

ARTÍCULOS ORIGINALES

REVISTA ARGENTINA DE SALUD PÚBLICA

FECHA DE RECEPCIÓN: 23 de agosto de 2021

FECHA DE ACEPTACIÓN: 10 de diciembre
de 2021

FECHA DE PUBLICACIÓN: 12 de abril de 2022

FUENTE DE FINANCIAMIENTO: Ninguna.

*AUTOR DE CORRESPONDENCIA:

santiago.torales@gmail.com

Registro Nacional de Investigaciones en
Salud N°: IS003551

ENCUESTA DE OPINIÓN Y SATISFACCIÓN A LOS PARTICIPANTES DE LA CONVOCATORIA SALUD INVESTIGA SOBRE EL PROCESO DE EVALUACIÓN DE LAS PROPUESTAS DE INVESTIGACIÓN

Survey of applicant's opinion and satisfaction regarding the evaluation process of research proposals for "Salud Investiga" grants call

* Santiago Blas Torales¹. Magíster y Especialista en Economía y Gestión de la Salud.

Manuel Lago¹. Magíster en Efectividad Clínico Sanitaria con orientación en Investigación y Gestión.

Oscar Gabriel Ianovsky¹. Magíster en Efectividad Clínico Sanitaria con orientación en Investigación y Gestión.

¹ Dirección de Investigación en Salud, Secretaría de Acceso a la Salud, Ministerio de Salud de la Nación, Argentina.

RESUMEN. INTRODUCCIÓN: La revisión por pares de propuestas de investigación es el "estándar de oro" utilizado por la mayoría de las agencias que financian investigación en salud para juzgar su calidad. Sin embargo, este proceso ha recibido numerosas críticas de la comunidad internacional. La Dirección de Investigación en Salud financia propuestas de investigación mediante la convocatoria a becas Salud Investiga, utilizando para la asignación de mérito el procedimiento de revisión por pares. En línea con lo ocurrido con otras agencias financiadoras públicas, el proceso ha sido blanco de objeciones informales por parte de la comunidad de investigación local. Hasta la fecha no se ha relevado de manera formal la opinión respecto del proceso de evaluación. El objetivo fue evaluar, en los postulantes a la Convocatoria 2020, la opinión sobre la experiencia con los revisores, la constructividad, claridad y lenguaje de los comentarios, la satisfacción global con la revisión y la confiabilidad de las puntuaciones. MÉTODOS: Encuesta electrónica voluntaria autoadministrada. RESULTADOS: La tasa de respuesta fue del 39%. Se observó mayoría de acuerdo con aquellos ítems que expresaban opiniones/percepciones positivas respecto del proceso de revisión por pares y mayoría de desacuerdo con los ítems que reflejaban opiniones/percepciones negativas. DISCUSIÓN: Predominó la satisfacción con el proceso de revisión, aunque su grado difirió marcadamente entre los que obtuvieron una beca y aquellos que no.

PALABRAS CLAVE: Encuestas y Cuestionarios; Revisión por Pares; Proyectos de Investigación; Salud Pública; Financiación Gubernamental; Argentina

ABSTRACT. INTRODUCTION: Peer review of research proposals is the "gold standard" used by most agencies that fund health research to judge its quality. However, this process has received much criticism from the international research community. The "Dirección de Investigación en Salud" funds research proposals through calls for "Salud Investiga" grants, using the peer review procedure for the assignment of merit. In line with other public funding agencies, the process has received informal objections from the local research community. To date, the opinion regarding the evaluation process has not been formally evaluated. The objective was to evaluate the opinion of applicants for 2020 call about the experience with reviewers, the constructiveness, clarity and language of the comments, the overall satisfaction with the review and the reliability of the scores. METHODS: Self-administered voluntary electronic survey. RESULTS: The response rate was 39%. A majority of agreement was observed with those items that reflected positive perceptions/opinions regarding the peer review process, as well as a majority of disagreement with items that reflected negative perceptions/opinions. DISCUSSION: Overall, satisfaction with the process prevailed, although the degree varied significantly between those who obtained the scholarship and those who did not.

KEY WORDS: Surveys and Questionnaires; Peer Review; Research Projects; Public Health; Government Financing; Argentina

INTRODUCCIÓN

La excelencia en la investigación depende, en parte, de la calidad de los procedimientos utilizados para seleccionar las propuestas de investigación a ser financiadas. Las instituciones públicas que financian investigación se enfrentan al desafío de formular y mantener los mejores procedimientos para evaluar la calidad y el potencial de las propuestas. Se trata de una tarea exigente, ya que cada propuesta es científicamente única y se origina en diversas culturas de investigación¹.

La revisión por pares de propuestas es el “estándar de oro” para garantizar la calidad de las investigaciones e implica involucrar a revisores expertos en el proceso de decidir qué solicitudes se recompensan con apoyo financiero. Si bien este proceso comparte aspectos con la revisión por pares de manuscritos de investigación, difiere en los atributos que evalúa, la perspectiva y principalmente en los objetivos: en la revisión editorial, informar a los comités respectivos del valor de una investigación para merecer su publicación y, en la revisión de proyectos, informar al órgano financiador del valor y potencial de una investigación para producir conocimiento¹. El procedimiento de revisión por pares, utilizado por la mayoría de las agencias financiadoras para seleccionar el mejor *portfolio* de investigaciones, asegura que el conjunto de propuestas escogidas sean del más alto mérito científico y que la investigación satisfaga las necesidades de conocimiento y permita informar prácticas, programas o políticas. Por lo tanto, es necesario que la revisión por pares sea efectiva, eficiente y sostenga la producción de investigaciones de alta calidad².

A pesar de ser considerado el estándar para la evaluación de propuestas de investigación, el sistema de evaluación por pares ha recibido numerosas críticas por parte de la comunidad de investigación internacional. Dichas críticas se han centrado en la poca transparencia del proceso, la baja experiencia de los revisores con relación al tema del proyecto, la pobre correlación de las puntuaciones entre revisores de un mismo proyecto y la tendencia a penalizar propuestas innovadoras, riesgosas o de frontera del conocimiento, así como aquellas postuladas por investigadores jóvenes³. Como consecuencia de estas objeciones, las agencias financiadoras internacionales han abordado desde hace más de una década la reforma de los procesos de revisión. Un estudio de casos de financiadores internacionales efectuado por la corporación RAND Europa, que fue publicado en 2018 e involucra a seis grandes financiadores —el Instituto Canadiense de Investigación en Salud (CIHR), los Institutos Nacionales de Salud de EE.UU. (NIH), el Consejo Médico de Investigación de Gran Bretaña (MRC), el Consejo de Investigación y Salud Australiano (NHMRC), el Instituto Nacional de Investigación en Salud de Gran Bretaña (NIHR) y la Organización para la Investigación y el Desarrollo en Salud de los Países Bajos (NOHRD)—, revela las iniciativas y reformas desarrolladas por dichas agencias para mejorar el proceso de revisión por pares y explicita los insumos utilizados para informar los procesos de reforma,

que incluyen encuentros y talleres con la comunidad de investigación, paneles de revisión, auditorías, sistemas de monitoreo y encuestas a los postulantes, revisores y directivos de las agencias financiadoras^{3,4}.

La Dirección de Investigación en Salud (DIS) del Ministerio de Salud de la Nación (MSAL) de Argentina, ex Comisión Nacional Salud Investiga, financia desde hace 19 años propuestas de investigación en Salud Pública. A lo largo de este periodo ha utilizado distintas estrategias para priorizar los lineamientos, decidir las distintas modalidades de financiación y definir el proceso de evaluación (evaluación individual por revisores independientes, o la anterior más evaluación por panel de expertos). La DIS cuenta con un *pool* de evaluadores independientes, identificados a partir de su experiencia y publicaciones, los cuales son invitados a evaluar las propuestas en ocasión de cada convocatoria y en forma *ad honorem*, de manera tal que donan una importante parte de su tiempo a este servicio. Para ello se les asignan propuestas alineadas con su área de experiencia, completan una grilla de evaluación y emiten un puntaje y comentarios. Todos los proyectos de investigación son sometidos a un procedimiento de revisión efectuado por dos expertos, luego se elabora un ranking promediando las puntuaciones adjudicadas por cada revisor y las subvenciones se otorgan según el orden de mérito resultante de ese proceso. Esta secuencia de revisión, en línea con lo ocurrido a otras agencias financiadoras públicas, ha recibido objeciones informales de la comunidad de investigación local. Hasta la fecha no se ha relevado de manera formal una opinión de los postulantes respecto de este proceso. Por ello, en sintonía con lo efectuado por otras agencias financiadoras, la DIS planificó una autoevaluación con el propósito de obtener información que permita guiar intervenciones de mejora en el proceso, tomando como referencia esta última convocatoria. Como parte de dicho diagnóstico se elaboró una encuesta con el objetivo principal de evaluar la opinión de los postulantes a la Convocatoria Salud Investiga (SI) 2020 respecto de la experiencia de los revisores, la confiabilidad, constructividad, claridad y lenguaje de los comentarios y la satisfacción global con la revisión. Adicionalmente, como objetivo secundario, se procedió a evaluar la confiabilidad de las puntuaciones entre revisores.

MÉTODOS

Población

Se consideró población blanco a todos los postulantes a la Convocatoria SI 2020-2021 que alcanzaron la etapa de evaluación por pares. El marco de muestreo utilizado fue la lista de registrados como postulantes que aprobaron la evaluación administrativa y de pertinencia con las líneas temáticas de la convocatoria.

Diseño del instrumento de encuesta y descripción de la fuente del registro de puntuaciones

Para la elaboración de esta encuesta se efectuó una

búsqueda de artículos publicados en inglés y castellano en Google Scholar, Biblioteca Virtual en Salud, SciELO, PubMed y Tripdatabase, con los términos de búsqueda "grant peer review", "grant peer review survey", "revisión por pares", "revisión de proyectos", "revisión de propuestas", "investigación", "research proposals", "salud", "survey", "encuesta", "quality", "evaluation" y "evaluación", utilizando los operadores AND y OR, mapeo automático de términos y también con la estrategia de búsqueda "bola de nieve" y el "artículo perla".

Se recuperaron 139 citas; luego de eliminar duplicados, se seleccionaron solo aquellas que referían a encuestas de evaluación del sistema de revisión por pares. Se accedió al texto completo de las citas y se eligieron solo aquellas que publicaban el instrumento de encuesta. Se seleccionaron finalmente 2 instrumentos de encuesta^{1,5}. Luego de revisar y evaluar los dominios y los ítems de dichas encuestas, se escogió un grupo de ítems, los cuales fueron traducidos del inglés original al español. Finalmente, después de varias iteraciones entre los miembros del equipo coordinador de la Dirección a cargo de la investigación, se seleccionaron un conjunto de ítems priorizados que pertenecían a los dominios de opinión, percepciones y satisfacción de los postulantes. Se incorporaron algunos ítems adicionales a partir de sugerencias del equipo. El instrumento final quedó compuesto por 15 ítems, en escala de Likert, destinados a relevar la percepción/opinión de los postulantes encuestados respecto de: a) Pertinencia de la evaluación (es decir, oportunidad, adecuación y conveniencia) con los ítems: "En la evaluación se juzgaron aspectos ajenos al manuscrito del proyecto"; "En la evaluación se expresan opiniones personales no basadas en la evidencia"; "Los comentarios reflejan que el o los evaluadores entendieron el significado de mi propuesta"; "Globalmente, los comentarios fueron pertinentes"; b) Percepción de la experiencia de los evaluadores con los ítems: "La evaluación efectuada refleja conocimiento experto por parte de los evaluadores"; c) Constructividad de los comentarios (comentarios tendientes a mejorar la propuesta): "En la evaluación se comentan las fortalezas y debilidades de los métodos propuestos"; "En la evaluación se identificaron las fortalezas y debilidades de mi propuesta"; "La evaluación efectuada me ayudará a mejorar mi propuesta"; d) Uso de lenguaje inadecuado por parte del revisor (entendido como lenguaje ofensivo o vulgar, o referido al autor y no a la propuesta) con el ítem: "La evaluación presenta lenguaje inadecuado"; e) Claridad y exhaustividad de los comentarios: "La evaluación fue clara y fácil de leer"; "La evaluación estaba estructurada y organizada"; "La evaluación fue detallada y completa"; f) Consistencia entre evaluadores con los ítems: "Los comentarios de los evaluadores acerca de mi propuesta fueron muy diferentes"; "Los comentarios de los evaluadores acerca de mi propuesta fueron contradictorios entre sí"; y g) Satisfacción global con la evaluación: "Globalmente ¿cuán satisfecho o insatisfecho está Ud. con la evaluación

efectuada a su propuesta de investigación?". De los 15 ítems incluidos, 10 se presentaron como afirmaciones positivas y 5 como afirmaciones negativas, más un ítem con campo abierto en donde se solicitó a los postulantes expresar sus opiniones y sugerencias respecto de la Convocatoria en general y del proceso de evaluación en particular, contenido que tras su correspondiente análisis será motivo de otro reporte. Se incorporaron, además, como variables a relevar: edad, sexo, línea de investigación postulada y estado de ganador de la beca. Para el diseño operacional de la encuesta se utilizó la aplicación basada en la web REDCap (*Research Electronic Data Capture*), que permitió la captura y posterior almacenamiento de los datos.

Para satisfacer el objetivo secundario se utilizaron los registros de las puntuaciones adjudicadas por los revisores a cada propuesta de investigación, los cuales se encontraban almacenados en la base de datos de la Convocatoria.

Muestra

Para incrementar la participación se envió la invitación y luego, un recordatorio. Una carta de cubierta informaba a los receptores del correo respecto del resguardo de la información vertida y su tratamiento confidencial. Uno de los investigadores, que no participó del análisis, identificó y consignó el estado de ganador o no ganador de la beca y luego desvinculó la información de identidad de la tabla de datos. De tal modo, los datos fueron anonimizados y los responsables del análisis desconocieron la identidad de los participantes.

Se identificó a los postulantes utilizando como fuente los registros de la Convocatoria. El 23 de noviembre de 2020 se enviaron por correo electrónico enlaces de REDCap a los 318 postulantes que habían accedido a la instancia de evaluación por pares. La encuesta se mantuvo activa hasta el 3 de diciembre de 2020.

Para el objetivo secundario se identificaron las puntuaciones asignadas por cada revisor a cada una de las propuestas presentadas en la Convocatoria, las cuales fueron tratadas estadísticamente.

Análisis estadístico

Una vez cerrada, la base de datos fue analizada para identificar datos incompletos, incorrectos, inexactos o no pertinentes, y ser depurada a continuación.

Los resultados de acuerdo y satisfacción se expresaron como el porcentaje de la fuerza de acuerdo con cada una de las afirmaciones, mediante una gráfica de barras divergentes de tamaño proporcional a la categoría de acuerdo.

Para la evaluación de la confiabilidad de las puntuaciones se efectuó análisis de correlación-concordancia de Lin y del coeficiente de correlación intraclase.

Los análisis se efectuaron mediante los paquetes tm (versión 0.7-8) y Likert (versión 1.3.5) del *software R* (versión 3.6.2, R Development Core Team/R Foundation for Statistical Computing, Viena, Austria), Concord (Stata 12).

El protocolo de investigación que originó el manuscrito presentado no requirió aprobación de un Comité de Ética de Investigación por tratarse de una evaluación del funcionamiento de un programa/servicio del Ministerio de Salud.

RESULTADOS

Se recibieron 124 encuestas (tasa de respuesta 39%), de las cuales 4 carecían de respuestas, por lo que no se consideraron válidas para su análisis. De las 120 respuestas válidas, 28 (26%) eran de investigadores que obtuvieron becas. La Figura 1 presenta la frecuencia de comentarios de los evaluadores. Un 4% de los encuestados respondieron que no habían recibido comentario alguno por parte de los evaluadores.

La Tabla 1 refleja la distribución de respuestas por edad, sexo y tipo de postulación.

Todos los gráficos de barras divergentes presentan en los bordes derecho e izquierdo del área los porcentajes acumulados de respuestas que expresan acuerdo y desacuerdo, respectivamente. En el total de encuestados, se observó más acuerdo que desacuerdo con aquellos ítems que reflejaban opiniones o percepciones positivas respecto del proceso de revisión por pares: por ejemplo, "La evaluación estaba estructurada y organizada", 51% de acuerdo vs. 35% de desacuerdo; "La evaluación efectuada refleja conocimiento experto por parte de los evaluadores", 56% vs. 32%; "En la evaluación se identificaron las fortalezas y debilidades de mi propuesta", 63% vs. 31%; "Globalmente, los comentarios fueron pertinentes", 65% vs. 28%; "La evaluación efectuada me ayudará a mejorar mi propuesta", 61% vs. 26% (ver Figura 2).

Se observó, en la totalidad de encuestados, más desacuerdo que acuerdo con los ítems que reflejaban percepciones u opiniones negativas respecto de la revisión por pares: "En la evaluación se juzgaron aspectos ajenos al manuscrito del proyecto", 63% de desacuerdo vs. 23% de acuerdo; "En la evaluación se expresan opi-

niones personales no basadas en la evidencia", 59% vs. 27%; "Los comentarios de los evaluadores acerca de mi propuesta fueron contradictorios entre sí", 51% vs. 27% (ver Figura 3).

En relación con la satisfacción global, un 53% de los encuestados expresó que se encontraba "muy" o "algo satisfecho" con la evaluación, mientras que un 41% manifestó insatisfacción respecto a ella (ver Figura 4).

Se analizó la respuesta luego de particionar la muestra entre los que obtuvieron becas y quienes no accedieron a ellas. En el grupo que obtuvo becas se observó un grado más importante de acuerdo que desacuerdo con todos los ítems que reflejaban opiniones o percepciones positivas, comparado con el observado en toda la muestra: a modo de ejemplo, "La evaluación estaba estructurada y organizada", 54% de acuerdo vs. 21% de desacuerdo; "La evaluación efectuada refleja conocimiento experto por parte de los evaluadores", 82% vs. 7% (ver Figura 5). A su vez, en el grupo que no obtuvo becas se observó un grado más importante de desacuerdo que acuerdo con todos los ítems que reflejaban opiniones o percepciones negativas, comparado con el observado en toda la muestra: a modo de ejemplo, "En la evaluación se juzgaron aspectos ajenos al manuscrito del proyecto", 88% de desacuerdo vs. 4% de acuerdo; "En la evaluación se expresan opiniones personales no basadas en la evidencia", 82% vs. 4%; "Los comentarios de los evaluadores acerca de mi propuesta fueron contradictorios entre sí", 78% vs. 9% (ver Figura 6).

En el grupo que no obtuvo becas, los ítems que reflejaban opiniones o percepciones positivas mostraron afirmaciones con más acuerdos que desacuerdos (por ejemplo, "La evaluación efectuada refleja conocimiento experto por parte de los evaluadores", 48% vs. 41%) y otras con más desacuerdos que acuerdos (por ejemplo, "La evaluación fue detallada y completa", 39% vs. 48%, respectivamente) (ver Figura 7).

En el grupo que no obtuvo becas, predominaron los

FIGURA 1. Presencia de comentarios de los evaluadores.

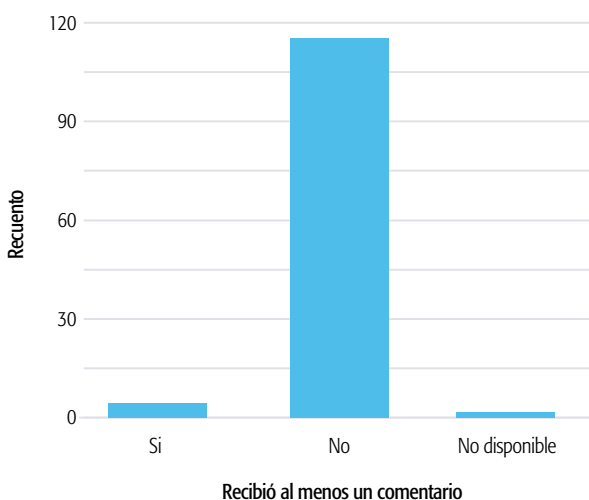


TABLA 1. Características demográficas y de la postulación.

Característica	N (%)
Mujeres	80 (67)
Edad	
20-29	14 (12)
30-39	32 (27)
40-49	41 (34)
≥50	33 (30)
Tipo de beca	
1) Individual, Clínica, área general	9 (7,5)
2) Individual, Clínica, COVID-19	14 (12)
3) Individual, Salud Pública, área general	15 (13)
4) Individual, Salud Pública, COVID-19	27 (23)
5) Estudios Multicéntricos, investigación en pandemia COVID-19 con enfoque a servicios de salud	17 (15)
6) Estudios Multicéntricos, investigación en pandemia COVID-19 con enfoque a determinantes sociales	16 (14)

FIGURA 2. Acuerdo con afirmaciones positivas, total de encuestados.

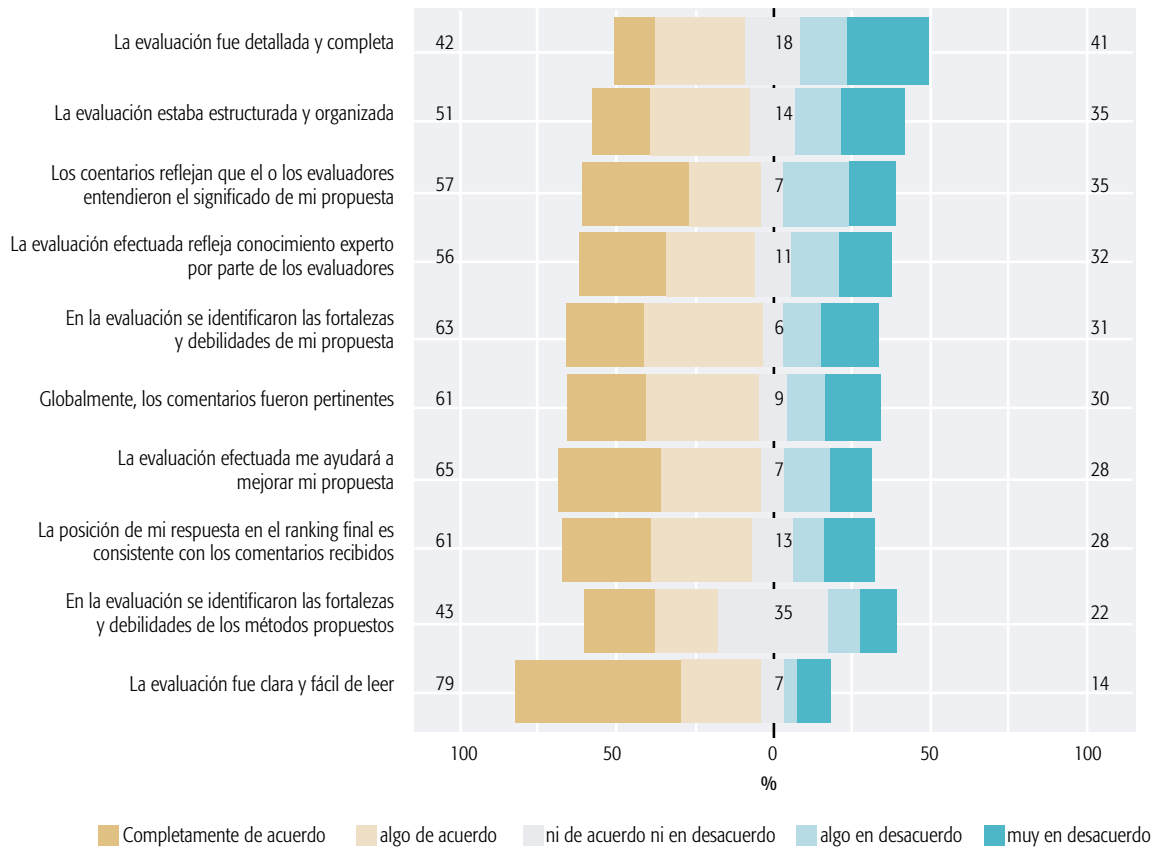


FIGURA 3. Acuerdo con afirmaciones negativas, total de encuestados.

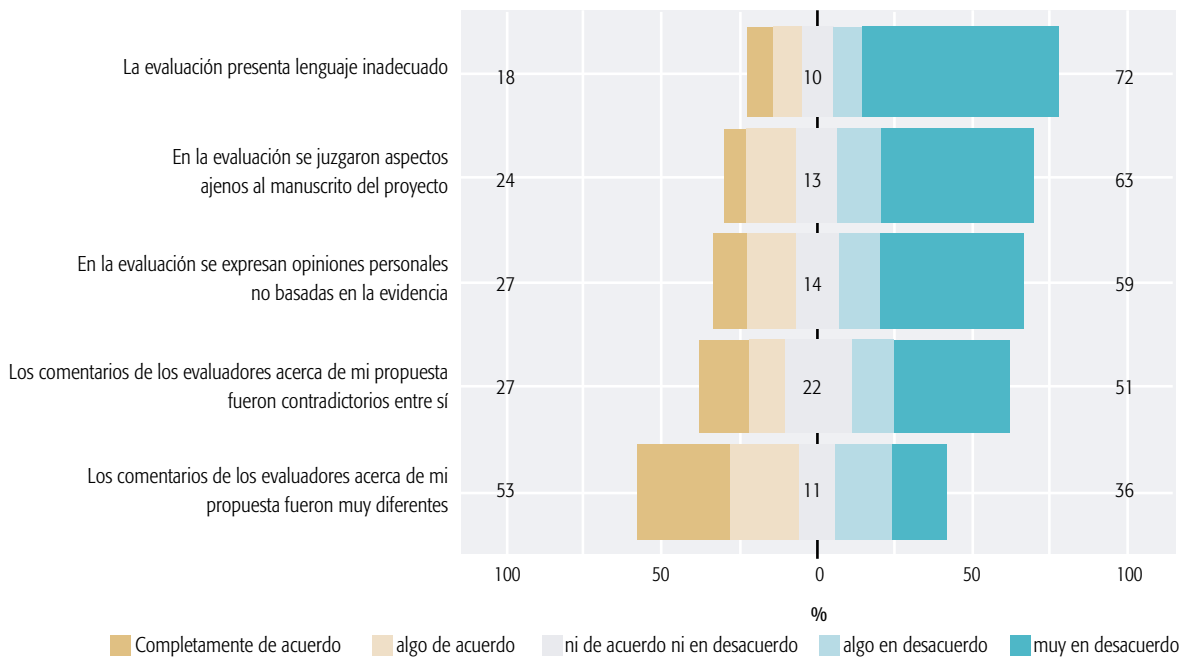


FIGURA 4. Satisfacción global con la evaluación, total de encuestados.

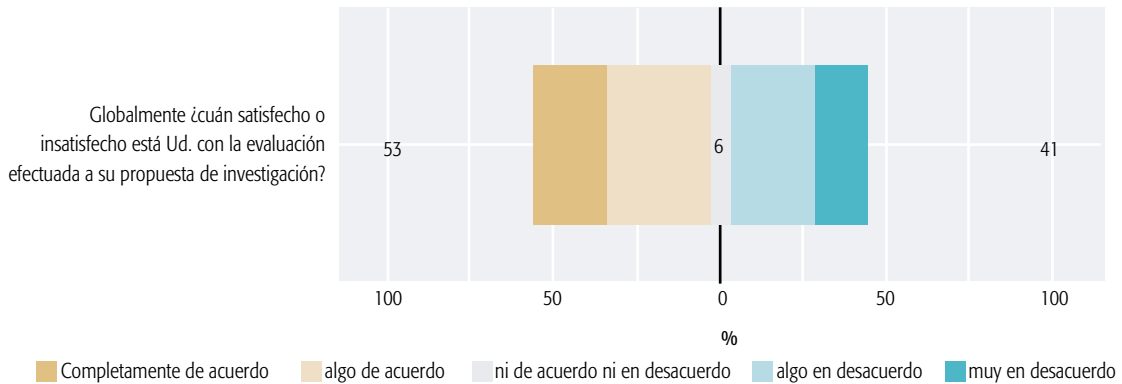


FIGURA 5. Grado de acuerdo o desacuerdo con afirmaciones positivas, grupo que obtuvo becas.

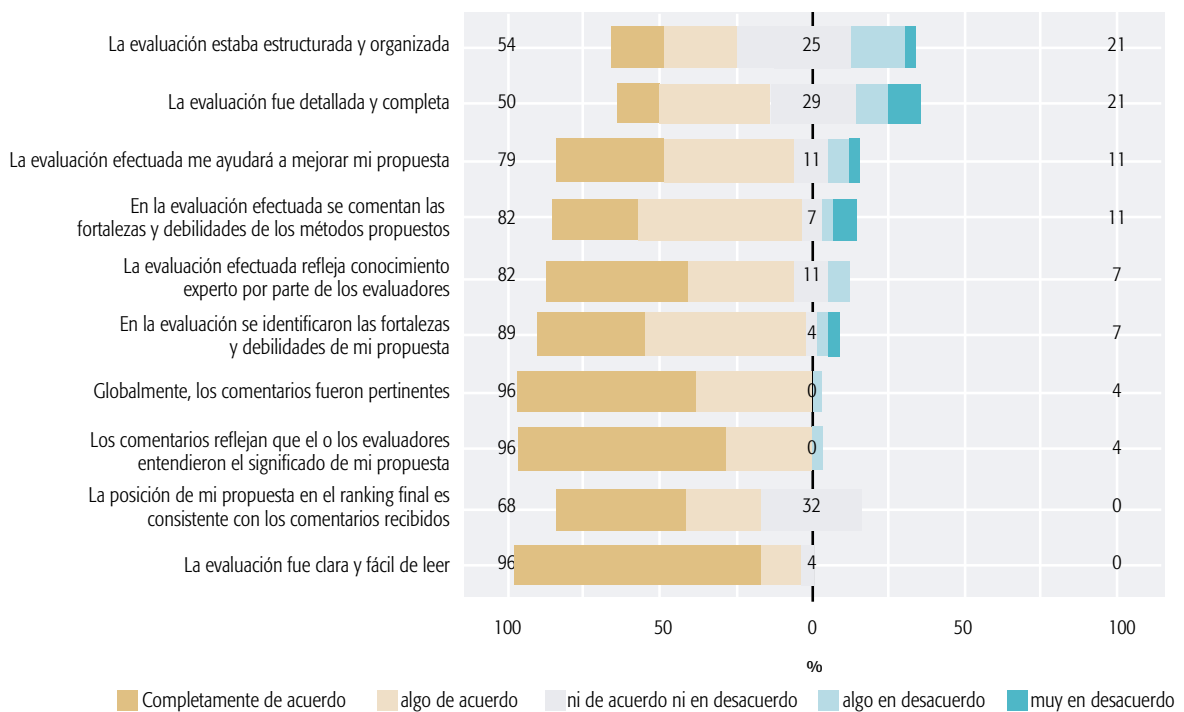
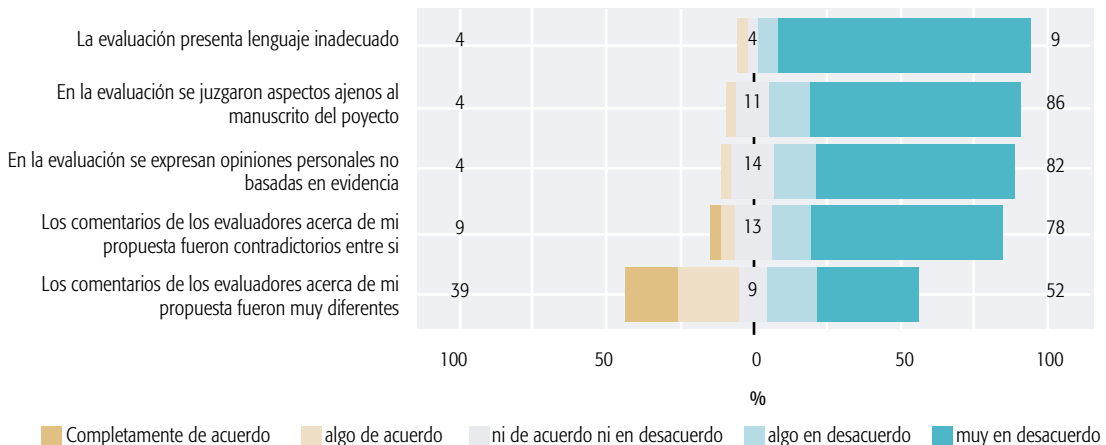


FIGURA 6. Grado de acuerdo con afirmaciones negativas, grupo que obtuvo becas.



desacuerdos en los ítems que reflejaban opiniones o percepciones negativas (por ejemplo: "Los comentarios de los evaluadores acerca de mi propuesta fueron contradictorios entre sí", 42% vs. 34% (ver Figura 8).

Las Figuras 9 y 10 reflejan el grado de satisfacción o insatisfacción con la evaluación recibida por parte de quienes obtuvieron y no obtuvieron becas, respectivamente. Para la evaluación de la confiabilidad entre revisores de los puntajes adjudicados a cada proyecto, se procesaron 321 observaciones apareadas. La media de las puntuaciones fue de 645. La Figura 11 presenta una gráfica de dispersión, y otra con el acuerdo y sus límites, observándose: Correlación Concordancia [IC95%] rho_c; 0,377 [0,283; 0,472

]; Coef. de Corr. Intraclase X *Boostraping* media[IC95%]: 0,379 [0,271; 0,477]

DISCUSIÓN

La tasa de respuesta (39%) fue similar a la observada en encuestas de características parecidas: la de los Institutos Nacionales de Salud de los EE.UU. sobre revisión por pares (*NIH Peer Review*), efectuada a postulantes en 2015 y publicada en 2017, tuvo una tasa de respuesta del 36%⁵. En este caso se trató de una encuesta dirigida a investigadores de los cuales un subconjunto participa reiteradamente de las convocatorias y a los cuales se les está solicitando opinión para mejorar un sistema que los afecta directamente. La tasa

FIGURA 7. Grado de acuerdo con afirmaciones positivas, grupo que no obtuvo becas.

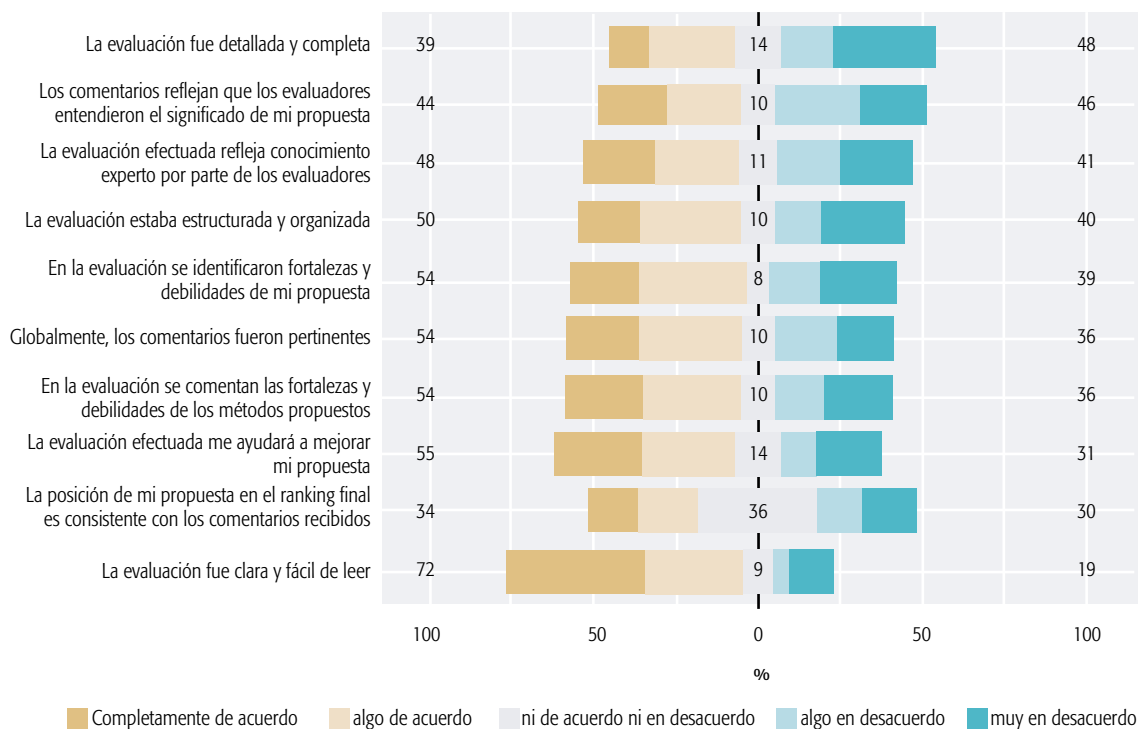


FIGURA 8. Grado de acuerdo con afirmaciones negativas, grupo que no obtuvo becas.

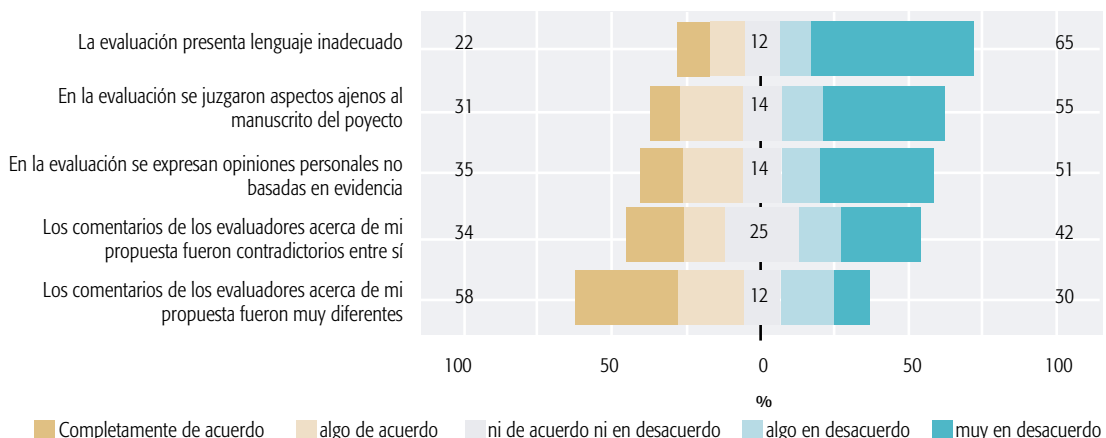


FIGURA 9. Satisfacción global, grupo que obtuvo becas.

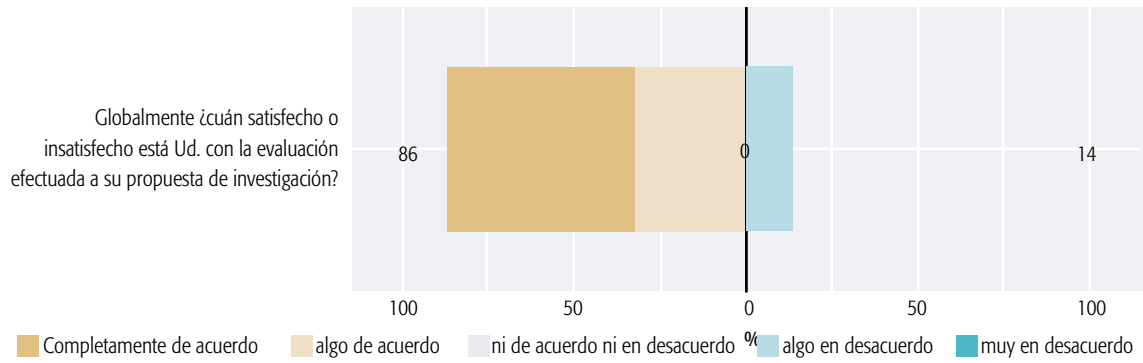


FIGURA 10. Satisfacción global, grupo que no obtuvo becas.

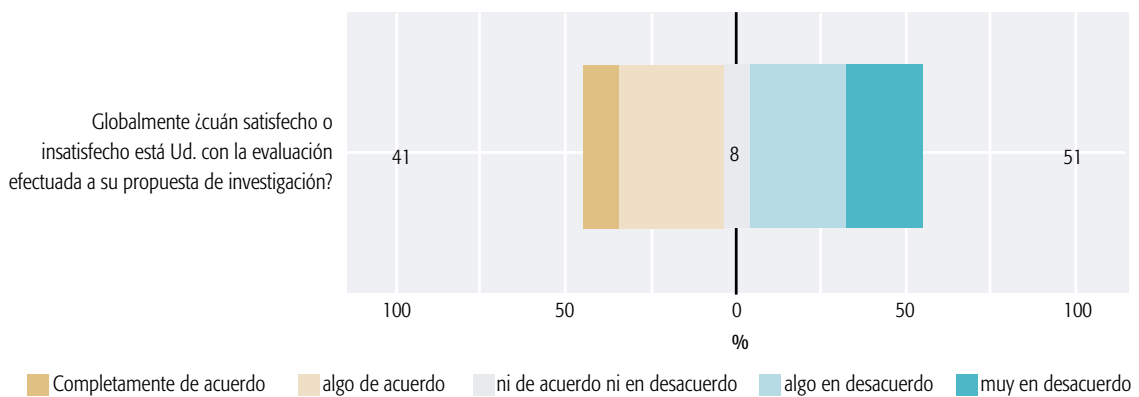
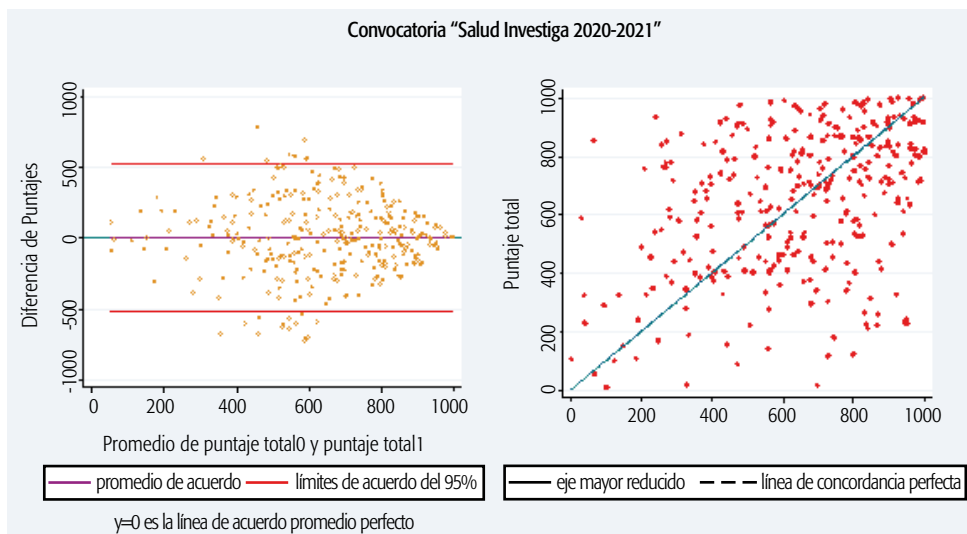


FIGURA 11. Evaluación de la confiabilidad de las puntuaciones.



de respuesta observada podría reflejar cierto desinterés y escepticismo respecto a la posibilidad de modificar aspectos del proceso de revisión por pares.

Globalmente hubo un porcentaje mayor de acuerdo que de desacuerdo con los ítems que reflejaban percepciones u opiniones positivas respecto de la revisión. Se observó un porcentaje mayor de desacuerdo que de acuerdo en aquellos ítems que reflejaban percepciones u opiniones negativas (Figuras 2 y 3), situación que se polarizó al estratificar por la obtención o no de becas (Figuras 6 y 7), lo que posiblemente evidencia el efecto de la recompensa⁶.

Específicamente, en relación con los ítems *“Los comentarios de los evaluadores acerca de mi propuesta fueron muy diferentes”* o *“Los comentarios de los evaluadores acerca de mi propuesta fueron contradictorios entre sí”*, hubo un 58% de acuerdo en que los comentarios fueron juzgados como muy diferentes y 34% de acuerdo en que fueron evaluaciones contradictorias (Figuras 2 y 3). Otras encuestas también encontraron hallazgos similares y los adjudicaron a la diversidad de áreas de formación de los evaluadores asignados y su distinta perspectiva¹.

Respecto de los ítems *“La evaluación fue detallada y completa”* y *“La evaluación estaba estructurada y organizada”*, se identificaron como aspectos relevantes la ausencia de comentarios por parte de algunos revisores, el bajo detalle y estructura de la evaluación y la inconsistencia entre las revisiones. Se observó un desacuerdo del 41% con el ítem que expresaba que la evaluación era detallada y exhaustiva, y un desacuerdo del 35% con el ítem que expresaba que los comentarios estaban estructurados y organizados. Entre los ganadores, ambos ítems fueron los que recibieron la mayor proporción de desacuerdos (21%) con relación al resto de los ítems con connotación negativa, lo que refleja un aspecto de la evaluación que resulta problemático y requiere solución. Las respuestas podrían obedecer a déficits en la selección de revisores, al escaso tiempo disponible de los revisores para una evaluación más detallada, a una falta de incentivo general para la revisión, a herramientas de evaluación inapropiadas o a una combinación de estos problemas.

La mayoría de los postulantes cuya propuesta fue financiada (82%) acordó que había sido evaluada por un revisor con conocimiento experto. Aquellos postulantes cuya propuesta no fue financiada expresaron un acuerdo mucho menor (42%). El nivel global de acuerdo con dicho ítem (*“La evaluación efectuada refleja conocimiento experto por parte de los evaluadores”*) fue del 56% (Figura 2). A modo de comparación, una pregunta similar en la encuesta del NIH sobre peer review publicada en 2017⁵ (*“my application was evaluated by reviewers with the appropriated expertise”*) reveló acuerdo del 91% y 53% entre propuestas finalmente financiadas o no, respectivamente, lo que muestra cierta similitud en el patrón de respuestas inducido por haber obtenido o no la beca y, según los autores, refleja el efecto de la recompensa sobre la opinión.

La mayoría de los postulantes cuya propuesta fue

financiada (96%) acordaron que el evaluador la había entendido. Aquellos postulantes cuya propuesta no accedió a la financiación acordaron en un 44% y estuvieron en desacuerdo en un 46% con dicho ítem (*“Los comentarios reflejan que el o los evaluadores entendieron el significado de mi propuesta”*). Hubo un 57% de acuerdo global. Ítems similares de la encuesta del NIH (*“The reviewers understood the significance of the proposed research”* y *“The reviewers understood my proposed approach”*) evidenciaron respectivamente 62% y 57% de acuerdo.

Los tres ítems referidos a la constructividad de los comentarios de los evaluadores —1) *“En la evaluación se identificaron las fortalezas y debilidades de mi propuesta”*, 2) *“En la evaluación se comentan las fortalezas y debilidades de los métodos propuestos”* y 3) *“La evaluación efectuada me ayudará a mejorar mi propuesta”*— mostraron respuestas que manifestaron acuerdo y desacuerdo con cada afirmación de 63% y 31%, de 61% y 30% y de 61% y 26%, respectivamente (Figura 2), con marcada polarización al estratificar entre ganadores (89%, 82% y 79% de acuerdo y 7%, 11% y 11% de desacuerdo) y no ganadores (54%, 54% y 55% de acuerdo y 39%, 36% y 31% de desacuerdo, respectivamente), lo que refleja un efecto marcado de la recompensa.

En relación con el ítem de la consistencia entre los comentarios recibidos y la posición final en el ranking, hubo un acuerdo de 43% y un desacuerdo de 22%, así como un porcentaje importante (33%) que no pudo expresar una posición probablemente por carencia de información respecto al ranking. Como era esperable, hubo un 0% de desacuerdo con la afirmación entre los ganadores y un 30% entre los no ganadores. La inconsistencia entre la puntuación y los comentarios es un fenómeno que ya se ha visto en otras convocatorias; se ha juzgado que obedecía a un efecto del instrumento de puntuación, el cual fue reformulado en sucesivas instancias y probablemente requiera una revisión más integral.

En lo que respecta a aquellos ítems que expresaban afirmaciones con connotación negativa, globalmente hubo mayoría de desacuerdo de los postulantes. Se observó un 18% de acuerdo y un 72% de desacuerdo con *“La evaluación presenta lenguaje inadecuado”*, lo cual refleja que el lenguaje utilizado fue apropiado. El 24% de acuerdo y 63% de desacuerdo con el ítem *“En la evaluación se juzgaron aspectos ajenos al manuscrito del proyecto”* y el 27% de acuerdo y 57% de desacuerdo con la afirmación *“Se expresan opiniones personales no basadas en la evidencia”* pueden reflejar cierta inexperiencia o arbitrariedad en la evaluación.

Se observó una satisfacción global moderada con la evaluación: un 53% manifestó que se encontraba “muy” o “algo satisfecho”, mientras que un 41% expresó insatisfacción, lo que refleja la necesidad de efectuar intervenciones para mejorar la calidad de las evaluaciones (Figura 4).

Adicionalmente, un análisis cuantitativo efectuado respecto del acuerdo entre evaluadores de los puntajes

adjudicados reveló una correlación-consistencia un poco por encima de otros estimadores publicados⁷ (Figura 11). En particular, el trabajo de Jayasinghe⁸ demuestra una correlación de confiabilidad de solo 0,21 para las Humanidades y Ciencias Sociales, y una correlación aún más baja (0,19) para las ciencias en general. Del mismo modo, Fogelholm⁹ encontró una confiabilidad de alrededor de 0,23 entre evaluadores en contextos de investigación médica.

La diversidad de antecedentes y disciplinas de los revisores reduce la coherencia de las calificaciones. Lobb identificó un coeficiente de correlación intracase de 0,12 al comparar puntuaciones efectuadas por revisores de distinta formación. Señaló que el nivel de acuerdo entre los expertos de diferentes disciplinas era considerablemente más bajo que entre los jueces de la misma disciplina, lo que significa que la presencia de varios profesionales de la misma área podría sesgar los resultados de la financiación¹⁰.

Si bien en la literatura las opiniones y los esfuerzos apuntan a lograr mayor consistencia en las evaluaciones⁷, no es claro el significado de dichas diferencias y si realmente son un aspecto cuestionable o deseable. Evaluadores de distinta formación reflejan distintas miradas que podrían "compensar" la puntuación, evitando sesgarla hacia una sola perspectiva o marco cognitivo. Para poder optar por una u otra estrategia, es decir, adjudicar una propuesta a evaluadores con igual o distinta formación, se requeriría definir en función de qué desenlace se evalúa y medir cómo impacta la elección de evaluadores de una misma o de distinta formación disciplinaria sobre dicho desenlace.

Aunque se observaron resultados similares a los relevados en otros países, las causas de la limitada tasa de participación deberían ser indagadas. Cerca del 60% de la población blanco de la encuesta no respondió. Dado que se desconocen cuáles son los factores de los sujetos que están efectivamente relacionados con la adherencia a la encuesta y a la propensión a participar en nuevas convocatorias, no es clara la aplicabilidad de los resultados observados a los futuros participantes. Es llamativa, en este contexto, la divergencia entre la frecuencia de la queja informal y la baja propensión a participar; un incremento de la comunicación de la DIS con la comunidad de investigadores podría ayudar a resolver estos problemas.

Una limitación del presente estudio es que se utilizó un instrumento diseñado *ad hoc*, que no fue sometido a un proceso de validación formal en su desempeño. Por lo tanto, resultan desconocidas su precisión y exactitud en la medición de los constructos de interés, así como las implicancias predictivas.

La moderada satisfacción con el proceso de evaluación por pares de los proyectos citados indica la necesidad de comenzar a diseñar un conjunto de intervenciones destinadas a modificar aspectos específicos.

Algunos factores contextuales podrían haber afectado la calidad de la revisión: dada la gran cantidad de propuestas recibidas en la última convocatoria a las becas Salud Investiga y al corto plazo para su evaluación debido a las

vicisitudes epidemiológicas y sociales de la pandemia de COVID-19, el sistema de revisión se ha visto severamente afectado por el incremento del rechazo a revisar por parte de los evaluadores, la consiguiente sobrecarga de los revisores que aceptaron participar, el incremento de revisores nuevos reclutados de urgencia y la ausencia de un número apropiado de personas con experiencia en campos muy específicos de la Salud Pública. Parecería haber un límite en la cantidad de trabajo que se puede efectuar con el *pool* de revisores actual. La dificultad de encontrar revisores en una convocatoria que requiere más de 600 revisiones en un corto periodo de tiempo pone en evidencia este límite operativo. Convocatorias más acotadas, a más largo plazo y con mayores incentivos podrían ser potenciales soluciones a algunos de estos aspectos.

Se identifican como relevantes la ausencia de comentarios por parte de algunos revisores, el bajo detalle y estructura de la evaluación y la inconsistencia entre las revisiones. Algunos aspectos podrían resolverse con el desarrollo de guías más estructuradas, que incorporen reglas de conducta con la evaluación y una perspectiva centrada en las propiedades de los proyectos relevantes para la DIS. El diseño de una encuesta a revisores para evaluar sus críticas, opiniones, necesidades e ideas dirigidas a reformular el proceso, la adopción de criterios explícitos para su selección, el diseño de incentivos para incrementar la participación y el detalle y exhaustividad de la evaluación, así como la posibilidad de proveer entrenamiento continuo a los revisores serían iniciativas útiles para afrontar el problema.

Es probable que se necesite, además, algún cambio en los tipos, tiempos y magnitudes de las convocatorias: por ejemplo, ventanilla abierta, subvención a proyectos y no subsidios, propuestas iniciadas por el investigador y otras por el MSAL, o bien contar con paneles de evaluación consolidados, incentivados, constituidos *a priori* y no emergentes, mecanismos de evaluaciones *ex ante* y no *ex post*, y otras posibles intervenciones.

Esas políticas alinearían aún más los procesos de la DIS con los que efectúan otras agencias de financiación en el mundo que gerencian fondos públicos, en pos de garantizar investigaciones de calidad, vinculadas con políticas de salud priorizadas y con resultados aplicables a la misma comunidad que las requiere y subvenciona.

RELEVANCIA PARA POLÍTICAS E INTERVENCIONES SANITARIAS

La información provista por esta encuesta constituye uno de los insumos clave para decidir qué modificaciones o reformas se deben hacer al proceso de evaluación de los proyectos de investigación a subsidiar por el MSAL.

RELEVANCIA PARA LA FORMACIÓN DE RECURSOS HUMANOS EN SALUD

Se espera que la difusión de los resultados a los postulantes y revisores, y a la comunidad científica en general,

retroalimente el proceso de mejora del sistema de revisión por pares.

RELEVANCIA PARA LA INVESTIGACIÓN EN SALUD

Se impone la realización de investigaciones adicionales que relevan las opiniones e iniciativas del cuerpo de evaluadores, insumo crucial para informar el proceso de mejora del sistema de revisión por pares.

AGRADECIMIENTOS: a la Profesora en Ciencias Antropológicas Carla Carbonelli, y a la Socióloga Mariana Fernandez, ambas compañeras de la Dirección de Investigación en Salud del MSAL, por su inestimable colaboración en el diseño y gestión electrónica de la encuesta.

DECLARACIÓN DE CONFLICTO DE INTERESES: No hubo conflicto de intereses durante la realización del estudio.

CONTRIBUCIÓN DE LOS AUTORES: Todos los autores han efectuado una contribución sustancial a la concepción o diseño del artículo o a la adquisición, análisis o interpretación de los datos; han participado en el diseño del trabajo de investigación o en la revisión crítica de su contenido intelectual; han intervenido en la aprobación de la versión final que vaya a ser publicada y son capaces de responder respecto a todos los aspectos del artículo de cara a asegurar que las cuestiones relacionadas con la exactitud o integridad de cualquier parte del trabajo han sido adecuadamente investigadas y resueltas.

Cómo citar este artículo: Torales SB, Lago M, Ianovsky OG. Encuesta de opinión y satisfacción a los participantes de la Convocatoria Salud Investiga sobre el proceso de evaluación de las propuestas de investigación. *Rev Argent Salud Publica*. 2022;14:e73. Publicación electrónica 12 de Abr de 2022.

REFERENCIAS BIBLIOGRÁFICAS

¹ Publons. Grant Review in Focus. Global State of Peer Review Series [Internet]. [lugar desconocido]: Publons; 2019 [citado 6 Ene 2022]. Disponible en: <https://publons.com/community/gspr/grant-review>

² European Science Foundation. European Peer Review Guide. Integrating Policies and Practices into Coherent Procedures [Internet]. Estrasburgo: ESF; 2011 [citado 6 Ene 2022]. Disponible en: https://www.esf.org/fileadmin/user_upload/esf/European_Peer_Review_Guide_2011.pdf

³ Guthrie S, Ghiga I, Wooding S. What do we know about grant peer review in the health sciences? An updated review of the literature and six case studies [Internet]. Santa Mónica (CA): RAND Corporation; 2018 [citado 6 Ene 2022]. Disponible en: https://www.rand.org/pubs/research_reports/RR1822.html

⁴ NIH Grants & Funding. Continuous Review of Peer Review [Internet]. Bethesda (MD): NIH; 2017 [citado 6 Ene 2022]. Disponible en: <https://grants.nih.gov/policy/peer/continuous-review.htm>

⁵ National Institutes of Health, Office of Extramural Research. Enhancing Peer Review - Survey Results Report [Internet]. Bethesda (MD): NIH; 2017 [citado 6 Ene 2022]. Disponible en: <https://grants.nih.gov/grants/peer/files/Final-Phase-III-Enhancing-Peer-Review-Report.pdf>

⁶ Cameron J, Pierce WD. Reinforcement, Reward, and Intrinsic Motivation: A Meta-Analysis. *Rev Educ Res*. 1994;64(3):363-423. doi: 10.3102/00346543064003363.

⁷ Sattler DN, McKnight PE, Naney L, Mathis R (2015) Grant Peer Review: Improving Inter-Rater Reliability with Training. *PLoS One*. 2015;10(6):e0130450. doi: 10.1371/journal.pone.0130450.

⁸ Jayasinghe UW, Marsh HW, Bond N. A multilevel cross-classified modelling approach to peer review of grant proposals: The effects of assessor and researcher attributes on assessor ratings. *J R Stat Soc Ser A Stat Soc* [Internet]. 2003 [citado 6 Ene 2022];166(3):279-300. Disponible en: <https://publons.com/publon/4887394/>

⁹ Fogelholm M, Leppinen S, Auvinen A, Raitanen J, Nuutinen A, Vaananen K. Panel discussion does not improve reliability of peer review for medical research grant proposals. *J Clin Epidemiol*. 2012;65(1):47-52. doi: 10.1016/j.jclinepi.2011.05.001.

¹⁰ Lobb R, Petermann L, Manafo E, Keen D, Kerner J. Networking and knowledge exchange to promote the formation of transdisciplinary coalitions and levels of agreement among transdisciplinary peer reviewers. *J Public Health Manag Pract*. 2013;19(1):E9-E20. doi: 10.1097/PHH.0b013e31823991c2.



Esta obra está bajo una licencia de *Creative Commons* Atribución-NoComercial-CompartirIgual 4.0 Internacional. Reconocimiento – Permite copiar, distribuir y comunicar públicamente la obra. A cambio se debe reconocer y citar al autor original. No comercial – esta obra no puede ser utilizada con finalidades comerciales, a menos que se obtenga el permiso.