REVISTA
CES MEDICINA

Revisión de tema

El índice de fragilidad y sus características en ensayos clínicos aleatorizados de diabetes mellitus

The Fragility Index and its characteristics in randomized clinical trials of diabetes mellitus

David Benavides-Zora [1] ✉ ORCID, Sara Vásquez-Martínez [2] ✉ ORCID, Jorge Hernando Donado [3] ✉ ORCID

[1] Médico general, Universidad de la Sabana, Anestesiólogo, Universidad CES, Colombia.
[2] Médica general, Universidad CES, Colombia.
[3] Epidemiólogo Clínico, Hospital Pablo Tobón Uribe, profesor titular Universidad Pontificia Bolivariana, Colombia.

Fecha correspondencia:
Recibido: marzo 01 de 2022.
Revisado: mayo 16 de 2022.
Aceptado: mayo 19 de 2022.

Forma de citar:
Benavides-Zora D, Vásquez-Martínez S, Donado JH. El índice de fragilidad y sus características en ensayos clínicos aleatorizados de diabetes mellitus. Rev CES Med. 2022; 36(2): 106-121. https://dx.doi.org/10.21615/cesmedicina.6632

Open access

Ética de publicaciones
Revisión por pares
Gestión por Open Journal System
DOI: 10.21615/cesmedicina.6632
ISSNe: 2215-9177
ISSN: 0120-8705

Publica con nosotros

Resumen

Introducción: para valorar la robustez de los resultados se ha propuesto una herramienta llamada el Índice de Fragilidad (IF), esta se define como el mínimo número de pacientes que se tienen que cambiar de "No eventos" a "Eventos" en el grupo de intervención para que un resultado estadísticamente significativo pase a no significativo, evidenciando que entre menor sea el IF, los resultados serán más frágiles. Diferentes autores han encontrado que la significancia de los resultados de muchos Ensayos Clínicos Controlados (ECA) dependen de pocos eventos. El objetivo del estudio fue evaluar el IF de los ECA en diabetes mellitus de cinco de las revistas médicas de mayor impacto a nivel mundial. **Metodología:** se realizó búsqueda electrónica en PubMed, para identificar ECA en Annals of Internal Medicine, BMJ, The Lancet, The New England Journal of Medicine y JAMA. Se revisaron los ECA en pacientes con diabetes mellitus o prediabetes y se calculó el IF para cada desenlace según el método descrito por Walsh et al, usando tablas de contingencia 2x2. Se planeó usar el coeficiente de correlación de Spearman para evaluar la correlación entre el IF y el tamaño de la muestra, el número de eventos, el valor de p y el tiempo de seguimiento. Se evaluó la significancia de todos los resultados con un valor de p<0,05. **Resultados:** la mediana del IF fue 11, y en tres estudios (7,3%) se encontró que el resultado no era estadísticamente significativo después de recalcular la p con el

test exacto de Fisher. Se encontró relación directa leve entre el número de eventos y el IF (Rho= 0,343, p= 0,02) y correlación moderada inversa entre el valor de p y el IF (Rho= -0,632, p= 0,000). No se encontró correlación estadísticamente significativa entre el tamaño de muestra, tiempo de seguimiento y pérdidas con el IF. **Conclusiones:** en los ECA sobre diabetes, los resultados estadísticamente significativos dependen de pocos eventos, evidenciado por un bajo valor en el IF, los valores de esta medición están relacionados de forma directa con el número de eventos, e inversa con el valor de p.

Palabras clave: índice de fragilidad; diabetes mellitus; ensayos clínicos controlados.

Abstract

Introduction: to evaluate the robustness of the results, a tool called the Fragility Index (FI) has been proposed, which is defined as the minimum number of patients that have to be changed from "No events" to "Events" in the intervention group to change a statistically significant to nonsignificant result. Showing that among a lower Fragility Index, the results of the trial will be less robust or more fragile. Different authors have found that the significance of the results of many controlled clinical trials (RCTs) depend on very few events. The objective of the study is to evaluate the FI of controlled clinical trials in diabetes mellitus in five of the general medical journals with the greatest impact factor worldwide. **Methods:** an electronic search was conducted in PubMed to identify randomized clinical trials in The Annals of Internal Medicine, BMJ, The Lancet, The New England Journal of Medicine and JAMA. Clinical trials were reviewed with diabetic or prediabetic patients and the FI was then calculated for each outcome according to the method described by Walsh et al, using 2x2 contingency tables. A priori was planned to use the Spearman correlation coefficient to evaluate the direct correlation between the Fragility Index and sample size, number of events, p-value and follow-up time. The significance of all the results was evaluated with a value of $p < 0.05$. **Results:** the median Fragility Index was 11, and in three studies (7.3%) the result were not statistically significant after recalculating the p value with Fisher's exact test. A slight direct relationship between the number of events and the Fragility Index (Rho = 0.343, p = 0.02) was found and a moderate inverse correlation was observed between the p value and the FI (Rho = -0.632, p = 0.000). No statistically significant correlation was found between sample size, follow-up time and losses with the FI. **Conclusions:** in controlled clinical trials on diabetes, we found that the statistically significant results depend on a few events, evidenced by a low value in the Fragility Index. The values of this measurement are related to the number of events and negatively to the p value.

Keywords: fragility index; diabetes mellitus; controlled clinical trials.

Introducción

Para que un resultado de un ensayo clínico sea válido éste no debe tener errores sistemáticos ni aleatorios. Para esto, desde hace décadas se viene tomando el valor de $p < 0,05$ como punto

de corte para hablar de significancia estadística, lo cual implica la probabilidad de obtener los resultados bajo la hipótesis nula [1,2]. Se ha venido cuestionando este valor cuando hay pocas diferencias en el valor de p (ejemplo, P= 0,051 y 0= 0,049) [3], e incluso han propuesto valores menores para obtener una evidencia más fuerte [4], además de sus múltiples limitaciones ya que este valor puede ser influenciado por un pequeño cambio en el número de eventos [1, 3, 5–8], es por esto que no debe ser la única medida para hablar de la significancia estadística de un resultado [9].

Se ha visto que inicialmente muchos Ensayos Clínicos Aleatorizado (ECA) muestran efectos estadísticamente significativos pero posteriormente se demuestra que son inefectivos, es más, existen reportes que hasta el 16% de los resultados tendrían efectos mucho menores a los reportados [10]. Es por eso, que es natural preguntarse: ¿Qué pasaría si los números observados fueran levemente modificados?.

Existen muchos factores en un ECA que pueden cambiar desenlaces estadísticamente significativos a no significativos, tales como errores de datos, pérdida en el seguimiento, retiro temprano, sesgos en la evaluación, entre otros, y que a pesar de mínimas variaciones en estos datos, el resultado seguirá siendo estadísticamente significativo y con similar magnitud del efecto, este concepto se conoce como robustez de los resultados [11].

Para valorar la robustez de los resultados, se ha propuesto una herramienta llamada el Índice de Fragilidad (IF), la cual se define como el mínimo número de pacientes que se tienen que cambiar de “No eventos” a “Eventos” en el grupo de intervención para cambiar un resultado estadísticamente significativo a no significativo, evidenciado en que el valor de p sea igual o mayor a 0,05 en el test exacto de Fisher, mostrando que entre un menor Índice de Fragilidad, los resultados de este ensayo serán menos robustos o más frágiles [1, 3, 9, 12–14].

La idea de reportar el IF no es reciente. Feinstein propone en 1990 la unidad del IF [15] y posteriormente Walter perfecciona el concepto en 1991 [6]. Se ha visto un renovado y creciente interés sobre este concepto en los últimos años debido a su fácil realización e interpretación.
Diferentes autores han encontrado que la significancia de los resultados de muchos ECAs dependen de muy pocos eventos y esto ha sido estudiado en distintas áreas como cardiología [16], cuidado intensivo [3], ortopedia [7, 11], otorrinolaringología [17] y urología [18]. Y recientemente oncología, trauma, cirugía de cadera y uso de esteroides en COVID [9, 12, 19, 20].

Nuestro objetivo es evaluar el IF de los ensayos clínicos controlados en diabetes mellitus en cinco de las revistas médicas generales con mayor factor de impacto a nivel mundial e identificar características de los ensayos clínicos que pueden estar relacionado con los valores del Índice de Fragilidad.

Metodología

Identificación de los estudios

Se realizó una búsqueda electrónica en PubMed, para identificar ensayos clínicos aleatorizados en cinco revistas de alto factor de impacto a nivel mundial: Annals of Internal Medicine, British Medical Journal, The Lancet, The New England Journal of Medicine (NEJM) y Journal of the American Medical Association (JAMA) con los siguientes términos de búsqueda.

- "diabetes mellitus"[MeSH Terms] AND ("The New England journal of medicine"[Journal] OR "Lancet"[Journal] OR "JAMA"[Journal] OR "BMJ"[Journal] OR "Annals of internal medicine"[Journal]) AND "Randomized Controlled Trial"[Publication Type] AND ("2006/01/01"[PDAT] : "2016/12/31"[PDAT] AND "humans"[MeSH Terms]).

Se revisaron ensayos clínicos involucrando humanos dentro del periodo de tiempo en los últimos diez años, correspondiendo entre el 01 de enero de 2006 hasta el 31 de diciembre de 2016. Dos revisores independientes evaluaron todos los resúmenes identificados, incluyendo todos los artículos con los siguientes criterios de inclusión y exclusión:

Criterios de inclusión:

1. Pacientes con Diabetes mellitus o prediabetes.
2. Desenlace como variable dicotómica.
3. Desenlace primario debe ser estadísticamente significativo con valor de P <0,05, e intervalo de confianza que excluya al valor nulo.
4. Propósito del ensayo clínico de superioridad.
5. Grupos paralelos con aleatorización 1:1 a tratamiento y control.

Criterios de exclusión:

1. Ensayos clínicos cluster o crossover.
2. Ensayos clínicos de no-inferioridad o equivalencia.

Datos

Los datos fueron extraídos para cada ECA por dos revisores por una tabla estandarizada en Excel 2016. Se compararon y se intentó llegar a un acuerdo por consenso, de no lograrse, se revisaron con un tercer autor.

Los datos extraídos de cada estudio fueron: el nombre de la revista, fuente de financiación, el valor de p, el tamaño de la muestra, el número de eventos, si el grupo control es placebo o fármaco activo, el tiempo de seguimiento y tipo de cegamiento de la intervención.

Índice de Fragilidad

El Índice de Fragilidad para cada desenlace fue calculado según el método descrito por Walsh et al (1), usando tablas de contingencia 2x2. Luego se añade un valor en el grupo expuesto o de intervención de "No eventos" a la celda de "Evento" (y substrayendo un "no evento" del mismo grupo para mantener el número total de pacientes constante) si el desenlace es negativo y se recalcula el valor de p del test exacto de Fisher (2 colas), se añaden sucesivamente eventos hasta que el valor de p sea mayor o igual a 0,05. Si el desenlace es positivo se resta un valor en el grupo expuesto de "Eventos" y se suma a la celda de "No evento" volviendo a recalcular el valor de p hasta que este deje de ser significativa (Figura 1). Para la evaluación de los valores del test exacto de Fisher y el IF se realizó un algoritmo en R Studio que permitía visualizar la totalidad de las iteraciones de los valores de p hasta encontrar el IF (Figura 2). De existir discordancia se volvía a correr el algoritmo y se buscaba de forma conjunta el valor en el cual a+f y b-f tenía un valor de p≥0,05.

El número mínimo de eventos totales que se tuvieron que añadir para lograr un valor de p mayor o igual a 0,05 es el Índice de Fragilidad.

Resultado del Ensayo		
	Evento	No Evento
Tratamiento	a	b
Control	c	d
Test exacto de Fisher p<0,05		

Cálculo del IF		
	Evento	No Evento
Tratamiento	a ± f	b ∓ f
Control	c	d
Test p de Fisher ≥ 0,05		

Desenlace negativo [b]		
	Evento	No Evento
Tratamiento	a + f	b - f
Control	c	d
Test p de Fisher ≥ 0,05		

Desenlace positivo [a]		
	Evento	No Evento
Tratamiento	a - f	b + f
Control	c	d
Test p de Fisher ≥ 0,05		

Figura 1. Cálculo del IF.

[a] Ejemplo: supervivencia, curación, remisión, mejoría.
[b] Ejemplo: mortalidad, morbilidad, complicaciones, recaída.

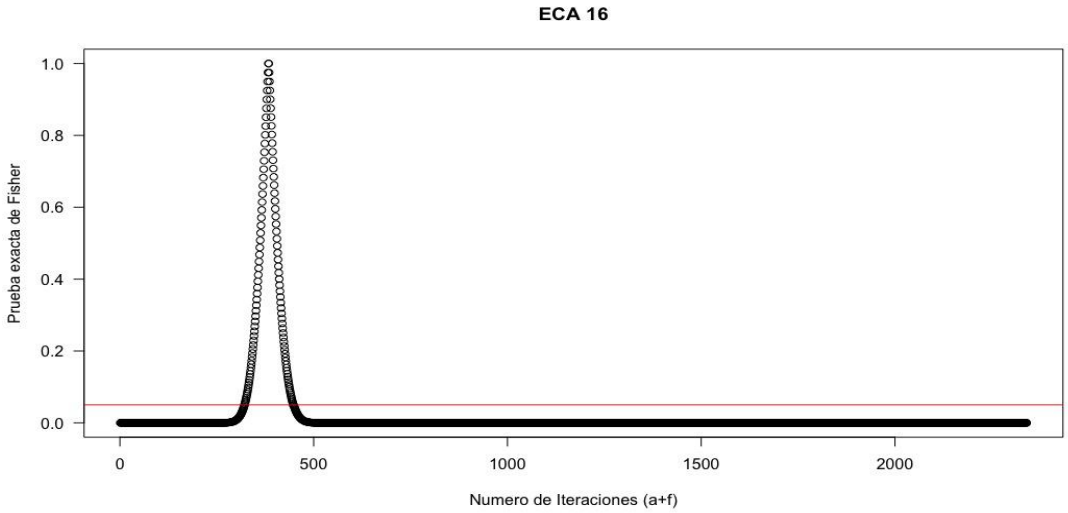

Figura 2.

La Figura 2 representa en la abscisa (eje x) el número de iteraciones (a+f) y en la ordenada (eje y) cada círculo representa un valor de p según la prueba exacta de Fisher. La línea roja corresponde a p=0,05, el primer corte sobre la gráfica, cuando sobrepasa el valor de p=0,05, es el número de iteraciones en que deja de tener significancia estadística o índice de fragilidad, el segundo corte muestra el número de iteraciones al cual vuelve a haber significancia con el desenlace contrario.

Análisis estadístico

Se emplearon estadísticos descriptivos para resumir las características de los estudios y el IF y CF de los ECA. Se reportaron las variables continuas como medias y medianas con desviaciones estándar y rangos intercuartílicos (IQRs) respectivamente. Se calculó el valor de la prueba de Kolmogorov-Smirnov para medir la distribución de las variables. Se planeó a priori usar el coeficiente de correlación de Spearman para evaluar la correlación directa entre el Índice de Fragilidad y el tamaño de la muestra, el número de eventos, el valor de p y el tiempo de seguimiento. Se evaluó la significancia de todos los resultados con un valor de $p<0,05$ con dos colas considerándolo significativo. Todos los análisis se realizaron usando Excel 2016 y SPSS V.26.

Se realizó coeficiente Kappa de Cohen para evaluar la concordancia entre los dos revisores para la búsqueda y calcular el IF. Un Kappa de 0 a 0,2 representa acuerdo leve, 0,21 a 0,4 acuerdo justo, 0,41 a 0,60 acuerdo moderado, y 0,61 a 0,80 acuerdo sustancial. Un valor por encima de 0,8 es considerado un acuerdo casi perfecto [21].

Resultados

Selección de estudios

Inicialmente la búsqueda de datos arrojó 250 artículos, luego cada autor aplicó criterios de inclusión y exclusión a los resúmenes obteniendo 48. Los autores revisan el texto completo encontrando que cinco de ellos eran sub-estudios sobre desenlaces secundarios y en dos ensayos clínicos no mencionan significancia estadística o el número de pacientes con el desenlace entre los grupos. Se obtiene una inclusión final de 41 desenlaces a evaluar (Figura 3). Se encuentra un Índice de Kappa de la búsqueda de 0,78 (IC 0,67; 0,88) y del cálculo del IF 0,91 (IC 0,83; 0,98).

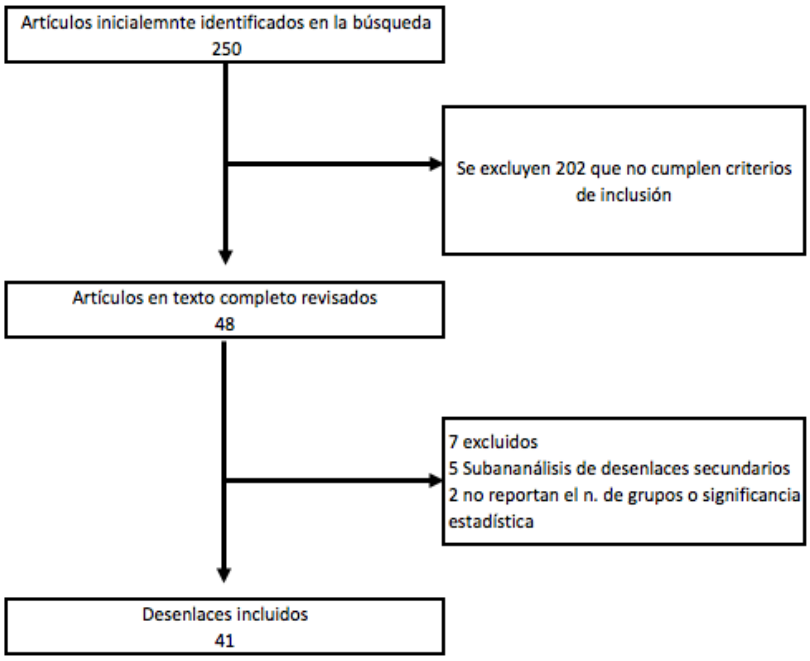

Figura 3. Búsqueda.

La Tabla 1 resume las características de los estudios. La mediana del tamaño de la muestra fue de 278 (IQR 101-2456,5), la mayoría de los artículos encontrados fue del NEJM (51,2%), no se encontró ningún ECA que cumplieran con los criterios de inclusión en BMJ. La mediana del seguimiento fue de 3 años (IQR 1,26-4,6), Encontrándose muy pocas pérdidas de pacientes analizados con una mediana de 0 (IQR 0-5,5). El 73% de los ensayos fueron financiados por la industria, el 78% tuvo tratamiento activo en el grupo control y el 61% de los estudios no tuvieron enmascaramiento.

Tabla 1. Características de los estudios.

Característica	Número (n=41)
Revista, n (%)	
Annals of Internal Medicine	2 (4,9)
JAMA	6 (14,6)
The Lancet	12 (29,3)
NEJM	21 (51,2)
BMJ	0 (0)
Año de publicación, n (%)	
2006 - 2008	10 (24,4)
2009 - 2011	7 (17,1)
2012 - 2014	18 (43,9)
2015 - 2016	6 (14,6)
Financiación, n (%)	
Si	30 (73,2)
No	11 (26,8)
Enmascaramiento, n (%)	
No ciego	25 (61)
Ciego	2 (4,9)
Doble ciego	14 (34,1)
Control	
Placebo	*9 (22)*
Activo	*32 (78)*
Tamaño de muestra, mediana (IQR) [min-max]	278 (101-2456,5) [40-11140]
Pérdidas, mediana (IQR) [min-max]	0 (0-5,5) [0-26]
Seguimiento (años), mediana (IQR) [min-max]	3 (1,26-4,6) [0,005-27]
Índice de Fragilidad, mediana (IQR) [min-max]	11 (6-19) [0-323]
Valor de p exacta de Fisher, mediana (IQR) [min-max]	0,002 (0,000047-0,012129) [0-0,09]

Índice de fragilidad

La mediana del Índice de fragilidad fue de 11, y en tres estudios (7,3%) se encontró que el resultado no era estadísticamente significativo después de recalcular la p con el test exacto de Fisher, resultando en un Índice de Fragilidad de cero. No hubo diferencia en los valores del IF entre revistas; el NEJM tiene una mediana de 9 al igual que Lancet, mientras que Annals tiene una mediana del IF de 12 y JAMA de 13. No hubo diferencia del IF entre estudios financiados y no financiados. Se encontró aumento de los valores del IF entre mayor número de desenlaces se reportará, con una mediana del IF de 8 (IQR 2-11) cuando había menos de 26 eventos, y una mediana de 32 (IQR 3-84) si había un número de eventos mayor a 425. También se observa diferencias en el valor del Índice de Fragilidad entre estudios con valores de p cercanos a 0,05

y valores de p más bajos, con una mediana del IF de 3 en los estudios con p más cercano a 0,05 y de 14 en los valores más bajos de p. Se resumió las características del IF entre subgrupos en la Tabla 2.

Tabla 2. IF por subgrupos.

Características de los Estudios	**Mediana del IF (IQR)**
Revista	
Annals	*12 (11-12)*
Lancet	*9 (3-9)*
NEJM	*9 (5-16)*
JAMA	*13 (8-38)*
Financiación	
No	*11 (3-15)*
Si	*11 (6-26)*
Enmascaramiento	
No ciego	*8 (5-15)*
Ciego	*168 (12-323)*
Doble ciego	*12 (1-26)*
Control	
Placebo	*13 (1-19)*
Activo	*10 (6-16)*
Muestra, n	
0-100	*8 (4-11)*
101-278	*12 (7-15)*
279-2.456	*9 (6-24)*
2.546-11.140	*28 (3-84)*
Número total de eventos	
0-26	*8 (2-11)*
27-95	*9 (6-14)*
96-424	*14 (9-24)*
425-3524	*32 (3-84)*
Seguimiento, años	
0-1,25	*7 (4-14)*
1,26-3	*15 (11-38)*
3,1-4,6	*14 (9-29)*
4,7-27	*7 (3-11)*
Valor de p exacta de Fisher	
0-0,000046	*14 (11-19)*
0,000047-0,002186	*16 (8-38)*
0,002186-0,012129	*9 (7-15)*
0,01213-0,0971	*3 (0-5)*

En la prueba Rho de Spearman se encontró una relación directa leve entre el número de eventos y el Índice de Fragilidad (Rho= 0,343, p= 0,02) y se observó una correlación moderada inversa entre el valor de p y el IF (Rho= -0,632, p= 0,000) (22).

No se encontró correlación estadísticamente significativa entre el tamaño de muestra, tiempo de seguimiento y pérdidas con el IF (Tabla 3).

Tabla 3. Correlación directa entre las características y el Índice de Fragilidad.

Variable	Coeficiente de correlación (Rho)	Valor de p
Valor de p	*-0,632*	*0,000* [a]
Tamaño de muestra	*0,281*	*0,075*
Número de eventos	*0,343*	*0,028* [a]
Pérdidas	*-0,099*	*0,537*
Tiempo de seguimiento	*-0,017*	*0,914*

[*] Valor de p<0.05.

Discusión

Actualmente dentro de la epidemiología clínica, viene en discusión el valor de p como única medida para valorar significancia estadística. Basados en la discrepancia de resultados entre ensayos clínicos controlados y que la significancia depende de muy pocos eventos dentro del estudio, se esta estudiando cada vez mas el índice de fragilidad, como medida objetiva para valorar la robustez de los ensayos clínicos.

En 41 ECAs sobre diabetes publicados en las revistas de mayor factor de impacto a nivel mundial se encontró que añadiendo pocos eventos al desenlace eliminaba la significancia estadística del estudio, evidenciado en que la mediana del IF fue de 11, y en un 7,3 por ciento el Índice de Fragilidad fue de cero cuando se recalculaba la p mediante la prueba exacta de Fisher.

Además de esto, encontramos que el IF tenía una asociación moderada negativa con el valor de p, y una asociación directa leve con el número de eventos. Resultando que el factor independiente que contribuye a una mayor robustez en los ensayos clínicos es un bajo valor de p calculado con el test exacto de Fisher.

Relación con artículos previos

Nuestro estudio muestra que en un 7,3% de los ensayos clínicos el IF es cero, Walsh et al. [1] reporta un IF de cero en un 10%. Esto evidencia que aproximadamente 1 de cada 10 ensayos

clínicos no tienen significancia estadística después de recalcular el valor de p. También hay similitud en la mediana del IF. Se evidenció una mediana de 11 en el IF comparada con la de Walsh et al. Que reportó 11. Además, se encontró relación directa entre el número de eventos y el IF, sin embargo, la asociación de este estudio fue leve y la del estudio de comparación, moderada. No se encontró asociación entre el IF y el tamaño de muestra comparado con otros estudios.

Ridgeon et al. [3], Khan et al. [11] y Evaniew et al. [7] realizan el cálculo del IF en ECAs sobre cuidado intensivo, cirugía deportiva y cirugía de columna, respectivamente, encontrando mayor fragilidad en estos estudios con una mediana del IF de 2, y un IF de cero entre 17% y 20%. Ridgeon et al. asocian este índice al tamaño del ensayo, levemente al número de centros que fue realizado, y negativamente al valor de p tal cual como nosotros hemos encontrado. No encontraron relación si hubo cegamiento al igual que nosotros. Mientras que Khan et al. y Evaniew et al. encuentran una asociación fuerte y moderada respectivamente con los valores bajos de p y el IF, sin encontrar relación con el tamaño de la muestra al igual que nosotros. Estos resultados subrayan la discrepancia en el tamaño entre los estudios de esos campos con los evaluados en el presente estudio.

Docherty et al. [16] evalúa ECAs en falla cardiaca, siendo el estudio que ha evaluado ensayos clínicos más robustos con una mediana del IF de 26, encontrando un artículo con IF de 0. Ellos encuentran asociación moderada con el valor de p y el valor del IF, sin encontrar relación con el tamaño de muestra, número de eventos, el número de pérdidas y el tipo de desenlace.

Nuestro estudio cae en una zona especial descrita por Walter [6], refiriéndose a ella como Zona 2, en donde los ensayos clínicos son estadísticamente significativos, cuantitativamente significativos pero frágiles. En estos casos se recomendaría dudar del veredicto de la significancia estadística a pesar de lo que aparenta el valor de p o el intervalo de confianza. Mostrándonos que los clínicos están basando las decisiones de la práctica clínica en pocos eventos debido a su alta fragilidad, lo que podría explicar la dificultad en reproducir los resultados de los ensayos [4, 10].

Las principales limitaciones del índice de fragilidad: son la necesidad de aleatorización 1:1 y el hecho de que deba ser usado en variables dicotómicas, excluyendo de esta manera, variables continuas y asignaciones de 2:1 o mayores [12, 17]. Por otro lado, en la actualidad no se cuenta con puntos de corte o valores que sugieran fragilidad o robustez en el resultado arrojado(12).

Implicaciones

Es probable que muchos clínicos no tengan entrenamiento sustancial en estadística y que la interpretación de los valores de p y de los intervalos de confianza sea limitada y hecha de manera intuitiva [3]. Por lo cual se creó una nueva aproximación que demuestra que tan fácil puede cambiar la significancia de un resultado basado en únicamente el valor de p [2]. El Índice

de Fragilidad es un estadístico que ha evolucionado con el tiempo y se ha destilado a una forma sencilla y fácil de usar que permite al lector de forma rápida evaluar la robustez de los resultados, en los cuales pequeños cambios entre los pacientes no cambian la magnitud del estimado ni su significancia. Es por esto por lo que consideramos que el IF es un estadístico simple y práctico de realizar con una fácil interpretación que permite ayudar a analizar la confiabilidad de la significancia estadística de los resultados.

Por otro lado, la validez de un resultado también depende de la precisión de este, y para mejorar la precisión en un ECA se requiere aumentar el número de eventos o desenlaces, y es por esto que, los investigadores aumentan el tamaño de la muestra como medida indirecta, para elevar el número de desenlaces y así hallar una diferencia significativa, sin embargo, el cuándo el número de pérdidas es mayor que el IF, también se vería la fragilidad de los resultados [23]. Sin embargo incluso en ECAs con muestra numerosa se ha encontrado fragilidad como es el caso de LIMIT-2 [3]. Además en este estudio encontramos ECAs con valores altos del IF en estudios con muy poca muestra pero con importante número de desenlaces, como el caso de los ensayos clínicos de cirugía bariátrica, lo que nos corrobora que más importante que el tamaño de la muestra es el número de eventos. Es por eso que aconsejamos que se deba visualizar todo el panorama al evaluar los resultados de un estudio, que hayan ocurrido suficientes eventos en vez del tamaño de la muestra, que se realice el IF en los estudios que se pueda calcular y que el valor de p sea el menor posible como es sugerido por algunos autores, debido a la creciente evidencia que el límite de 0,05 del valor de p no provee evidencia fuerte ni conclusiva en contra de la hipótesis nula [4].

Los resultados de este articulo nos dan una herramienta adicional para interpretar el valor de p de forma más dinámica además del valor por sí sólo, permitiéndonos tomar en cuenta otros factores que en ocasiones son pasados por alto, tales como el número de eventos y de la muestra. La moderada asociación del IF con valor de p nos dice que en un futuro puede ser necesario tener como referencia un menor valor de p para aumentar la robustez de los resultados y así aumentar su confiabilidad para disminuir la necesidad de otros ensayos clínicos de confirmar o desmentir los desenlaces.

Teniendo en cuenta todo esto, nos añadimos a lo que muchos autores refieren sobre recomendar el uso de este estadístico para los ECA [1, 3, 7, 9, 11, 12, 17, 24], debido a su utilidad para asistir al clínico en determinar la confianza del resultado, mejorar el entendimiento de los resultados de los ensayos y de los valores de p, evidenciar que aunque haya un estimado de mayor magnitud estos pueden ser frágiles como por ejemplo en los ensayos clínicos pequeños, pudiendo de esta manera resolver dudas que no son intuitivas [3, 24].

A su vez, encontramos como cada vez son más los estudios que sugieren la inclusión del IF en los resultados de los RCT y ahora también, en las GPC y metaanálisis, de las distintas áreas y especialidades en el área de la salud [12, 14, 24].

Limitaciones

El uso del IF tiene limitaciones importantes, una de estas es la no aplicabilidad en desenlaces continuos, por lo cual no se puede generalizar su uso teniendo en cuenta que en diabetes muchos estudios tienen desenlaces continuos tales como el nivel de la hemoglobina glicosilada, o desenlaces de funcionalidad.

En nuestro estudio hemos analizado una moderada cantidad de ensayos, debido a esto no se puede realizar una clara asociación entre las características de los estudios y el IF.

Carter [5] realiza una crítica al Índice de fragilidad, y afirma que debido a su asociación al valor de p y al número de eventos, este Índice puede ser más robusto únicamente con aumentar el tamaño de la muestra; y refieren que no en todos los ensayos se puede aumentar el tamaño de la muestra para tener un IF adecuada debido a que hay recursos limitados y estos están destinados para una muestra balanceada con la eficacia esperada, por lo tanto es esperable que los resultados dependan de unos pocos eventos. Sin embargo Stern [4] afirma que no es necesario un gran cambio en el tamaño de la muestra para disminuir el valor de p de 0,05 a 0,001.

Otra de las críticas realizadas al IF es que no se puede caer en el error de caracterizar un ensayo de robusto o no robusto únicamente con el valor del IF sin tener en cuenta otros factores tales como el diseño del estudio, la mitigación de sesgos, la magnitud del estimado puntual, entre otros [5, 25].

Conclusiones

En nuestro estudio encontramos en los ensayos clínicos controlados sobre diabetes, en 5 de las revistas con mayor factor de impacto que los resultados estadísticamente significativos dependen de unos pocos eventos, evidenciado por un bajo valor en el Índice de Fragilidad; y que los valores de esta medición están relacionados con el número de eventos y negativamente con el valor de p.

El presente artículo adiciona evidencia para el uso del IF como un estadístico práctico y fácil de interpretar que ayuda a valorar la robustez y la confiabilidad de los resultados de un ensayo clínico. Sin embargo se deben hacer estudios adicionales para aclarar: ¿cómo se debe interpretar el IF?, ¿Cómo se interpretaría si el IF es de 3, 40 o de 100?, ¿Qué punto de corte debemos tomar para valorar un ensayo clínico como robusto o frágil? Y ahora también, ¿se iniciará la implementación del IF en variables continuas? y ¿esto aportará solidez en la evidencia actual?.

Bibliografía

1. Walsh M, Srinathan SK, Mcauley DF, Mrkobrada M, Levine O, Ribic C, et al. The statistical significance of randomized controlled trial results is frequently fragile : a case for a Fragility Index. J Clin Epidemiol. 2014;67(6):622-8.

2. Bennett DA. How to Distinguish between Statistically Significant Results and Clinically Relevant Results. Front Neurol Neurosci. 2016;39:37-49.

3. Ridgeon EE, Young PJ, Bellomo R, Mucchetti M, Lembo R, Landoni G. The Fragility Index in Multicenter Randomized Controlled Critical Care Trials*. Crit Care Med J. 2016;44(7):1278-84.

4. Sterne JAC, Smith GD. Sifting the evidence — what ' s wrong with significance tests ? Br Med J. 2001;322(January):226-30.

5. Carter RE, Mckie PM, Storlie CB. The Fragility Index : a P -value in sheep ' s clothing ? Eur Heart J. 2016;1-3.

6. Walter SD. STATISTICAL SIGNIFICANCE AND FRAGILITY CRITERIA FOR ASSESSING A DIFFERENCE OF TWO PROPORTIONS. J Clin Epidemiol. 1991;44(12):1373-8.

7. Evaniew N, Files C, Smith C, Bhandari M, Ghert M, Walsh M, et al. The fragility of statistically significant findings from randomized trials in spine surgery : a systematic survey. Spine J. 2015;15(10):2188-97.

8. Trimmel H, Landsteiner K, Hospital G, Voelckel WG, Ahmed W, Health S, et al. Does Sample Size Matter When Interpreting the Fragility Index? Crit Care Med J. 2016;44(1):1142-11431. Trimmel H, Landsteiner K, Hospital G,.

9. Tignanelli CJ, Napolitano LM. The Fragility Index in Randomized Clinical Trials as a Means of Optimizing Patient Care. JAMA Surg. 1 de enero de 2019;154(1):74.

10. Contribution O. Contradicted and Initially Stronger Effects in Highly Cited Clinical Research. J Am Med Assoc. 2017;294(2):218-28.

11. Khan M, Evaniew N, Gichuru M, Habib A, Ayeni OR, Bedi A, et al. The Fragility of Statistically Significant Findings From Randomized Trials in Sports Surgery: A Systematic Survey. Am J Sports Med. 2016;1-7.

12. Del Paggio JC, Tannock IF. The fragility of phase 3 trials supporting FDA-approved anticancer medicines: a retrospective analysis. Lancet Oncol. agosto de 2019;20(8):1065-9.

13. Caldwell J-ME, Youssefzadeh K, Limpisvasti O. A method for calculating the fragility index of continuous outcomes. J Clin Epidemiol. agosto de 2021;136:20-5.

14. Goerke K, Parke M, Horn J, Meyer C, Dormire K, White B, et al. Are results from randomized trials in anesthesiology robust or fragile? An analysis using the fragility index. Int J Evid Based Healthc. marzo de 2020;18(1):116-24.

15. Contrast FORA, Two OF. The unit fragility index : an additional appraisal of " statistical significance ". J Clin Epidemiol. 1990;43(2):201-9.

16. Docherty KF, Campbell RT, Jhund PS, Petrie MC, Mcmurray JJ V. How robust are clinical trials in heart failure ? Eur Heart J. 2016;1-10.

17. Skinner M, Tritz D, Farahani C, Ross A, Hamilton T, Vassar M. The fragility of statistically significant results in otolaryngology randomized trials. Am J Otolaryngol. enero de 2019;40(1):61-6.

18. Tzelves L, Chatzikrachtis N, Lazarou L, Mourmouris P, Pinitas A, Tsirkas K, et al. Fragility index of urological literature regarding medical expulsive treatment. World J Urol. octubre de 2021;39(10):3741-6.

19. Maldonado DR, Go CC, Huang BH, Domb BG. The Fragility Index of Hip Arthroscopy Randomized Controlled Trials: A Systematic Survey. Arthrosc J Arthrosc Relat Surg. junio de 2021;37(6):1983-9.

20. Vargas M, Marra A, Buonanno P, Coviello A, Iacovazzo C, Servillo G. Fragility Index and Fragility Quotient in Randomized Controlled Trials on Corticosteroids in ARDS Due to COVID-19 and Non-COVID-19 Etiology. J Clin Med. 14 de noviembre de 2021;10(22):5287.

21. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data Data for Categorical of Observer Agreement The Measurement. Biometrics. 2013;33(1):159-74.

22. Rosa María Martínez Ortega, Leonel C. Tuya Pendás MMOrtega. El coeficiente de correlacion de los rangos de spearman caracterizacion. Rev Haban Cienc Méd Habana. 2009; VIII(2).

23. Khan MS, Ochani RK, Shaikh A, Usman MS, Yamani N, Khan SU, et al. Fragility Index in Cardiovascular Randomized Controlled Trials. Circ Cardiovasc Qual Outcomes [Internet]. diciembre de 2019 [citado 9 de febrero de 2022];12(12). Disponible en: https://www.ahajournals.org/doi/10.1161/CIRCOUTCOMES.119.005755

24. Lin L. Factors that impact fragility index and their visualizations. J Eval Clin Pract. abril de 2021;27(2):356-64.

25. Desnoyers A, Wilson BE, Nadler MB, Amir E. Fragility index of trials supporting approval of anti-cancer drugs in common solid tumours. Cancer Treat Rev. marzo de 2021; 94: 102167.