

# Las variables en el proceso de investigación en salud: importancia, clasificación y forma de presentación en protocolos de investigación\*

Miguel Ángel Castro-Jiménez, MD MSc<sup>\*</sup>

Luis Alfonso Díaz-Martínez, MD MSc<sup>†§</sup>

## Resumen

La planeación adecuada de una investigación en ciencias de la salud es necesaria para determinar la información que debe ser recolectada y que permitirá lograr los objetivos del estudio. Esta información conforma el grupo de variables que permiten, a su vez, realizar una aproximación *a priori* a la metodología de análisis de los datos. Este artículo académico busca aclarar la terminología empleada al hablar de variables y mostrar una forma sencilla de su presentación en un protocolo o proyecto de investigación. [Castro-Jiménez MA, Díaz-Martínez LA. *Las variables en el proceso de investigación en salud: importancia, clasificación y forma de presentación en protocolos de investigación*. MedUNAB 2009; 12: 151-156].

**Palabras clave:** Variables epidemiológicas, Metodología, Factores de riesgo, Proyectos de investigación.

## Introducción

La necesidad de establecer la etiología de los eventos en salud ha abierto la posibilidad de desarrollar métodos que permiten al ser humano moderno aproximarse al conocimiento de la realidad. La epidemiología y la bioestadística, dos ciencias paralelas y relacionadas con la ciencia médica, aportan herramientas para permitir a los científicos descubrir las causas de las enfermedades, conocer el impacto de su aparición sobre las diferentes poblaciones o personas y determinar las medidas de intervención apropiadas para su control. El uso adecuado de

## Summary

A careful planning of any epidemiological or clinical research, including thesis in any area of the health sciences, is necessary to determine what information should be gathered in order to achieve the study's objectives. This information defines the set of variables that, in turn, make an *a priori* approach to the methodology of data analysis. This academic paper is aimed to clarify the terminology used when talking about variables and show a simple way to describe them on a research project or protocol. [Castro-Jiménez MA, Díaz-Martínez LA. *Variables in health research process: importance, classification, and show way in research protocols*. MedUNAB 2009; 12: 151-156].

**Key words:** Confounding variable, Methodology, Risk factors, Research design.

estas alternativas metodológicas requiere tanto un diseño de investigación bien planeado que ayude a definir la información que se recolectará, como un conocimiento estadístico y epidemiológico básico, aunque bien fundamentado, por parte del investigador.

Cada día se publica una gran cantidad de artículos con métodos de análisis novedosos y especializados que abren más la brecha del *no conocimiento* entre quienes deben utilizarlos, incluso en los investigadores más consagrados. Como no es posible que las mismas pruebas estadísticas puedan utilizarse en todos los datos porque su uso depende

\* El Dr. Díaz es el Editor de MedUNAB, por lo que el proceso de este artículo lo realizó el Comité Editorial de la revista, donde uno de sus miembros actuó como editor *ad hoc*.

<sup>\*</sup> Grupo de Virología, Instituto Nacional de Salud, Bogotá, Colombia.

<sup>†</sup> Centro de Investigaciones Biomédicas, Facultad de Ciencias de la Salud, Universidad Autónoma de Bucaramanga, Bucaramanga, Colombia.

<sup>§</sup> Departamento de Pediatría, Escuela de Medicina, Facultad de Salud, Universidad Industrial de Santander, Bucaramanga, Colombia.

**Correspondencia:** Dr. Castro-Jiménez, Grupo de Virología, Red Nacional de Laboratorios. Instituto Nacional de Salud. Avenida Calle 26 No. 51-20 - Zona 6 CAN, Bogotá, Colombia. E-mail: mcastro2505@yahoo.es

Artículo recibido: agosto 2 de 2009; aceptado, noviembre 10 de 2009.

de la forma como estos fueron recolectados y, por lo tanto, del tipo de variable, es necesario, entonces, conocer algunas definiciones básicas acerca de las variables, pues en gran medida son éstas las que definen cuáles pruebas deberían ser usadas. Esta es una revisión basada en diversos textos de epidemiología y estadística<sup>1-7</sup> y en la experiencia propia. Este artículo busca aclarar la terminología empleada al hablar de variables y mostrar una forma sencilla de su presentación en un protocolo o proyecto de investigación.

## Terminología básica

Algunas definiciones que deben tenerse en cuenta al momento de desarrollar un proyecto o tesis o al plantear cualquier investigación clínica o epidemiológica son:

**Variables.** Son aquellas características que cambian en un mismo individuo, grupo, comunidad o población *con el avance del tiempo* o que varían entre los individuos, grupos, comunidades o poblaciones comparadas *en el mismo instante*. Cada situación (por ejemplo, cada consulta que haga un paciente o cada factura que se genere), individuo o grupo que se analice recibe el nombre de unidad de observación.<sup>8</sup>

Para ilustrar esto, imaginemos un estudio que se realizará entre pacientes con diagnóstico de cáncer de pulmón que asisten a un hospital. Se utiliza una encuesta para determinar la presencia o el antecedente de contacto con factores como el consumo de derivados del tabaco, alcohol y marihuana; esto último constituye lo que se llama la exposición. Las respuestas de estos pacientes van a ser comparadas con las de otros sin esta enfermedad. En este ejemplo se observan diferentes variables que pueden ser objeto de análisis: la presencia de la enfermedad, los antecedentes de tabaquismo, consumo de alcohol y de marihuana. Si este estudio involucrara varios hospitales o municipios, la institución tratante y el municipio de atención también pueden ser variables de interés.

Otros ejemplos de variables que usualmente se incluyen en los protocolos de investigación son lugar de residencia, antecedentes patológicos, estado civil, nivel socio-económico, adherencia a un tratamiento formulado por el médico o la creencia en Dios. En fin, son todos datos que deben recolectarse para el cumplimiento del objetivo de la investigación.

**Definición operacional.** Es la aclaración del significado de cada una de las variables de un estudio, incluyendo todos los detalles del procedimiento que llevará a establecer el valor de cada variable en cada unidad de observación.<sup>9</sup> Aunque la mayoría de las variables de uso frecuente pueden ser de fácil comprensión, existen otras que no lo son y, por lo tanto, requerirán de una definición estandarizada. Por ejemplo: imaginemos un estudio específico, en el que se requiere tener el dato de la variable *temperatura corporal*, la cual se

podría definir así: es el grado de calor superficial de una persona medida con una exactitud de 0.01 grados centígrados mediante el uso de un termómetro digital Temperar<sup>®</sup>, previamente calibrado, ubicado en la axila derecha.

Ahora bien, en ciertas circunstancias, una variable puede tener diversas formas de ser definidas operacionalmente. Un ejemplo muy útil es la edad de un paciente; la definimos, por lo general, como el tiempo que transcurre entre el nacimiento y el momento que se da el suceso que es analizado. Sin embargo, podemos recolectar este dato de diferente manera, algunos más precisos que otros: podemos preguntarle simplemente cuántos años tiene, o indagar sobre la fecha de nacimiento y calcular la edad a partir de esa fecha y la del suceso en observación o, incluso, pedir cualquier documento de identidad que corrobore esto. Todo depende del grado de precisión que se necesita.

**Dato.** Es el valor real, tan aproximado como se necesite, que toma la variable estudiada en cada unidad de observación del estudio y que se obtiene luego de un proceso específico de recolección de datos; estos procesos pueden ser entrevista directa, pruebas de laboratorio o información de la historia clínica, entre otros. Un dato puede ser cuantitativo (tiene 74 años y una presión sistólica de 110 mm Hg) o cualitativo (se encuentra en el grupo de mayores de 60 años y es normotenso).

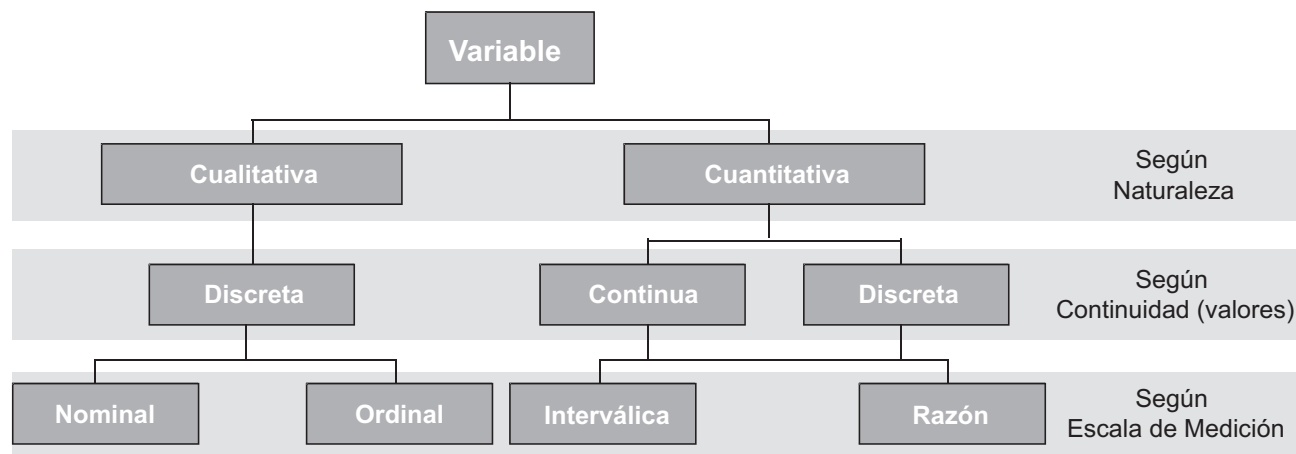
## Clasificación de las variables

Cuando el investigador necesita información de una variable específica entre diferentes unidades de observación, debe planear la mejor forma de recolección, clasificarla y establecer cuáles pruebas estadísticas podría utilizar en su análisis.

Una misma variable puede clasificarse de distintas maneras, de acuerdo con sus características, dentro de una determinada investigación. La figura 1 muestra un diagrama que resume estas clasificaciones y sus relaciones. Estas son:<sup>8,9</sup>

- Según su naturaleza, en *cualitativas o cuantitativas*
- Según los valores (continuidad) que toma, en *discretas y continuas*
- Según la escala de medición, en *nominal, ordinal, interválica y de razón (o proporcional)*
- Según la función desempeñada, en *dependiente, independiente y de confusión*.

**Variable cualitativa.** Es aquella que expresa una característica que no es posible medir ni cuantificar; sólo se puede calificar. Es decir, los posibles valores que tome la variable (también llamados categorías) no son medibles dado que no implican magnitud (no se expresan en unidades de medida como metros, años o decibeles) ni es cuantificable (no podemos decir cuántas veces pasa algo).



**Figura 1.** Diferentes clasificaciones de las variables y sus relaciones\*

\* Según el objetivo y diseño del estudio clínico o epidemiológico, cualquier tipo de variable podría desempeñar una función como dependiente, independiente o de confusión

En ocasiones, una variable puede tener sólo dos categorías, por lo que se llaman también dicotómicas. Un ejemplo de pregunta que representa la variable que se va a indagar y sus posibles respuestas son: ¿tuvo fiebre antes de llegar al hospital?: *sí o no*; ¿cómo está la presión arterial sistólica de este paciente?: *baja, normal o alta*.

Debe tenerse en cuenta que, usualmente, a la hora de almacenar los datos para su posterior análisis, a cada opción de respuesta se le puede asignar un código numérico. Así, en las variables cualitativas, este número es un resumen de lo que significa cada posible respuesta (o categoría) de la variable, pero no es una cantidad. Por ejemplo, al codificar la variable *sexo* se puede asignar el valor 1 (uno) al encuestado “masculino” y 2 (dos) o 0 (cero) al “femenino”; no por esto la variable *sexo* deja de ser cualitativa, ni mucho menos la categoría “1” será más que la “0” o se podrán hacer operaciones con ellas. Sería muy ingenuo sacar el promedio de los ceros y unos que codifican la variable *sexo* de una investigación. Dependiendo del programa que va a ser utilizado para el análisis, generalmente se prefiere codificar como cero (0) o dos (2) a la ausencia de enfermedad o de exposición a algún factor específico; en cambio, el uno (1) usualmente indica presencia de enfermedad o antecedente de exposición al factor de riesgo (por ejemplo: no fumador → 0; fumador → 1). En cualquier situación, si usa una codificación propia, debe tener claro lo que significa cada código para cada variable.

Una variable que usualmente genera confusión es la que indica el estado clínico de una enfermedad, la cual se clasifica como 0, 1, 2, 3 (o en sus números romanos correspondientes); aunque es una numeración que implica un gradiente, es arbitraria y representa diferentes niveles de avance no cuantificables de la enfermedad en el individuo; una ampliación sobre estos se puede ver en el apartado sobre variables ordinales.

**Variable cuantitativa.** Es aquella cuyos datos provienen de una característica que permite ser cuantificada o medida

(siempre que se mide algo se expresa en unidades de medida); usualmente está definida por un valor numérico según los valores reales de la variable. Por ejemplo, la edad es cuantitativa cuando se aclara que el paciente tiene cinco, siete o cien años; o el número de matrimonios que una persona ha tenido en su vida, los cuales son 0, 1, 2 o los que haya decidido tener.

**Variable discreta.** Es una variable cuantitativa que sólo permite valores enteros, sin que exista la posibilidad de intermedios entre dos enteros. Por ejemplo, la variable *número de hijos* permite decir cero, uno, dos, tres u otro valor, pero siempre es un entero. Es posible llegar a considerar a las variables discretas también como variables cualitativas porque en determinado momento se tiene o no la característica. Por ejemplo: el color de ojos puede ser azul, verde o café, sin dar la posibilidad de valores intermedios.

**Variable continua.** Es una variable cuantitativa que puede tomar cualquier valor, sea entero o fraccionario, entre los valores límites mínimo y máximo. Por ejemplo: el peso: 64,3 kilogramos; la estatura: 173.5 centímetros; la temperatura del paciente: 37.5 grados centígrados; el valor del colesterol HDL: 45.5 mg/dL. Un detalle que permite identificar a las variables continuas es que el valor entre cada unidad es el mismo (hay el mismo kilómetro de diferencia entre medir 45 y 46 km que entre 73.4 y 74.4 km).

Una recomendación importante es que, durante la recolección de datos de las variables, particularmente las continuas o las discretas, el dato se capture tal cual como se genera, crudo, sin agruparlo. Si es necesario formar categorías o grupos, lo mejor es hacerlo durante la fase de análisis. Por ejemplo, en un trabajo se decidió evaluar en mujeres con resultado indeterminado de la mamografía, si había diferencias en el resultado de la biopsia (maligna o benigna) según la edad. La recolección de la información fue realizada a partir de los datos obtenidos por la historia clínica. En estas circunstancias, es mejor recolectar la información sobre la edad a partir de la historia clínica o del

interrogatorio directo a la paciente, tal cual como se halla. Más adelante, durante el análisis de los datos, se decidirá qué es mejor, si utilizar la información de la edad en forma continua o en categorías (por ejemplo, ubicarla en el grupo de 35 años o menos). Es posible que resulte que no hay diferencias importantes en las mujeres al agrupar *a priori* la edad, pero, al recolectarse en forma cuantitativa *pueden explorarse otras formas de agrupar* a las pacientes por edad (por ejemplo, menor o igual a 30 años y más de 30 años) o, simplemente, analizar la edad en forma continua, algo que no se podría hacer si al momento de recolectar la edad fue incluida en agrupaciones ya definidas. No hacerlo así, es desperdiciar una oportunidad de análisis con un potencial de mayor valor y utilidad.

**Variable nominal.** Es una variable cualitativa en la que sus opciones de respuesta no tienen un orden de importancia definido ni representan un gradiente; es decir, ninguna posible respuesta es más relevante, buena o mala que otra, aún si se expresan en números. Tal es el caso de la diabetes tipo 1 y diabetes tipo 2, en las que el número no representa ni una cantidad ni algún tipo de gradiente y pueden presentarse en cualquier orden ya que ninguno denota mayor importancia; son tan solo caracteres alfanuméricos de igual significado que cualquier otro. Ahora bien, a veces cuando se montan las bases de datos, estas categorías se recodifican con números, que igualmente solo diferencian una característica de otra pero sin que representen una cantidad o un orden; por ejemplo: el sexo puede definirse como hombre (1) y mujer (0), o mujer (1) y hombre (0).

**Variable ordinal.** En estas variables existe un orden implícito en los datos, a manera de gradiente, aunque este orden no significa que las distancias entre las diferentes categorías de la variable sean iguales. Un ejemplo es la clasificación de los estadios clínicos de los tumores, que pueden ser clasificados como estadio 0 (*in situ*), estadio I (invasor temprano), estadio II (invasor intermedio), estadio III (invasor avanzado) y IV (metástasis a distancia). Esta variable indica severidad clínica del tumor pero no significa que el estadio IV sea el doble de avanzado que el estadio II, ni que la distancia (o aumento en severidad) que existe entre estadio II y III sea la misma que entre estadio III y IV. Otro ejemplo de una variable ordinal es la respuesta del paciente ante una terapia analgésica, esto es, mejoría total, mejoría parcial, sigue igual o empeora.

**Variable interválica.** Son datos recolectados en forma cuantitativa pero donde el valor de inflexión entre positivos y negativos (el llamado cero) de la escala es arbitrario. El mejor ejemplo es la temperatura medida en grados centígrados, donde 0° (cero grados centígrados) no implica ausencia de temperatura.

**Variable de razón o proporcional.** En éstas, los datos fueron recolectados en forma continua pero el cero de la escala es real; no hay valores posibles bajo ese punto. Por ejemplo: cero (0) gramos indica la ausencia de peso o al medir temperatura en grados Kelvin, el cero es absoluto.

**Variable dependiente.** Es la variable principal que se está estudiando y de la que se quiere saber algunas características tales como causas, factores pronósticos, frecuencia, etc. Usualmente es aquella que está inmersa en el objetivo general o en el título de una investigación. También puede ser llamada variable de resultado, de salida, desenlace o, simplemente, enfermedad (u otro evento) en estudio.

**Variable independiente.** Es aquella que puede explicar el porqué del comportamiento de una variable dependiente. También puede llamarse exposición, variable de explicación, factores de riesgo o factores protectores.

Esto se puede ilustrar con un ejemplo. Imaginemos un estudio cuyo objetivo es determinar las causas de hipertensión arterial en la población de estrato 2 de Bogotá, Colombia. Lo que se quiere saber (la variable dependiente) es frecuencia de ausencia de hipertensión (HTA), así como entender cuáles factores pueden explicar su ocurrencia, como pueden ser tabaquismo, consumo de bebidas alcohólicas, estrés, dieta de esa población, entre otras (las variables que explican el comportamiento de la dependiente, o sea, las variables independientes). De otra manera:

*HTA = tabaquismo + alcohol + edad + respuesta al estrés + obesidad + tipo de dieta + ... ..*

Ahora, imagínese otro estudio que se hace en forma casi simultánea en la misma población y en el que se busca determinar las causas de otra enfermedad cardiovascular, el infarto agudo del miocardio (IAM); en éste se recolectan como variables independientes a la edad y los antecedentes de hipertensión, diabetes y tabaquismo. Se podría generar una expresión similar a la anterior:

*IAM = edad + HTA + diabetes + tabaquismo + tipo de dieta + alcohol + ...*

Obsérvese que la HTA pasó de ser la variable dependiente en el primer estudio a ser variable independiente o de explicación en el segundo. Puede observar también que a pesar de que las variables dependientes (HTA e IAM) son diferentes, algunas variables independientes son iguales; es decir, una variable independiente puede explicar la aparición de más de una enfermedad. Esto es común en la investigación clínica y epidemiológica.

**Variable de confusión.** Para entender mejor lo que es una variable de confusión, imaginemos un estudio cualquiera en la que se evalúa la posible relación entre un evento (I) de interés (como una enfermedad o el uso de un servicio de salud) y una exposición (V) que se cree que es un posible factor de riesgo (o protector). Es posible que exista una segunda exposición (C) que se conoce que está asociada tanto con I como con V.

Por lo tanto, encontrar una asociación entre V e I puede ser malinterpretada si no se tiene en cuenta la presencia de C, la

cual confunde la verdadera relación que existe entre V e I. Por eso a C se le llama variable de confusión. Un criterio adicional que es importante para definir si una variable es o no un factor de confusión es que no puede encontrarse en la vía causal entre I y V, es decir, no puede ser una lesión precedente a la enfermedad (I).

Por ejemplo, en un estudio se encontró que a mayor número de hijos, mayor era el riesgo de hipertensión arterial en las mujeres. Los investigadores no se explicaban esta aparente relación, la cual no había sido informada previamente en la literatura; sin embargo, fue desmentida cuando analizaron el papel que juega la edad tanto en la probabilidad que una mujer tenga hijos como que sea hipertensa.

Se sabe que a mayor edad hay mayor probabilidad de tener esta patología; también es cierto que la probabilidad de tener hijos aumenta a mayor edad. Cuando los investigadores analizaron la asociación entre el número de hijos y la presencia de hipertensión arterial pero ajustando por la edad, se observó que el aparente mayor riesgo dado por el número de hijos desaparecía. La edad cumple entonces el papel de factor de confusión.

El ajuste o tratamiento de la confusión durante el análisis de los datos se puede lograr por medio de dos métodos llamados estratificación y regresión; igualmente, es posible preverla al planear una investigación y usar otras tres estrategias metodológicas, el apareamiento (o emparejamiento), la aleatorización y la restricción de la población.<sup>10</sup> Una explicación más detallada está por fuera del objetivo de este artículo.

## La presentación de las variables en un proyecto de investigación

Las variables se definen, clasifican y presentan dentro de la sección “Materiales y métodos” o “Metodología” de un proyecto de investigación o en sus subsecuentes informes. En la tabla 1 se puede apreciar la aproximación propuesta en este artículo. Brevemente, esta tabla contiene las siguientes columnas: variable, unidad de medición, nombre dado en la base de datos, tipo (clasificación de la variable), definición para el estudio, valores límites (en cualitativas se incluirían los códigos asignados a cada respuesta) y, en forma opcional, sobre todo para estudiantes que están dando sus primeros pasos académicos por la investigación, un ejemplo de su interpretación.

## Análisis de los datos generados en las variables

El análisis de los datos usualmente inicia con la descripción de los resultados de cada variable (llamado también análisis univariado); luego se pasa a hacer cruces, por lo general, mostrando el comportamiento de cada variable independiente según las categorías de la dependiente (análisis bivariado) y finalmente, se procede a técnicas más sofisticadas como son los análisis estratificado y multivariado. Abundar en estos tópicos va más allá del alcance de este escrito, pero es necesario hacer algunas

**Tabla 1.** Ejemplo de definición de variables en un estudio realizado en menores de 15 años para establecer causas de hipertensión arterial pediátrica.

Variable (unidad de medida)	Tipo de variable	Definición para el estudio	Valores límites (códigos)	Ejemplo de interpretación
Edad (años) En la base: edadnino	Cuantitativa, discreta, de razón, independiente	Edad <i>en años cumplidos</i> referida por la persona cargo del niño a estudio	0-14 años	La edad media de las niñas del estudio fue de 10 años
Azúcar en sangre (mg/dl) En la base: glicem8h	Cuantitativa, continua, de razón independiente	Cifra que muestra el nivel de glicemia en cada sujeto luego de 8 horas de ayuno y observada en el informe de glicemia realizado por el laboratorio clínico contratado.	40-600 mg/dL* Hipo: <70 1 Normo: 70-130 2 Hiper: >130 3	La glicemia de uno de los sujetos de investigación fue de 140 mg/dl
Sexo En la base: masculino	Cualitativa, nominal, discreta, independiente	Sexo del menor informado por la persona a cargo del menor	Femenino 0 Masculino 1	El 50% de los participantes fueron de sexo masculino
Presión arterial En la base: hta	Cualitativa, nominal, discreta, dependiente	Es la clasificación de la HTA según lo referido por la madre	Alta 1 Normal 2 Baja 3	La HTA estuvo presente en el 20% de los niños

\* La división de este ejemplo es arbitraria y no pretende representar los que son aceptados en la actualidad.

anotaciones relevantes con miras a que el lector pueda abordar de una mejor manera estas temáticas en otros documentos.

El análisis univariado es terreno de la llamada estadística descriptiva e incluye el uso de proporciones y de medidas de tendencia central y de dispersión.<sup>11</sup> Igualmente, su presentación se hace por medio de tablas o gráficas.<sup>12</sup> Por otro lado, el objetivo básico del análisis bivariado es ver si hay diferencias significativas en una variable según las categorías de una segunda. Usualmente se trabaja con muestras, es decir, con una fracción que se espera sea representativa de todas las unidades de observación de la población de donde se toma la muestra. Entonces el objetivo del análisis bivariado no es solo ver las diferencias en los resultados de los datos de la variables independientes según las categorías de la dependiente, sino también establecer hasta qué punto las diferencias que se observen es producto del azar que está implícito en un proceso de muestreo adecuado.

Esta consideración incluye el concepto de error tipo I, que es aquel que se comete cuando se aceptan como válidas para la población diferencias que se aprecian en la muestra, cuando en la realidad, estas diferencias no son ciertas sino que son producto del proceso de muestreo. De esta forma, las pruebas estadísticas generan una probabilidad, que no es más que el riesgo de cometer el error tipo I, y que es más conocido como el *valor de p*, o simplemente, *p*. Hay que tener en cuenta que este concepto solo es aplicable en condiciones en donde el proceso de muestreo es adecuado, porque en los casos en que este proceso es inadecuado (la palabra técnica es *sesgado*),<sup>13</sup> las pruebas estadísticas no tienen sentido alguno, ya que este fenómeno es un error sistemático y no aleatorio. La prueba estadística más adecuada depende del tipo de datos que se va a analizar, del diseño del estudio y la forma como se seleccionaron las unidades de observación, y del cumplimiento de ciertas premisas que deben tener los datos frente algunas pruebas, temas que van más allá del alcance de este artículo.

## Conclusión

La selección, recolección y definición adecuadas de las variables en un estudio clínico o epidemiológico no sólo aportan la información requerida para cumplir el objetivo del trabajo, sino que permiten visualizar previamente la validez del enfoque metodológico propuesto y el tipo de análisis estadístico que se requiere. Bajo esta perspectiva, es necesario el conocimiento amplio de qué son las variables, su definición y usos, lo cual fue el objetivo de este manuscrito.

## Referencias

1. Woodwark M. Epidemiology study design and data analysis. London; Chapman & Hall/CRC, 1999.
2. Norman GR, Streiner dL. Bioestadística. México: Mosby/Doyma Libros, 1996.
3. Hill AB. Principios de estadística médica. Madrid: Librería ElAteneo, 3 ed, 1965.
4. Milton JS. Estadística para biología y ciencias de la Salud. México, McGraw-Hill, 2001.
5. Altman DG. Practical statistics for medical research. London; Chapman & Hall, 1991.
6. Daniel WW. Bioestadística: base para el análisis de las ciencias de la salud. Bogotá: Editorial Limusa y Noriega Editores, 2000.
7. Dos Santos-Silva I. Epidemiología del cáncer: principios y métodos. Lyon: Agencia Internacional de Investigación sobre el Cáncer, 1999.
8. Altman DG, Bland JM. Units of analysis. Br Med J 1997; 314:1874.
9. Moreno-Altamirano A, López-Moreno S, Corcho-Berdugo A. Principales medidas en epidemiología. Salud Pub Mex 2000; 42:337-48.
10. Bland JM, Altman DG. Matching. BMJ 1994; 309:1128.
11. Altman DG, Bland JM. Presentation of numerical data. BMJ 1996; 312:572.
12. Altman DG, Bland JM. Quantiles, quintiles, centiles, and other quantiles. BMJ 1994; 309: 996.
13. Hernández-Ávila M, Garrido F, Salazar-Martínez S. Sesgos en estudios epidemiológicos. Salud Pub Mex 2000; 42:438-46.