

Generalizabilidad o validez externa de los ensayos clínicos en psiquiatría*

Mayumi Okuda¹
Carlos Blanco²

Resumen

Introducción: Para aprobar el uso comercial de un medicamento en Estados Unidos, la Food and Drug Administration (FDA) exige, por lo menos, dos ensayos clínicos de fase III que demuestren su seguridad y eficacia. En estos estudios, sin embargo, los resultados pueden verse influenciados por la composición de sus muestras. *Objetivo:* Describir los problemas de generalizabilidad de los resultados de ensayos clínicos en psiquiatría, debido a la composición de las muestras. *Método:* Breve descripción del efecto del uso de criterios de elegibilidad muy restrictivos en la generalizabilidad de los resultados de los ensayos clínicos. También se exponen ejemplos de estudios que han examinado este tema tanto en muestras clínicas como en muestras comunitarias. *Conclusión:* Tanto en muestras clínicas como en muestras comunitarias, varios estudios han demostrado que, usualmente, una gran proporción de adultos con trastornos psiquiátricos son excluidos, debido a los criterios de elegibilidad comúnmente utilizados en los ensayos clínicos. Éstos suelen reclutar muestras altamente selectivas, en lugar de adultos que podrían ser más representativos de pacientes típicos. Los ensayos clínicos deben evaluar cuidadosamente los efectos de sus criterios de elegibilidad sobre la generalizabilidad de sus resultados, ya que esto puede limitar su representatividad.

Palabras clave: ensayos clínicos, selección de paciente, protocolos, psiquiatría.

Title: Generalizability or External Validity of the Clinical Trials in Psychiatry

Abstract

Introduction: To approve the prescription of a medication in the United States, the Food and Drug Agency (Food and Drug Administration (FDA)) requires at least two Phase III clinical trials demonstrate the safety and efficacy of the medication. In these studies, however, the results can be influenced by the composition of their samples. *Objective:* To describe issues with the generalizability of results from clinical trials in psychiatry due to the use of highly selected samples. *Methods:* We briefly describe the effects of stringent eligibility criteria on the results of clinical trials. We summarize studies that have examined this issue both in clinical

.....
* Este estudio fue financiado por las becas DA019606 y DA023200 del Instituto Nacional de Salud de los Estados Unidos de América.

¹ Médica. New York State Psychiatric Institute/Department of Psychiatry, College of Physicians and Surgeons of Columbia University. Nueva York, Estados Unidos.

² Médico. PhD. New York State Psychiatric Institute/Department of Psychiatry, College of Physicians and Surgeons of Columbia University. Nueva York, Estados Unidos.

and community samples. *Conclusion:* Both in clinical and community samples, studies have shown that a significant proportion of adults with psychiatric disorders are typically excluded by eligibility criteria commonly used in clinical trials. This tends to recruit highly selective samples, rather than adults who might be more representative of typical patients. Clinical trials should carefully evaluate the effects of eligibility criteria on the generalizability of their results. The use of stringent exclusionary criteria could limit the representativeness of results.

Key words: Clinical trials, patient selection, protocols, psychiatric.

Introducción

En las últimas décadas se ha incrementado el relieve puesto en la práctica de la medicina basada en la evidencia (MBE), definida como “el uso consciente, explícito y juicioso de la mejor y más actualizada evidencia para la toma de decisiones en el cuidado de los enfermos” (1). Generalmente, la mejor evidencia suele proceder de los ensayos clínicos aleatorizados.

Para aprobar el uso comercial de un medicamento en Estados Unidos, el Food and Drug Administration (FDA) exige, por lo menos, dos ensayos clínicos de fase III que demuestren su seguridad y eficacia. En estos estudios, sin embargo, los resultados pueden verse influenciados por la composición de sus muestras. En vista de que suele haber una gran restricción en la participación de individuos en este tipo de estudios, debido a los

criterios de inclusión y exclusión, estos estudios pueden no necesariamente ofrecer resultados precisos, en términos de su eficiencia en la práctica real, especialmente si las características que excluyen a los participantes están asociadas con peores desenlaces clínicos (2).

Por esto en años recientes ha surgido una gran preocupación por la generalizabilidad o validez externa de los resultados de tales estudios (3,4). Como resultado de esta inquietud, han aparecido diferentes análisis que intentan cuantificar la generalizabilidad o validez externa de los resultados de los ensayos clínicos.

Estudios sobre la generalizabilidad de los ensayos clínicos en muestras clínicas

Un estudio realizado con la intención de examinar la representatividad de las muestras de los ensayos clínicos comparó las características clínico-demográficas y la respuesta al tratamiento de los participantes reclutados por medio de anuncios contra las características de los pacientes referidos de ámbitos clínicos (5). Utilizando este método, este estudio encontró pocas diferencias clínicamente relevantes entre las dos poblaciones, lo que apoyó la generalizabilidad de los estudios de eficacia.

Sin embargo, al comparar las poblaciones que participan en estudios

clínicos con aquellas que no participan, otros estudios han encontrado resultados menos alentadores. Así es como estudios como el conducido por Stirman y colaboradores (6) encontraron que cerca de la mitad de los pacientes vistos en prácticas privadas de psicoterapia no habrían podido participar en un ensayo clínico, ya que habrían sido excluidos por alguno de los criterios de exclusión.

De modo similar, Zimmerman y colaboradores (7) encontraron que según una lista de los criterios de elegibilidad que suelen ser más comúnmente utilizados en la gran parte de ensayos clínicos de depresión mayor, tan sólo el 14% de los pacientes con el trastorno que atendían a la consulta externa de un hospital universitario habrían podido participar en un ensayo clínico típico de este trastorno. Entre los hallazgos del estudio se encontró, además, que criterios de exclusión, como la presencia de trastornos de ansiedad concomitantes, habrían excluido a un gran porcentaje de sujetos. Debido a que se ha demostrado que la presencia de este tipo de comorbilidad influye fuertemente en los desenlaces clínicos, este hallazgo pone en cuestionamiento la posibilidad de extrapolar los resultados de los ensayos clínicos tradicionales a los individuos en la comunidad que sufren la enfermedad.

Un estudio reciente publicado en *JAMA* (8) examinó el tipo y la ex-

tensión de los criterios de exclusión en los estudios clínicos aleatorizados publicados en las principales revistas médicas entre 1994 y el 2006. Encontró que la presencia de enfermedades médicas, la edad y el uso de medicamentos de uso común excluyen a una gran parte de participantes potenciales. Otros criterios, como el sexo y el nivel socioeconómico también excluyen a una proporción importante de individuos. Más aún, el estudio encontró que el uso de algunos de estos criterios de exclusión estaba pobremente justificado. De manera similar, otros investigadores, como Humphreys y colaboradores (9), han encontrado que algunos criterios de exclusión son excesivamente restrictivos y que ofrecen muy poco beneficio en términos de seguridad y validez interna, porque limitan gravemente la generalizabilidad de los resultados de los ensayos clínicos. Por ende, dado que no todos los estudios reportan los criterios de elegibilidad que utilizan o la manera como se definen, esto les dificulta a otros investigadores repetir dichos criterios.

Recientemente, un equipo de investigadores del proyecto *Sequenced Treatment Alternatives to Relieve Depression (STAR*D)*, el estudio clínico prospectivo más grande que se ha realizado hasta la fecha, examinó la generalizabilidad de sus resultados (2). Al subdividir a los sujetos de la muestra en aquellos que reunían los criterios de elegibilidad para

un estudio de eficacia típico (que utiliza criterios más restrictivos) y en aquellos que eran elegibles usando criterios más amplios, el estudio demostró que los desenlaces clínicos diferían de acuerdo con el subgrupo al que pertenecían. El estudio encontró que las tasas de remisión y respuesta al tratamiento eran mayores en el grupo que se asemejaba a un estudio de eficacia, donde los individuos tendían a ser más jóvenes, a tener mayor nivel educativo y mayores ingresos, a ser con mayor probabilidad de raza blanca, a no ser latinos, a estar empleados, a estar casados y a tener seguro médico privado.

Además de estos estudios, un meta-análisis de estudios publicados entre 1990 y 1998 sobre la eficacia de diversas psicoterapias (10) demostró que aproximadamente el 68%, el 64% y el 65% de los individuos con depresión mayor, trastorno de pánico y trastorno de ansiedad generalizada, respectivamente, eran excluidos de este tipo de estudios. El estudio señaló que un gran porcentaje de estas exclusiones se debía a la presencia de trastornos comórbidos, que son comunes en escenarios clínicos.

Estudios sobre la generalizabilidad de los ensayos clínicos a la comunidad

Hasta hace poco no habría sido posible examinar la representatividad

de los resultados de estos estudios, en cuanto a la población general que padece los trastornos, por haber utilizado muestras de conveniencia para examinar la generalizabilidad de los resultados de los ensayos clínicos. Para superar las limitaciones de estos estudios previos, otra serie de estudios conducidos por miembros de nuestro equipo (11-13) examinó la proporción de individuos de la población general con el trastorno que pudieran haber sido incluidos en un estudio clínico típico del trastorno en cuestión. Esta serie de estudios examinó la proporción de individuos con depresión mayor, dependencia al alcohol y dependencia al cáñamo (*cannabis*).

Desde el punto de vista de políticas de salud, las muestras locales, de conveniencia y los individuos que buscan tratamiento no suelen representar adecuadamente a los individuos dentro de la población general que necesitan tratamiento, es decir, a la población de individuos con el trastorno en la comunidad, incluidos los que no buscan tratamiento. En cuanto mayor sea la proporción de individuos que necesitan tratamiento que califiquen para ser incluidos en un estudio clínico, mayor será la posibilidad de que sus resultados sean generalizables a la población afectada.

Para realizar estos estudios, utilizamos la Encuesta Epidemiológica Nacional sobre Alcohol y Condicio-

nes Relacionadas (NESARC, por su sigla en inglés), del 2001-2002, conducida por el Instituto Nacional sobre el Abuso de Alcohol y Alcoholicismo (NIAAA, por su sigla en inglés) (14,15). La encuesta, que es la de tipo epidemiológico más extensa, cuenta con una muestra representativa de la población civil no institucionalizada de Estados Unidos, que incluye a 43.093 individuos de más de 18 años que residen en casas habitacionales y cuarteles.

Esta encuesta incluye un espectro amplio de diagnósticos psiquiátricos de los ejes I y II, entre estos: trastornos por uso de sustancias (abuso y dependencia al alcohol, abuso y dependencia a las drogas y dependencia a la nicotina); trastornos del afecto (depresión mayor, distimia, trastorno afectivo bipolar); trastornos de ansiedad (trastorno de pánico, ansiedad social, fobia específica, trastorno de ansiedad generalizada), y trastornos de personalidad evitativo, dependiente, obsesivo-compulsivo, paranoide, histriónico y antisocial.

El cuestionario también incluye variables que miden el uso de sustancias, eventos estresantes y búsqueda de tratamiento para los problemas de salud mental. La tasa de respuesta fue del 81%. Los datos fueron ajustados para representar la población civil basada en el censo nacional del 2000. Los métodos de campo incluyeron entrenamientos

estructurados, supervisión y control de calidad, entre éstos llamadas aleatorias a los entrevistados para verificar los datos.

Para el desarrollo de la encuesta el instrumento utilizado fue el Alcohol Use Disorder and Associated Disabilities Interview Schedule-IV (AUDADIS-IV) del NIAAA (16). El AUDADIS es una entrevista diagnóstica estructurada aplicada cara a cara por entrevistadores legos. Esta proporciona diagnósticos de alguna vez en la vida y durante los últimos 12 meses según los criterios diagnósticos del *Manual diagnóstico y estadístico de los trastornos mentales* (DSM-IV). La validez y confiabilidad del instrumento han sido evaluados en varios estudios incluyendo el World Health Organization/National Institutes of Health International Study on Reliability and Validity y han sido previamente descritas en otras publicaciones (17).

Con el fin de examinar el grado con el que los grupos de individuos con trastornos psiquiátricos son representados en ensayos clínicos y las limitaciones en la generalizabilidad de sus resultados, aplicamos una serie de criterios de elegibilidad comúnmente utilizados a los individuos con un diagnóstico en los últimos 12 meses y en la submuestra de individuos que buscaron tratamiento para dichos trastornos durante ese período. Según el

trastorno en cuestión, usamos los criterios de elegibilidad que fueran considerados los más comunes en la mayoría de ensayos clínicos de dicho trastorno. Adaptamos los criterios según las preguntas disponibles en el cuestionario.

En el primer estudio se examinaron individuos con un episodio depresivo mayor en los últimos 12 meses (11) y empleamos los criterios que según Zimmerman y colaboradores (7) representan los más usados en los ensayos clínicos de depresión mayor. Éstos incluyen tener una historia positiva de un episodio de manía o hipomanía, tener un cuadro con rasgos psicóticos, riesgo suicida significativo, un trastorno por el uso de alcohol o drogas en el último año, un índice de gravedad de menos de 18 puntos en la Escala de Hamilton para Depresión (HAM-D), distimia concomitante, otros trastorno psiquiátricos comórbidos, tener un episodio de menos de cuatro semanas o más de dos años de duración y la presencia del trastorno límite de la personalidad.

En este estudio encontramos que entre los individuos con diagnóstico de depresión mayor, el 75,8% habrían sido excluidos por al menos un criterio de exclusión. Entre aquellos que habían buscado tratamiento, el 66,9% habrían sido excluidos por al menos un criterio. Aunque algunos criterios no excluían individuos, otros habrían excluido hasta a un

47,4% de los individuos. El estudio encontró que tener un trastorno psiquiátrico adicional y la duración del episodio depresivo eran los criterios que excluirían al mayor porcentaje de individuos.

En el segundo estudio se examinaron los individuos con dependencia al alcohol y utilizamos los criterios descritos como los más comúnmente reportados en los estudios de dependencia al alcohol, según Monahan y Finney (18,19). Éstos incluían la presencia de trastornos psiquiátricos comórbidos, haber recibido tratamiento recientemente, tener una condición médica, embarazo, tener una motivación baja o falta de adherencia, vivir en un lugar alejado al sitio de tratamiento, aislamiento social, dificultades económicas, tener un nivel educativo bajo y tener problemas legales. En este estudio encontramos que cerca de la mitad (50,5%) de los individuos con dependencia al alcohol y el 79,4% de aquellos que buscaron tratamiento fueron excluidos por al menos un criterio de exclusión (12). La motivación baja o la falta de adherencia, tener dificultades económicas y tener enfermedades médicas concomitantes fueron los criterios que excluyeron al mayor porcentaje de individuos.

Los individuos con dependencia al alcohol representan un grupo muy heterogéneo y un porcentaje considerable presenta comorbi-

alidad médica y psiquiátrica (20). Sin embargo, los pacientes con comorbilidad son precisamente los más frecuentemente excluidos de los ensayos que prueban intervenciones nuevas. Un estudio reciente, conducido por Humphreys y colaboradores (21), demostró que los criterios de elegibilidad comúnmente utilizados (comorbilidad psiquiátrica y médica, inestabilidad social, baja motivación/adherencia, problemas con drogas) en los ensayos de dependencia al alcohol producían sesgos entre moderados y graves en las estimaciones de los desenlaces y que, contrario a lo esperado, el uso de criterios restrictivos no incrementa el poder estadístico.

Salvo algunas excepciones, la mayoría de los criterios de los ensayos de dependencia al cáñamo son similares a los de los estudios de dependencia al alcohol. En un tercer estudio, examinamos a los individuos con dependencia al cáñamo y encontramos que la gran mayoría (aproximadamente el 80%) de los adultos con dependencia en la muestra habrían sido excluidos por al menos uno de los criterios (13): tener problemas legales, el uso de otras sustancias ilícitas y fumar menos de cinco cigarrillos a la semana. Es de notar, en cuanto a la presencia de problemas legales, que en la gran mayoría de los países, buena parte de los individuos que reciben tratamiento para problemas

con cáñamo son referidos por el sistema criminal (22).

Esta serie de estudios aportan a la noción de que los ensayos clínicos típicos excluyen a la mayoría de los adultos con el trastorno tanto en la comunidad, así como a aquellos que reciben tratamiento.

Comentario

Los ensayos clínicos suelen reclutar a pacientes “puros”, en vez de pacientes “típicos”, que no representan el “mundo real” de la práctica médica (23). Aunque el uso de criterios de elegibilidad restrictivos es apropiado en estudios iniciales de eficacia, se debe crear un balance entre la aplicación de estos criterios y el impacto que éstos pueden tener en la generalizabilidad de los resultados (24). Algunos investigadores, como DelBoca y Darkes (25), han propuesto estrategias efectivas para aumentar la representatividad de poblaciones que son típicamente excluidas, como ofrecer servicios de guardería gratuita o de bajo costo y subsidio de transporte.

El uso de criterios de inclusión más amplios podría aumentar la generalizabilidad de los resultados de los ensayos clínicos a la práctica y potencialmente reducir la respuesta al placebo y las tasas de remisión. Sin embargo, es probable que esto aumente el riesgo de efectos adversos. Los ensayos clínicos deben

justificar adecuadamente el uso de sus criterios de exclusión y calcular el impacto de sus criterios de elegibilidad en la generalizabilidad de sus resultados.

Conclusión

Los ensayos clínicos deben aumentar y mejorar la operacionalización de sus criterios de elegibilidad. Es importante que los investigadores que conducen estos ensayos reporten cómo se operan sus criterios y cómo éstos podrían influenciar la selección de sus participantes. Los sesgos en el reclutamiento pueden limitar la generalizabilidad de los resultados de los ensayos clínicos y reducir el conocimiento que se obtiene de ellos. Un mayor acento en criterios de selección más incluyentes moverá el campo de la investigación hacia el desarrollo de intervenciones dirigidas a la población psiquiátrica necesitada de atención.

Referencias

1. Evidenced-Based Medicine Working Group. Evidence-based medicine: a new approach to teaching the practice of medicine. *JAMA*. 1992;268(17):2420-3.
2. Wisniewski SR, Rush AJ, Nierenberg AA, Gaynes BN, Warden D, Luther JF, et al. Can phase III trial results of antidepressant medications be generalized to clinical practice? A STAR*D report. *Am J Psychiatry*. 2009;166(5):599-607.
3. National Institute on Drug Abuse. American Recovery and Reinvestment Act of 2009 (Recovery Act): 07-DA-102 Development of methodologies and scientific tools for improving and/or assessing the external validity of randomized clinical trial (RCT) results to known populations [Internet]. Bethesda: National Institute on Drug Abuse; 2009 [citado: 24 mayo 2010]. Disponible en: <http://www.drugabuse.gov/Recovery/NIDAChallenge.html>
4. Rothwell PM. External validity of randomized controlled trials: "to whom do the results of this trial apply?" *Lancet*. 2005;365(9453):82-93.
5. Rapaport MH, Frevert T, Babior S, Zisook S, Judd LL. A comparison of demographic variables, symptom profiles, and measurements of functioning in symptomatic volunteers and an outpatient clinical population. *Psychopharmacol Bull*. 1995;31(1):111-4.
6. Stirman SW, DeRubeis RJ, Crits-Christoph P, Brody PE. Are samples in randomized controlled trials of psychotherapy representative of community outpatients? A new methodology and initial findings. *J Consult Clin Psychol*. 2003;71(6):963-72.
7. Zimmerman M, Mattia JI, Posternak MA. Are subjects in pharmacological treatment trials of depression representative of patients in routine clinical practice? *Am J Psychiatry*. 2002;159(3):469-73.
8. Spall HGV, Toren A, Kiss A, Fowler RA. Eligibility criteria of randomized controlled trials published in high-impact general medical journals: A systematic sampling review. *JAMA*. 2007;297(11):1233-40.
9. Humphreys K, Weingardt KR, Harris AHS. Influence of subject eligibility criteria on compliance with National Institutes of Health guidelines for inclusion of women, minorities, and children

- in treatment research. *Alcohol Clin Exp Res.* 2007;31(6):988-95.
10. Westen D, Morrison K. A multidimensional meta-analysis of treatments for depression, panic, and generalized anxiety disorder: an empirical examination of the status of empirically supported therapies. *J Consult Clin Psychol.* 2001;69(6):875-99.
 11. Blanco C, Olfson M, Goodwin RD, Ogburn E, Liebowitz MR, Nunes EV, et al. Generalizability of clinical trial results for major depression to community samples: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *J Clin Psychiatry.* 2008;69(8):1276-80.
 12. Blanco C, Olfson M, Okuda M, Nunes EV, Liu SM, Hasin DS. Generalizability of clinical trials for alcohol dependence to community samples. *Drug Alcohol Depend.* 2008;98(1-2):123-8.
 13. Okuda M, Hasin DS, Olfson M, Khan SS, Nunes EV, Montoya I, et al. Generalizability of clinical trials for cannabis dependence to community samples. *Drug Alcohol Depend.* 2010;111(1-2):177-81.
 14. Grant B, Moore TC, Shepard J, Kaplan K. Source and accuracy statement: wave 1 National Epidemiologic Survey on Alcohol and Related Conditions (NESARC). 2003. Disponible en www.niaaa.nih.gov.
 15. Grant BF, Dawson DA, Stinson FS, Chou PS, Kay W, Pickering R. The alcohol use disorder and associated disabilities interview schedule-IV (AUDADIS-IV): reliability of alcohol consumption, tobacco use, family history of depression and psychiatric diagnostic modules in a general population sample. *Drug Alcohol Depend* 2003;71(1):7-16.
 16. Hasin D, Carpenter KM, McCloud S, Smith M, Grant BF. The alcohol use disorder and associated disabilities interview schedule (AUDADIS): reliability of alcohol and drug modules in a clinical sample. *Drug Alcohol Depend.* 1997;44(2-3):133-41.
 17. Cottler LB, Grant BF, Blaine J, Mavreas V, Pull C, Hasin D, et al. Concordance of DSM-IV alcohol and drug use disorder criteria and diagnoses as measured by AUDADIS-ADR, CIDI and SCAN. *Drug Alcohol Depend.* 1997;47(3):195-205.
 18. Monahan SC, Finney JW. Explaining abstinence rates following treatment for alcohol abuse: a quantitative synthesis of patient, research design and treatment effects. *Addiction.* 1996; 91(6):787-805.
 19. Humphreys K, Weingardt KR, Horst D, Joshi AA, Finney JW. Prevalence and predictors of research participant eligibility criteria in alcohol treatment outcome studies, 1970-98. *Addiction.* 2005;100(9):1249-57.
 20. Hasin DS, Stinson FS, Ogburn E, Grant BF. Prevalence, correlates, disability, and comorbidity of DSM-IV alcohol abuse and dependence in the United States: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Arch Gen Psychiatry.* 2007;64(7):830-42.
 21. Humphreys K, Harris AH, Weingardt KR. Subject eligibility criteria can substantially influence the results of alcohol-treatment outcome research. *J Stud Alcohol Drugs* 2008;69(5):757-64.
 22. United Nations Office on Drugs and Crime. 2006. Bulletin on narcotics: review of the world cannabis situation. Disponible en: www.unodc.org/unodc/en/bulletin_on_narcotics.html.
 23. Marlatt GA. From hindsight to foresight: a commentary on project MATCH. En: *Changing addictive behavior: bridging clinical and public health strategies.* New York: Guilford; 1999. p. 45-66.

24. Humphreys K, Weisner C. Use of exclusion criteria in selecting research subjects and its effect on the generalizability of alcohol treatment outcome studies. *Am J Psychiatry*. 2000; 157(4):588-94.
25. DelBoca FK, Darkes J. Enhancing the validity and utility of randomized clinical trials in addictions treatment research: II. Participant samples and assessment. *Addiction*. 2007;102(8):1194-203.

Conflicto de interés: los autores manifiestan que no tienen ningún conflicto de interés en este artículo.

Recibido para evaluación: 14 de abril del 2010
Aceptado para publicación: 21 de junio del 2010

Correspondencia
Carlos Blanco
Department of Psychiatry
New York State Psychiatric Institute
College of Physicians and Surgeons
Columbia University
1051 Riverside Drive, New York, NY 10032
New York, Estados Unidos
cb255@columbia.edu