

Diseño de cohorte

Overview of cohort studies

Chiharu Murata,¹ Luisa Díaz-García,² Isabel Medina-Vera,² Silvestre García-de la Puente,³ Alejandro González-Garay⁴

Resumen

ANTECEDENTES: Un estudio de cohorte es un diseño epidemiológico que pretende estimar la relación de causalidad entre una o más variables y un desenlace de interés. Existen distintos tipos y sus aportaciones a la bibliografía científica han sido decisivos para identificar factores de riesgo para múltiples condiciones clínicas.

OBJETIVO: Analizar las principales características de este diseño para facilitar la interpretación de sus resultados y apoyar su correcta conducción.

DESCRIPCIÓN: Los estudios de cohorte se caracterizan por ser observacionales y longitudinales, sus observaciones permiten identificar claramente la relación causal entre variables (factores de riesgo) y desenlaces de interés (enfermedades, mortalidad, etc.). De acuerdo con la relación temporal de las variables estudiadas pueden clasificarse en cohortes prospectivas, retrospectivas o ambispectivas. La forma de reclutar a los sujetos también puede clasificarlas como incipientes, estáticas, dinámicas, etc. Las comparaciones pueden ser de carácter regional-local (cohortes internas) o, bien, permitir comparaciones con distintas poblaciones (cohortes externas). Los resultados de estos estudios han permitido conocer tasas de incidencia, riesgos relativos, riesgos atribuibles, factores de protección, etc. Aunque se consideran estudios ideales para establecer causalidad, casi siempre son costosos, largos y complicados de realizar.

CONCLUSIONES: Este diseño se considera el patrón de referencia para estimar causalidad entre variables y su aplicación práctica más común es la identificación de factores de riesgo. Conocer las características principales de los estudios de cohorte habilita para analizar críticamente sus contribuciones al campo de la Medicina.

PALABRAS CLAVE: Factores de riesgo; estudio longitudinal; prospectivo; retrospectivo; incidencia.

Abstract

BACKGROUND: A cohort study is an epidemiological design in which the frequency of a disease is compared between populations in order to measure the causality of different risk factors; however, due to their great variety of terms, they hinder the correct interpretation of the results.

OBJECTIVE: to analyze the main characteristics of the cohort study to facilitate the perform and interpretation of its results.

DESCRIPTION: this design is characterized by being a observational and longitudinal study; whose observations identify the causal relationship between the risk factors and the outcomes of interest. According to their variables, the cohorts can be classified as prospective, retrospective or ambispective. In this design, participants who do not present the event of interest at the beginning of the study are included; and after a time then they are compared to identify how many of them developed it and which were the risk factors that they had for developing the disease. Therefore, this type of study allows to know incidence of the event and determine the causal relationship of the different factors expressed by relative risk.

CONCLUSIONS: the cohort study is considered the best design to analyze the causality; However, it is costly and difficult to perform, because it uses different techniques to reduce the sources of bias to ensure the reliability of its results.

KEYWORDS: Risk factors; Longitudinal study; Prospective; Retrospective; Risk factors; Know incidence.

¹ Maestro en Rehabilitación Neurológica, Metodología de la Investigación.

² Maestra en Ciencias, Metodología de la Investigación.

³ Doctor en ciencias, jefe del Departamento, Metodología de la Investigación.

⁴ Doctor en Ciencias, Metodología de la Investigación.

Instituto Nacional de Pediatría, Ciudad de México.

Recibido: 5 de marzo 2018

Aceptado: 21 de septiembre 2018

Correspondencia

Alejandro Gabriel González Garay
pegasso.100@hotmail.com

Este artículo debe citarse como
Chiharu Murata, Díaz-García L, Medina-Vera I, García-de la Puente S, González-Garay A. Diseño de cohorte. Acta Pediatr Mex. 2018;39(6):368-375.



INTRODUCCIÓN

En epidemiología, una cohorte se refiere a un grupo de individuos que comparten ciertas características al inicio del estudio y se les da seguimiento en el tiempo hasta que tienen el evento de interés (muerte, enfermedad) con la finalidad de compararlos y determinar con qué frecuencia sucedió el evento de acuerdo con una determinada característica o exposición a un factor.^{1,2}

Por lo anterior, este diseño se considera el mejor de los estudios observacionales para analizar la causalidad de los fenómenos porque puede identificarse con toda precisión el momento (secuencia temporal) y las circunstancias necesarias (especificidad) para favorecer el evento de interés; sin embargo, debido a su gran variedad de términos, generan confusión y dificultan la interpretación correcta de los resultados; por eso se efectuó la revisión del diseño de cohorte.^{1,2}

Clasificación de los estudios de cohorte

Existen diferentes tipos de estudios de cohorte, dependiendo de la fuente de información, dirección, objetivo de la investigación, forma de reclutamiento y comparación de los participantes y tipo de evento de interés.

Por fuente de información. Se considera que una cohorte es “*prolectiva*” cuando la recolección de los datos de los participantes se obtiene por primera vez por el investigador (se recaba información a partir de observaciones de los individuos de estudio). Es “*retrolectiva*” cuando la información de los sujetos de estudio ya está consignada en alguna fuente de información y el investigador solo la consulta (como la información obtenida de registros clínicos, expedientes).¹⁻³

Por dirección. Los estudios de cohorte “*prospectivos*” son aquellos donde los individuos no

tienen el evento de interés durante su inclusión al estudio y son seguidos durante un tiempo para observar la frecuencia con la que aparece la enfermedad (participantes sanos que viven en la playa y se observa cuántos tendrán cáncer de piel en los próximos 5 años); por esto también reciben el nombre de estudios de *seguimiento o de incidencia*. Una variante de este tipo de cohorte son los estudios “*pronóstico*” donde los participantes inician el estudio con la enfermedad en el mismo estadio y son seguidos en el tiempo para observar su evolución (pacientes con cáncer de piel con tratamiento oncológico seguidos durante 5 años para identificar la frecuencia de muertes por cáncer).³

Los estudios de cohorte se consideran “*retrospectivos o históricos*” cuando los participantes tienen una enfermedad y se estudia qué factores tuvieron en el pasado que hubieran podido provocar la enfermedad; dependen de la calidad de los registros para establecer la exposición y el resultado (pacientes con cáncer de piel en quienes se investiga en registros previos cuáles pudieron haber sido los factores que favorecieron la evolución al cáncer).³

Por objetivo de la investigación. Cuando el principal objetivo del estudio de cohorte es la observación de un parámetro epidemiológico (tasa de incidencia de cáncer de piel en el país en 2015) se denomina “*descriptiva*”; a diferencia de una cohorte “*analítica*” en la que se compara la ocurrencia de un evento de interés entre un grupo de participantes que estuvieron expuestos a un factor de riesgo con respecto a los sujetos que no se expusieron (comparación de la incidencia de cáncer de piel entre participantes que utilizan bloqueador solar diariamente con respecto a quienes nunca se lo aplican).¹⁻³

Por forma del reclutamiento de sus participantes. Se considera cohorte “*fija o cerrada*” cuando el periodo de reclutamiento de los participantes

solo es fijado por los investigadores (solo durante el periodo de enero a julio se incluyeron individuos sanos que vivían en la playa y, posteriormente, se les dio seguimiento durante 5 años para identificar qué proporción de ellos resultó con cáncer de piel; todos los participantes permanecieron el mismo tiempo en la cohorte). La cohorte “dinámica”, por su parte, durante toda la conducción del estudio va reclutando más participantes; por esto el número de individuos puede variar a través del tiempo (durante 5 años se incluyó a una cohorte a individuos sanos que vivían en la playa para identificar qué proporción de ellos tuvo cáncer de piel). Todos los participantes tienen diferentes tiempos de seguimiento porque se van incorporando después del inicio de la cohorte.¹⁻³

Por tipo de comparación. Si los participantes de la cohorte sólo se comparan con sujetos geográficamente de la misma población se denomina cohorte interna (se compara la incidencia de cáncer de piel entre participantes que se aplican bloqueador solar diariamente con respecto a quienes nunca lo usan y que viven en la playa de Los Cabos). Si se comparan con individuos de poblaciones diferentes se considera cohorte externa (se compara la incidencia de cáncer de piel entre participantes que se aplican bloqueador solar diariamente con respecto de quienes nunca lo usan y viven en alguna de las playas de mexicanas).¹⁻³

Por tipo de evento de interés. Los estudios de cohorte pueden caracterizarse por analizar un solo evento de interés en cada sujeto hasta el final de su seguimiento; se denomina “evento simple” (se compara la incidencia de cáncer de piel entre participantes que usan bloqueador solar todos los días con respecto a los participantes que nunca se lo aplican; al presentarlo, esos participantes se retiran de la cohorte). A diferencia de la cohorte de “evento múltiple o recurrente” en la que el evento de interés puede repetirse

durante la estancia del participante en la cohorte (se compara la incidencia de cáncer de piel entre participantes que utilizan bloqueador solar *versus* los que nunca se lo aplican. Cuando a los sujetos les aparece el cáncer se retiran de la cohorte y se les administra tratamiento; posteriormente, al remitir la enfermedad vuelven a incluirse en la cohorte).^{3,4}

Conducción de los estudios de cohorte

La característica clásica de los estudios de cohorte es que se recluta a un grupo de participantes que no tienen el evento de interés, pero todos se encuentran en riesgo de tenerlo. Para su inclusión, los individuos se clasifican según sus factores de riesgo que podrían guardar relación con el resultado en al menos dos grupos:

A) Sujetos sin factores de riesgo de padecer el evento de interés (*grupo no expuesto*) que se compararán con los del **B)** participantes con al menos un factor que pudiera favorecer o proteger la aparición del desenlace (*grupo expuesto*). Posteriormente, se observan a lo largo del tiempo y se cuantifica la proporción de cuántos resultaron con el evento en un tiempo determinado (*incidencia*) en cada grupo para que posteriormente se comparen.³

Duración del seguimiento en cohortes

Uno de los principales retos de este diseño es la duración del estudio. Al ser tan prolongados se facilita la pérdida de los participantes a lo largo del seguimiento o el estudio se interrumpe debido a factores externos (agotamiento de los recursos económicos). Por esto es fundamental conocer el tiempo necesario para que se presente el evento de interés y determinar el tipo de inclusión de los participantes (*cohorte fija o dinámica*), para planificar los gastos antes del inicio del estudio.^{1,3}



Las posibles causas que facilitan la pérdida de participantes en la cohorte pueden deberse a:

- a. Abandono del estudio por falta de interés.
- b. Pobre factibilidad de mantener el apego de los participantes.
- c. Falta de incentivos a los participantes para continuar en el estudio.
- d. Muerte de los participantes por otros motivos.
- e. Pérdidas administrativas generadas por razones externas a la cohorte (agotamiento de los recursos económicos).^{4,5}

Fuentes de sesgo

A pesar de que el estudio de cohorte es uno de los diseños observacionales más sólidos para demostrar evidencia debido a que:

- Controla adecuadamente los errores sistemáticos durante su conducción.
- Permite inferir la causalidad de los fenómenos.
- Determina con precisión la secuencia temporal de los eventos y cuantifica los riesgos.

Sin embargo, tiene algunas debilidades que son fuentes de sesgo:^{3,5}

- *Sesgo de selección.* Sucede en caso de error en el procedimiento utilizado para constituir la población de estudio, que puede tener características aparentemente similares a la población general e impedir la generalización de los resultados (estudiar la mortalidad por cáncer de los trabajadores de una empresa de productos radioactivos y comparar los resultados con los de la población general).⁴
- *Sesgo de clasificación no diferencial.* Se relaciona con la forma en que se conforman

los grupos al inicio del estudio porque algunos participantes pueden tener el evento de interés, pero no detectado al momento de su inclusión al estudio debido a la falta o mal uso de los instrumentos utilizados (se incluyen participantes con cáncer de piel en una etapa muy inicial en la que aún no hay signos ni síntomas y falsamente se consideran sanos).

- Otro sesgo de selección se debe a la pérdida de los participantes durante el seguimiento en un grupo en particular (en la cohorte de cáncer de piel hay una pérdida de los participantes, sobre todo en el grupo que utiliza el protector solar, debido a que la mayoría eran turistas y solo participaron en el estudio cuando estuvieron vacacionando).⁴
- *Sesgo de información,* sucede cuando ésta se obtiene de manera diferente entre los grupos de estudio (los participantes del grupo expuesto son seguidos de manera más cuidadosa que el grupo no expuesto).⁴
- *Sesgo de desgaste,* cuando las pérdidas de los participantes exceden 20% del total incluido en la cohorte, se pierde poder estadístico para demostrar diferencias en los resultados (se incluyeron 100 participantes sanos para analizar la incidencia de cáncer de piel; 50 en el grupo expuesto y 50 en el no expuesto; al finalizar el estudio se perdieron 20 individuos del grupo expuesto y 15 del no expuesto; las pérdidas son de 35%).⁴
- *Sesgo de arrastre,* cuando existen diferencias significativas entre los participantes de los grupos al inicio de la cohorte (en el grupo que utiliza bloqueador solar 90% de los participantes son de raza afroamericana, mientras que en el grupo que no se aplica el bloqueador 80% son caucásicos; esto interfiere en el riesgo de cáncer de piel).⁴

Ventajas y desventajas

Como cualquier otro diseño, los estudios de cohorte tienen ventajas y debilidades con respecto a otros estudios observacionales, como se muestra en el **Cuadro 1**.⁴

Cuadro 1. Principales ventajas y desventajas de los estudios de cohorte

Ventajas	Desventajas
<ul style="list-style-type: none"> • Permiten evaluar directamente la incidencia y el riesgo relativo. • Permiten evaluar adecuadamente la relación causal porque existe una clara secuencia temporal de exposición y enfermedad. • Permiten evaluar los efectos del factor de riesgo de varias enfermedades. 	<ul style="list-style-type: none"> • Pueden ser muy costosos y requerir mucho tiempo para obtener el resultado. • El seguimiento puede ser difícil y las pérdidas durante ese periodo pueden influir en los resultados del estudio. • Pueden introducirse sesgos de información si la identificación de la enfermedad puede estar influida por el conocimiento del estado de exposición del sujeto.

Análisis estadístico

Para cuantificar los resultados en los estudios de cohorte se requiere generar un cuadro de contingencia, que debe contener la información de cuántos participantes, del grupo expuesto y no expuesto al factor de riesgo, tuvieron el evento de interés (**Cuadro 2**). Esto permite calcular tres tipos de cuantificaciones que se utilizarán posteriormente para obtener las medidas de frecuencia y magnitud del efecto.

A) Cuantificaciones

- **Razón:** es la comparación de dos cantidades mutuamente excluyentes [razón = $n_1: n_2$].⁶
- **Proporción:** es la determinación de la parte de interés ante la totalidad; el numerador es parte del denominador [proporción = $a/(a+b)$].⁶

Cuadro 2. Cuadro de contingencia

	Evento	Sin evento	
Grupo expuesto	a	b	$a+b = n_1$
Grupo no expuesto	c	d	$c+d = n_2$
	r1	r2	$a+b+c+d=N$

a: participantes expuestos con el evento de interés
 b: participantes expuestos sin el evento de interés
 c: participantes no expuestos con el evento de interés
 d: participantes no expuestos sin el evento de interés
 n1: total de participantes del grupo expuesto
 n2: total de participantes del grupo no expuesto
 r1: total de participantes con el evento de interés
 r2: total de participantes sin el evento de interés
 N: total de participantes en el estudio

- **Tasa:** es una proporción que expresa la frecuencia de un evento en un periodo determinado ajustado a una constante [tasa = $(r1/N)*k$].⁶

Ejemplo: con la finalidad de conocer la incidencia de reprobación de los alumnos inscritos en la materia de investigación en el año en curso, según el sexo, se recolectó la siguiente información: (**Cuadro 3**)

Cuadro 3. Ejemplo de reprobación según el sexo

Sexo	Reprobados	Aprobados	Total
Hombres	35	10	45
Mujeres	40	15	55
Total	75	25	100

- Razón: *existen 0.82 hombres por cada mujer inscritos en el curso* [$45:55 = 9:11 = 0.82:1$]
- Proporción: *46.6% de los reprobados en el curso eran hombres* [$35/75 = 0.466$]
- Tasa: *existen 7.5 reprobados por cada 10 alumnos inscritos a la materia de investigación en el año en curso* [$(35/100)*10 = 7.5$]

B) Medidas de frecuencia

La forma en que se analiza la frecuencia instantánea de un evento en un tiempo determinado es mediante la incidencia, que puede expresarse en dos formas:

- **Incidencia acumulada** es la proporción del número de participantes que tienen el evento por primera vez con respecto a los individuos en riesgo. $[I = a / n_1]$ (I:incidencia).⁷
- Para calcular su intervalo de confianza al 95% (técnica de estimación que permite establecer un par de valores, entre ellos la incidencia buscada con una confiabilidad del 95 %) solo se realiza la siguiente operación: $[IC95 \% = I \pm 1.96 \sqrt{I * (1-I) / n_1}]$ (IC95 %: intervalo de confianza al 95 %; I:incidencia; $z_{\alpha/2}$: valor de significancia a dos colas = 1.96; $\sqrt{\quad}$: raíz cuadrada).⁷
- **Tasa de incidencia** es la tasa del número de participantes con eventos nuevos, ajustada por una constante, por ejemplo: el tiempo total observado de todos los individuos en riesgo. $[TI = a / n_1 * k]$ (TI:tasa de incidencia; k:constante).⁷

C) Medidas de efecto

Estas medidas expresan la comparación de la incidencia de eventos entre los diferentes grupos que componen la cohorte:

- **Razón de riesgo (riesgo relativo)** es la proporción de la incidencia entre el grupo expuesto y no expuesto. $[RR = (a/n_1) / (c/n_2)]$
- Para calcular su intervalo de confianza se requiere realizar lo siguiente: $[IC95 \% = \text{antilogaritmo}(\log \text{natural}(RR) \pm (1.96 \sqrt{(1/a)-(1/n_1) + (1/c) - (1/n_2)}))] (IC95 \%: \text{intervalo de confianza al } 95\%; \log \text{ natural: } \log \text{aritmo natural; RR: razón de riesgo; } \sqrt{\quad}: \text{ raíz cuadrada}).⁷$

- **Diferencia de riesgo (riesgo atribuible)** es la diferencia de la incidencia del grupo expuesto con respecto al grupo no expuesto. $[RA \text{ o } DR = I_{\text{exp}} - I_{\text{no exp}}]$ (DR: diferencia de riesgo; I:incidencia; exp:grupo de expuestos; no exp: grupo no expuesto).⁷

En ambas medidas, si el cociente es mayor a 1 indica que el factor al que se exponen los participantes es “*riesgo*” de tener el evento; a diferencia de cuando el cociente es menor a 1, expresa que el factor de exposición es “*protector*”.

- **Proporción de riesgo atribuible** es la proporción de individuos con el evento que pueden prevenirse al eliminarse el factor de riesgo en el grupo expuesto. $[RA \% = (DR/I_{\text{exp}}) * 100]$ (RA:riesgo atribuible; DR:diferencia de riesgo; I:incidencia; exp:grupo expuesto).⁷

Otras medidas de efecto que suelen utilizarse en los estudios de cohorte, con la finalidad de ajustar múltiples factores confusores, son los modelos lineales con efectos mixtos, regresión logística, análisis de muestras repetidas, series de tiempo y supervivencia.⁷

Enseguida se ilustran esos conceptos mediante un ejemplo:

Con la finalidad de estimar la repercusión del tabaquismo en relación con la muerte por cáncer, se realizó un seguimiento del hábito tabáquico en hombres entre 40 y 70 años de edad durante 10 años, y se observó lo siguiente: (**Cuadro 4**)⁸

Cuadro 4. Ejemplo del impacto de tabaquismo y muerte

Exposición a Tabaquismo	# Muertos	# Vivos	Total
Fumadores	840	22,421	23,261
No fumadores	220	10,618	10,838
Total	1,060	33,039	34,099

#: número de participantes

- Razón; en los fumadores existen 3.8 muertes por cada muerte en los no fumadores [840:220 = 3.8:1]
- Proporción: 79% de las muertes sucedieron en los fumadores [840/1,060 = 0.79]
- Tasa de incidencia: existen 3.1 muertes por cada 100 decesos en 10 años [(1,060/34,099)*100 = 3.1]
- IA de los expuestos: existen 0.036 (IC95% 0.0336 - 0.0384) o 3.6% (IC95% 3.36% - 3.84%) muertes en los fumadores a 10 años [IA 840/23,261 = 0.036] [IC95% 0.036 ± 1.96 0.036 * (1-0.036)/23,261]
- IA de los no expuestos: existen 0.020 (IC95% 0.0178 - 0.0231) ó 2.02% (IC95% 1.78% - 2.31%) muertes en los fumadores a 10 años [IA 220/10,838 = 0.020] [IC95% 0.020 ± 1.96 √0.020*(1-0.020)/10,838]
- RR: existe un riesgo de 1.78 (IC95% 1.53 a 2.06) veces de morir debido al tabaquismo en comparación con los no fumadores [RR : (840/23261)/(220/10,838) = 1.78]
- IC95% RR: existe una probabilidad de 95% que el riesgo relativo se encuentre entre 1.53 y 2.06 sin deberse al azar [antilog((logn 1.78) ± 1.96√(1/840)-(1/23,261) + (1/220)-(1/10,838)) = antilog (.5766 ± (1.96 * 0.0748) = 1.537, 2.061]
- DR: 15 de cada 1000 fumadores fallecen debido a sus hábitos tabáquicos. [(840/23,261) - (220/10,838) = 0.0158]
- RA%: se evitaría 43.7% de muertes si se eliminara el tabaquismo [(0.0158/(840/23,261)) * 100]

Consideraciones éticas en los estudios de cohorte

Cuando la cohorte es prospectiva debe solicitarse la carta de consentimiento-asentimiento informado, siguiendo los lineamientos locales de

los comités de ética e investigación, a diferencia de la cohorte retrospectiva en la que solo es necesario declarar el manejo de la información de acuerdo con los lineamientos locales de los comités de ética e investigación, para asegurar su confidencialidad.

De acuerdo con la Ley General de Salud y la Norma Oficial Mexicana (NOM 012) de investigación en seres humanos debe considerarse el riesgo de las maniobras dependientes de los procedimientos de medición o cuantificación de la exposición.⁴

CONCLUSIONES

El diseño de cohorte es el mejor estudio observacional para demostrar la causalidad de los fenómenos debido a la minimización de sesgos importantes durante su conducción; sin embargo, tiene importantes desventajas: el costo y factibilidad para su conducción. Para garantizar la confiabilidad de sus resultados es necesario identificar y controlar los sesgos, para que de esta manera pueda evitarse la generación de inferencias apócrifas que puedan afectar la toma de decisiones por parte del personal sanitario.

Agradecimientos

Se agradece la colaboración de la M en C Anahí Anzo Osorio en la presente revisión.

REFERENCIAS

1. Argimon Pallás JP. Métodos de investigación clínica y epidemiológica, 4ª ed. Madrid: Elsevier 2012; 64-72.
2. Díaz Ceballos FJ. Introducción a los estudios de cohorte en epidemiología y al análisis de supervivencia. Med UNAB. 2005;8(1):43-53.
3. Lazcano-Ponce E, Fernández E, Salazar-Martínez E, Hernández-Avila M. Estudios de cohorte. Metodología, sesgos y aplicación. Salud Pública de México. 2000;42(3):230-241.

4. Szklo M, Nieto J. Epidemiology, 2nd ed. Beyond the Basics. Boston: Jones and Bartlett Publishers 2007; 45-104.
5. Bel Ferré MN, Inglés Novell M, J.L. Piñol Moreso JL. Estudios de cohorte. Fisioterapia. 2009;31(5):218-23.
6. Elandt-Johnson RC. Definition of rates: some remarks on their use and misuse. Am J Epidemiol. 1975;102(4):267-71.
7. Rothman KJ. Epidemiología Moderna. México: Ediciones Díaz Santos, 1987; 29-48.
8. Inoue M, Hanaoka T, Sasazuki S, Sobue T, Tsugane S. Impact of tobacco smoking on subsequent risk among middle-aged Japanese men and women: data from large-scale population-based cohort study in Japan- the JPHC study. Preventive Medicine. 2004; 38:516-22.

Consulte **Acta Pediátrica de México** en internet:

www.actapediatrica.org.mx



:actapediatricademexico



:@ActaPedMex