indicadores de tamaño de efecto también correlacionen de manera inversa o negativa con los coeficientes de confiabilidad, como al menos se ha demostrado en este trabajo con el Alfa de Cronbach.

La confirmación de los hallazgos y la ampliación de los mismos será tarea para futuros trabajos, queda pendiente el comportamiento esperado entre la significación estadística y la potencia en este tipo de casos, por lo pronto la evidencia obtenida permite señalar, como ya se ha planteado, que el tamaño de efecto puede ser una alternativa para observar indicios de confiabilidad o bien un algoritmo confirmatorio de los procedimientos tradicionales.

Referencias

- Arribas, M. (2004). "Diseño y validación de cuestionarios". *Matronas Profesión*. 5 (17). pp. 23-29. Recuperado de http://evidencia.com/wp-content/uploads/2014/07/validacion_cuestionarios.pdf
- Avilés M., R., Morales M., M., Sao A., A. y Cañedo A., R. (2004). "La Colaboración Cochrane en Cuba. Parte VII. Los metaanálisis: aproximaciones útiles para su comprensión". *ACIMEN*. 12 (4). Versión en línea en Scielo. Recuperado de <http://scielo.sld.cu/pdf/aci/v12n4/aci04404.pdf>
- Bono, R. y Arnau, J. (1995). "Consideraciones generales en torno a los estudios de potencia". *Anales de Psicología*, Vol. 11(2), pp. 193-202.
- Cárdenas C., M. y Arancibia M., H. (2014). "Potencia estadística y cálculo del tamaño del efecto en G*Power: complementos a las pruebas de significación estadística y su aplicación en Psicología". *Salud y Sociedad*. Vol. 5 (2), pp. 210-224. Recuperado de <https://doi.org/10.22199/S07187475.2014.0002.00006>
- Clark-Carter, D. (2002). *Investigación cuantitativa en psicología*. México: Oxford.
- Cohen, R. J. y Swerdlik, M. E. (2001). *Pruebas y evaluación psicológicas. Introducción a las pruebas y a la medición* [4ª. Edición]. México, D.F.: McGraw-Gill.
- Coolican, H. (2005). *Métodos de investigación y estadística en Psicología* [3ª Edición]. México: Editorial El Manual Moderno.
- Kerlinger, F. N. y Lee, H. B. (2002). *Investigación del Comportamiento*. México, D.F.: McGraw-Gill.
- Nadelsticher M., A. (1983). *Técnicas para la construcción de cuestionarios de actitudes y opción múltiple*. México: Instituto Nacional de Ciencias Penales.
- Nunnally, J.C. (2013). *Teoría psicométrica* [reimpresión]. México: Trillas.

- Pita F, S.; Pértegas D, S. y Rodríguez M., E. (2003). La fiabilidad de las mediciones clínicas: el análisis de concordancia para variables numéricas. *Cadernos de Atención Primaria*, Vol. 10 (4), pp. 290-296. Recuperado de https://revista.agamfec.com/wpcontent/uploads/2003/12/12InvestigacionN10_4.pdf
- Sanduvete, S. (2008). "Innovaciones metodológicas en la evaluación de la formación continua". [Tesis de Doctorado inédita]. Departamento de Psicología Experimental, Universidad de Sevilla. Recuperado de https://idus.us.es/bitstream/handle/11441/15496/Y_TD_PS295.pdf?sequence=1&isAllowed=y
- Sullivan, G. M. y Feinn, R. (2012). Using Effect Size-or Why the P Value Is Not Enough". *Journal of Graduate Medical Education*. September 2012, Vol. 4, No. 3, pp. 279-282. Recuperado el 25 de mayo de 2020, en <https://doi.org/10.4300/JGME-D-12-00156.1>
- Thorndike, R. L. y Hagen, E. P. (2012). *Medición y evaluación en psicología y educación*. México: Trillas.

Sobre los autores

Ennio Héctor Carro Pérez

Licenciado en Psicología Social, por la Universidad Autónoma Metropolitana (UAM). Doctor en Aprendizaje y Cognición por la Universidad de Sevilla, España. Coordinador del Centro de Investigación y Desarrollo Tecnológico Aplicado al Comportamiento (CIDETAC) de la Facultad de Derecho y Ciencias Sociales, de la Universidad Autónoma de Tamaulipas (UAT), Editor General de la Revista de Psicología y Ciencias del Comportamiento de esta misma casa de estudios. Miembro del Sistema Nacional de Investigadores, y del Cuerpo Académico “Investigaciones Sociales en Contextos Diversos” UAT-CA-109. Su investigación se ha centrado principalmente en Metodología básica de las ciencias del comportamiento y meta análisis en Salud, comportamiento preventivo y de riesgo.

Arturo Secundino Hernández Gómez

Doctorado en Comunicación Educativa por la Universidad de Baja California (UBC). Maestría en Publicidad por el Centro Avanzado de Comunicación (CADEC); Maestría en Educación Superior y Docencia por la Universidad del Noreste (UNE) y, Licenciado en Comunicación por la Universidad Autónoma de Tamaulipas (UAT). Miembro del Sistema Nacional de Investigadores. Líder del Cuerpo Académico “Investigaciones Sociales en Contextos Diversos” UAT-CA-109; Colaborador editorial de la Revista de Psicología y Ciencias del Comportamiento. Profesor de Tiempo Completo en la carrera de Comunicación de la Facultad de Derecho y Ciencias Sociales de la UAT. Líneas de investigación en comunicación para la salud, meta análisis y educación.

Isaías Martínez Trejo

Doctor en Gestión Estratégica de Negocios por la Universidad Autónoma de Tamaulipas, avalado en el PNPC-CONACYT. Maestro en Comercio Exterior, Contador Público Auditor de la Facultad de Comercio y Administración de Tampico de la UAT. Profesor de la Facultad de Derecho y Ciencias Sociales. Líneas de Investigación relacionadas con la Educación Superior, Gestión y Tecnología. Coautor del libro *M-LEARNING: el aprendizaje a través de la tecnología móvil en la universidad*. Integrante de la Comisión Dictaminadora en la zona sur de la Universidad Autónoma de Tamaulipas.

Arturo Sahagún Morales

Licenciado en psicología por la Universidad Autónoma de Tamaulipas (UAT). Fue becado como tesista en el proyecto “Rumores y creencias colectivas sobre la salud y la enfermedad en habitantes con distinto perfil sanitario” del cuerpo académico Investigaciones Jurídicas y Sociales, subvencionado por la UAT. Es doctorante en Psicología Social y Ambiental

por la Universidad Nacional Autónoma de México (UNAM) desde el 2017, siendo becado por el Consejo Nacional de Ciencia y Tecnología (Conacyt). Es colaborador del Centro de Investigación y Desarrollo Tecnológico Aplicado al Comportamiento, de la Facultad de Derecho y Ciencias Sociales de la UAT desde hace 8 años, trabajando en distintos proyectos de investigación, difusión y divulgación científicas, entre los que destacan la co-gestión editorial de la Revista de Psicología y Ciencias del Comportamiento (RPCC-UACJS) y del Museo de Ciencias del Comportamiento (MUCCOM). Ha impartido clases y cursos a nivel de licenciatura desde 2016, impartiendo materias como Métodos de investigación 1 y 2, Estadística inferencial, Proyectos de Inversión, Introducción a la triada Diseño de investigación - medición - análisis estadístico en psicología, Análisis estadístico y alcances de los diseños $N = 1$, entre otras.

Jocelyn Pamela Castelán Félix

Licenciada en Psicología por la Universidad Autónoma de Tamaulipas. Miembro del Centro de Investigación y Desarrollo Tecnológico Aplicado al Comportamiento de la Facultad de Derecho y Ciencias Sociales de la Universidad Autónoma de Tamaulipas. Colaborador editorial de la Revista de Psicología y Ciencias del Comportamiento.

Oscar Eliezer Mendoza de los Santos

Licenciado en Psicología por la Universidad Autónoma de Tamaulipas. Estudiante de la Maestría en Metodología de la Ciencia en el Centro de Investigaciones Económicas, Administrativas y Sociales del Instituto Politécnico Nacional. Miembro del Centro de Investigación y Desarrollo Tecnológico Aplicado al Comportamiento de la Facultad de Derecho y Ciencias Sociales de la Universidad Autónoma de Tamaulipas. Colaborador editorial de la Revista de Psicología y Ciencias del Comportamiento.

Amada Ampudia Rueda

Profesora de carrera de tiempo completo de la Facultad de Psicología de la UNAM, docente en los programas de licenciatura, especialidad, maestría y doctorado. Coordinadora de proyectos sobre maltrato infantil, perfil de agresores, delincuentes, violencia, problemas de adicciones, así como consumo de alcohol y manejo automovilístico. Miembro del Sistema Nacional de Investigadores (SNI).

Contribuciones a los métodos aplicados en la investigación en psicología y ciencias de la comunicación de Ennio Héctor Carro Pérez y Arturo Secundino Hernández Gómez, coordinadores. Publicado por la Universidad Autónoma de Tamaulipas y Colofón, se terminó de imprimir en marzo 2021 en los talleres de Litográfica Ingramex S.A. de C.V., Centeno 162-1, Granjas Esmeralda, 09810, Ciudad de México, México. El tiraje consta de 350 ejemplares impresos de forma digital en papel Cultural de 75 gramos, exteriores sobre cartulina Lustrolito brillante de 300 gramos. El cuidado estuvo a cargo del Consejo de Publicaciones UAT.

