

Anormalidad no significa enfermedad

El trabajo de Rodríguez, Torres, Nóbrega, Soto y Tinedo, sobre los valores normales de espirometría en la población venezolana de Valencia, de éste número de la Revista Colombiana de Neumología, merece una apreciación crítica con respecto a su validez, a la naturaleza de los datos obtenidos y a la aplicabilidad de ellos en la práctica clínica cotidiana.

Se trata en esencia, de un estudio de *Corte Transversal* (con componente analítico) de base comunitaria, en el que una población elegible de 980 estudiantes escolares urbanos, de 6 a 18 años, de seis escuelas, de la ciudad de Valencia, son contactados a través de

sus profesores para que sus padres o cuidadores den el consentimiento para la participación en el estudio.

La muestra estuvo finalmente constituida por una serie de individuos dispuestos a participar, que no cumplieron criterio de exclusión, y que fueron entrevistados, examinados y culminaron a cabalidad tres pruebas espirométricas consecutivas.

El proceso de selección aseguró, mediante la referencia escrita de los participantes o cuidadores, mediante el interrogatorio personal y mediante el examen físico, a unos individuos "sanos", no fumadores,

EDITORES INVITADOS

asintomáticos y sin antecedentes patológicos respiratorios, cardiovasculares o relacionados con enfermedades neuromusculares o deformidades de la caja torácica.

De esta manera la *población blanco*², (niños y jóvenes "sanos" de Valencia) se vio representada y restringida finalmente por una serie de 537 participantes.

Los estudios observacionales¹ como este, son proclives a dos tipos de errores, que atentan contra su validez interna y externa, los cuales se verifican en la selección de los individuos (hecho preponderante en el estudio) y en la medición de las variables. El sesgo de selección se refiere a una distorsión en la estimación de un fenómeno resultante de la manera como los sujetos participantes son seleccionados.

En el abordaje inicial, los autores no hacen manifiesto el marco muestral, esto es, la densidad poblacional finita de individuos del área urbana de Valencia de la que se haría el muestreo (*población de referencia*)², así como tampoco la de individuos a los cuales se van a hacer las inferencias clínicas (*población blanco*)².

Los problemas continúan cuando un estudio comunitario como este, no utiliza un sistema aleatorio para obtener su muestra (muestreo probabilístico) sino que lo hace por conveniencia de manera no probabilística.

Si uno está interesado en medir la frecuencia de un fenómeno en la población, pues sencillamente mide toda la población y establece el parámetro poblacional de ese fenómeno. Cuando quiera que existen poblaciones muy grandes para los propósitos de medición y en aras de hacer eficiente esta fase del proceso de investigación, se han diseñado estrategias de Muestreo (aleatorio y no aleatorio) que permiten entonces obtener estimadores de dicho parámetro.

Una muestra aleatoria simple es una selección de algunos miembros de la población de tal manera que cada individuo es independientemente seleccionado y tiene una igual probabilidad conocida (no cero) de ser escogido³.

Un muestreo poblacional de tipo aleatorio, garantiza, por efectos de azar, que aquellos individuos seleccionados son **representativos** de la población

mayor, que no fue medida, pero a la cual queremos extrapolar o generalizar los resultados o los valores obtenidos.

El estudio de Rodríguez y colaboradores "seleccionó" seis escuelas sin que se hiciera explícito por qué motivo fueron seis, así como tampoco por qué específicamente esas escuelas (y no otras) y qué tipo de características basales tenían. Como se anotó, la muestra fue supraselectiva de quienes aceptaron participar, cumplieron los criterios y finalizaron las pruebas (con una tasa de pérdida de 49% de la población contactada inicialmente). De esta manera, obviamente, no conformaron la muestra, niños valencianos que no estudiaban en esas seis escuelas, no colaboradores en estudios y encuestas (que generalmente tienen un perfil y un pronóstico diferentes) y que fueron incapaces de hacer las pruebas y por lo tanto, existió un grupo significativo de niños de la población blanco no representado adecuadamente por la muestra y a quien naturalmente no sería posible (aunque deseable) generalizar los resultados de este estudio.

De esta manera, la ausencia de un muestreo probabilístico y la presencia de una muestra de conveniencia, determinaron pérdida de validez interna y dificultaron de manera protuberante la posibilidad de generalización de los datos a poblaciones mayores (de la misma Valencia, de Venezuela y de Latinoamérica). El trabajo no recoge los "valores espirométricos normales para niños valencianos" sino los valores encontrados en un segmento selectivo de la población pediátrica valenciana.

En lo que respecta al proceso de medición es de resaltar los siguientes aspectos:

No aparecen registradas (ni tampoco tenidas en cuenta en el modelo) las características basales de los participantes en cuanto hace referencia a su nivel de escolaridad, estado nutricional, grado de ejercitación, complejidad física, etnia, raza, nivel socioeconómico, altura sobre el nivel del mar, contaminación ambiental, etc, y esto tiene como consecuencia no sólo la falta adecuada de control de otras variables de importancia pronóstica en la espirometría (como las precipitadas), sino que impide que el lector estime adecuadamente si la población de estudio es clínica, biológica o socio-demográficamente similar a la propia pero en otros escenarios.

² La población blanco es el grupo de individuos, con interés restringido clínico-demográfico, del que uno ha muestreado (no necesariamente de una manera representativa) y alrededor del cual uno desea hacer inferencias con respecto de la población de estudio².

La falta en considerar ciertas variables de importancia predictiva en un modelo multivariable, tiene implicaciones interesantes. Nosotros podemos perder precisión en las estimaciones de los modelos si introducimos en él un número indiscriminado de variables que necesitan ser consideradas por razones de validez, pero ganar precisión y perder validez al omitir esas variables². De esta manera hay siempre una transacción entre pérdida de validez cuando se controla por pocas variables y pérdida de precisión cuando se controla por muchas. En la práctica primero la validez y luego la precisión (la validez no debe ser sacrificada en función de la precisión).

Ahora bien, los autores tan solo contemplaron la talla y la edad en sus reglas de predicción, con el objeto de tener estimadores precisos, pero dejaron por fuera variables importantes que hubiesen dado a dicho modelo más validez. A pesar de no contar sino con estas dos variables y de tener un importante tamaño de muestra (537 individuos) la variación de los estimadores en la descripción univariable de los datos (establecida al realizar coeficientes de variación) fue muy importante: FVC mujeres: 30%, FVC varones: 39%, FEV1 Mujeres 30%, FEV1 varones: 39%, FEF 25%-75% mujeres: 33%, FEF 25%-75% varones: 41%, PEFr mujeres 30%, PEFr varones: 41%.

Entonces la falta de precisión se suma a los estimadores sesgados de este estudio.

Es muy llamativo cómo los estimadores de la función pulmonar tiene en todos los casos menor variabilidad en las mujeres. Esto quiere decir tan sólo que las pruebas espirométricas son más confiables o consistentes en ese género.

Las variables utilizadas y registradas no fueron definidas operativamente, ni tampoco los procesos de medición así como el grado de precisión de ellas (especialmente en lo que hace referencia a la edad y la talla, las dos variables predictivas finalmente modeladas). En otras palabras, no se sabe si la edad fue preguntada o chequeada del registro de nacimiento?, fue consignada en años cumplidos?, meses cumplidos? O, años, meses y días cumplidos?. El lector entenderá que sin adecuadas definiciones operativos un niño de 7 años, 11 meses y 29 días, podría ser registrado como de 7 años por unos, o como de 8 años por otros y no reflejarse la **verdadera** edad de este niño.

Fueron excluidos individuos agudamente enfermos, con antecedentes de enfermedades cardiovasculares

o respiratorias, con hábito de fumar (de pacientes o padres?), extremadamente delgados o extremadamente gordos y se presumió (erróneamente) que todo individuo que finalmente culminó la investigación era "sano" o "libre de enfermedad" sólo por el hecho de ser clínicamente asintomático y no poderle demostrar anomalías patológicas al examen físico y en el interrogatorio. No se debe perder de vista que las enfermedades o condiciones patológicas tiene una fase presintomática o preclínica (en ocasiones de años) que sigue a aquellas de causa suficiente para la generación de enfermedad 4,5.

Así las cosas, los investigadores perdieron una valiosa oportunidad de comparar entre grupos, determinar cómo son los valores espirométricos en las diferentes categorías poblacionales y clínicas, establecer diferencia clínica y estadísticamente significativas entre ellas y más allá, haber evaluado la utilidad de esta como una prueba diagnóstica o de tamizaje.

Llegando al punto medular de la discusión, cómo podemos interpretar y aplicar los resultados de este trabajo?

La palabra "normal" en medicina tiene varias acepciones⁶. Una de ellas, la más socorrida, y mal interpretada, es aquella en la que después de medir una variable en una población de tamaño significativo, los valores de dicha variable siguen una distribución acampanada que ha sido llamada "*Normal*" o *Gaussiana*. Dos desviaciones estándar arriba y abajo del promedio en dicha distribución contiene bajo su área el 95% de los valores, o sea que 5% de ellos quedan por fuera de dicha distribución. En otras palabras en toda muestra poblacional el 5% de los valores (y por supuesto los pacientes que dieron origen a dichos valores) son "anormales estadísticos". Ese 5% es tan sólo infrecuente, no enfermo. No se debe significar que dado que los valores de los individuos se puedan ubicar dentro de lo "normal" estén libres de enfermedad y por otro lado aquellos con valores "anormales" si lo estén. Lógicamente, es más probable que estén enfermos (o libres de una determinada condición) aquellos que ostenten valores lejanos (o cercanos respectivamente) de dichos valores "normales". A esta acepción se le refiere como *normalidad estadística*.

La homologación de criterio de normalidad estadístico como "sano o libre de enfermedad" tiene implicaciones serias en medicina. La primera es que por definición el 5% de la población estaría enferma, o

sea con una prevalencia del 5% para todas las patologías por igual (hecho no cierto). La segunda es que es posible que individuos verdaderamente enfermos ostenten valores “normales” y sean rotulados falsamente como “sanos”. En resumen la utilización de la normalidad estadística como criterio de enfermedad asegura una tasa fija de enfermedad en 5% para todas las enfermedades (con muchos falsos positivos) y un número significativo de escapes diagnósticos (falsos negativos)

A pesar de ello el establecimiento de pesos y tallas “normales”, valores normales densitometría ósea, valores normales en los ángulos acetabulares y valores normales espirométricos se siguen empleando, entre otros, para definir desnutridos de nutridos, osteoporóticos de no osteoporóticos, displásicos de no displásicos acetabulares, y enfermos respiratorios de sanos, como en el caso que nos ocupa, y plantear propuestas terapéuticos con base en ellos.

Haciendo un breve paréntesis, si bien existe evidencia de que la mayoría de las variables biológicas siguen una distribución no normal ⁷ (sesgada o asimétrica), tal como lo corroboran los autores en la distribución encontrada en sus valores espirométricos, la normalización estadística obtenida mediante el logaritmo de las variables a la que procedieron, no controla biológicamente y clínicamente el hecho que siguen siendo no normales, sino sólo que esto permitió la introducción de dichas variables en el modelo, sin violar las premisas necesarias para tales efectos en los *métodos multivariantes del modelo de predicción* ⁸. Por lo tanto, aún así normalizadas, la utilización de parámetros con Desviación Estándar de la medida, para diferenciar lo “normal” de lo “anormal”, de una variable **no normalmente distribuida** no hace cierta la descripción de su distribución real.

Las pruebas de función pulmonar, son esencialmente pruebas diagnósticas, o sea que intentan clasificar a un individuo dentro de categorías: Disfunción respiratorias (patrón restrictivo, obstructivo, etc) o sin Alteraciones.

Sin embargo es probable, que los parámetros poblacionales **reales** de ambas distribuciones se dispongan una a continuación de la otra a lo largo de los posibles valores, SIN UN CLARO INTERVALO o sin una clara brecha entre ellas, esto es, una proporción, a veces significativa, de individuos “sanos” puede sobreponerse o traslaparse en sus valores sobre la de los individuos “disfuncionales” o viceversa. Figura 1.

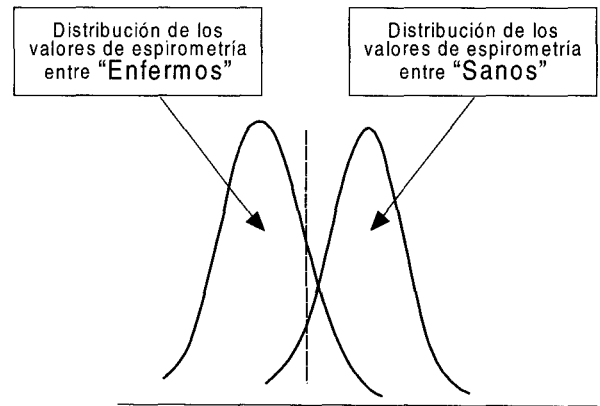


Figura 1. Posibles distribuciones reales (hipotéticas) de los valores de espirometría en individuos con enfermedad respiratoria y sin enfermedad respiratoria. La línea interrumpida señala el punto de corte en los valores espirométricos, que diferencian salud de enfermedad.

Ya que la mayoría de las variables biológicas tienen esta tendencia, se han diseñado técnicas estadísticas para establecer cuál es el mejor punto de corte (línea interrumpida en la figura) que pueda definir, con el menor número de errores (falsos positivos y negativos) quien padece o no una condición. Cómo se llega a esa instancia?. Se toma un grupo de individuos el cual es sometido a un proceso de evaluación por un mismo estándar de referencia (estándar de oro) que lo diferencia en dos subgrupos, aquellos claramente enfermos y aquellos claramente no enfermos. Unos y otros de manera enmascarada son ahora examinados por la nueva prueba bajo estudios que intentan llegar a la misma distinción y clasificación de los individuos. En los análisis se estima cuán lejos está la prueba bajo estudio en sus diagnósticos con respecto a los establecidos por el estándar de referencia (esa distancia es la sensibilidad y especificidad). En el caso que la prueba bajo estudio se exprese en términos continuos (como son, por ejemplo, los valores espirométricos) una técnica, la *Curva de Características Operativas del Receptor* (en inglés *ROC Curve*), define el punto en que la nueva prueba comete menos errores y de qué magnitud son dichos ellos.

La verdadera utilidad de conocer los valores normales de una variable (prueba), es que ello ayude a discriminar válida y precisamente a este grupo de los no normales o enfermos y el trabajo de Rodríguez y colaboradores carece de elementos metodológicos y de diseño para poder llegar a ese punto. No tuvieron un grupo control (que hubiese podido ser niños claramente enfermos) y de un patrón de referencia (posiblemente la evolución y evaluación de pacientes e

individuos sanos a lo largo del tiempo) de manera igual para ambos grupos. Así las cosas el trabajo se quedó a medio camino. No sabríamos cómo utilizar los resultados en la definición de enfermedad en individuos de 6 a 18 años en la práctica clínica cotidiana.

Para que se tenga una idea de las dificultades de aplicación de este trabajo, es muy probable que el espectro de individuos medidos, con sus errores de medición, se ubique de una forma no suavizada, sesgada y sobrepuesta sobre las dos curvas poblacionales como se muestra en la figura 2.

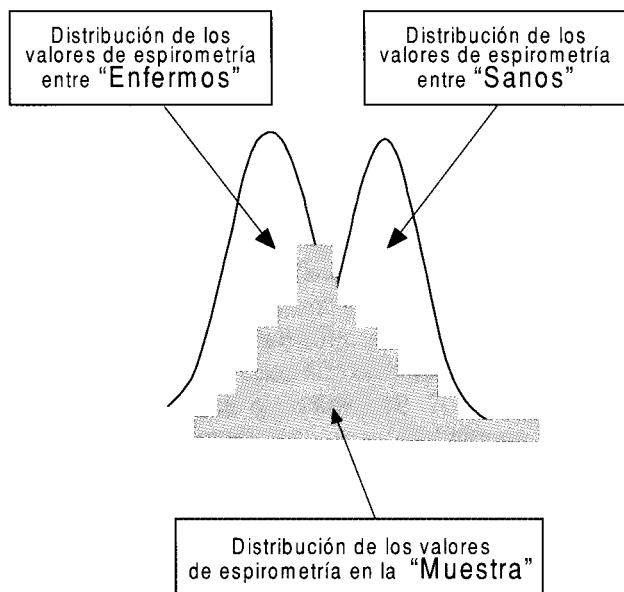


Figura 2. Posibles distribuciones de los valores de espirometría de la muestra (en gris) en relación con las posibles distribuciones de los valores espirométricos en verdaderamente enfermos y sanos.

Hay sin embargo un atenuante y un gran obstáculo en éste tipo de tópicos que exploran condiciones funcionales, dinámicas o en movimiento. La inexistencia de un estándar de oro más allá de la propia espirometría para definir la verdadera situación funcional y por ende las dificultades en los procesos de validación de estas técnicas o pruebas para su uso clínico.

En resumen, ciertos sesgos en la selección de los individuos (sobre todo la ausencia de un muestreo aleatorio), la falta de consideración de ciertas variables

importantes, posibles errores en la medición, interpretaciones no exactas de una distribución normal, hace que las reglas predictivas obtenidas por este meritorio trabajo, se vean atentadas en su validez y precisión y por ende asaltan dudas acerca de la posibilidad de generalizar los resultados incluso a otros sectores pediátricos de la población valenciana o venezolana.

*Jairo Echeverry Raad M.D. MSc (c). U.J.N. Corpas
Profesor asociado departamento de Pediatría e Instituto
de Investigaciones Clínicas Facultad de Medicina
Universidad Nacional de Colombia*

BIBLIOGRAFÍA

1. Feinstein AR. The architecture of cross-sectional research (conclusión) Clin Pharm & Therapeut 1978; 23:481-93
2. Kleinbaum DG, Kupper LL, Morgenster H. Validity of Epidemiologic research. In: Kleinbaum DG, Kupper LL, Morgenstern H. Eds. Epidemiologic Research: Principles and Quantitative Methods. Van Nostrand Reinhold 1982; New York. Chap: 10.
3. Rosner B. Estomation. In: Rosner B. Ed. Fundamental of Biostatistics. Duxbury Press Belmont 1990; Ca: Chap:6.
4. Rothman KJ, Greenland S. Causation and Causal Inference. In: Rothman KJ, Greenland S. Eds. Modern Philadelphia; Epidemiology. Lippincott-Raven 1998; Chap:2.
5. Echeverry J. Tuercas y tornillos de la Medicina Basada en Evidencias: El modelo epidemiológico moderno. En: Echeverry J. Ed. Haciendo un diagnóstico clínico desde la perspectiva de la Medicina Basada en Evidencias. Editorial el Manual Moderno (en prensa) Bogotá; 2003; Capítulo: 1
6. Echeverry J. Ardila E. Pruebas Diagnosticas y Procesos Diagnósticos. En: Ardila E. Sánchez R, Echeverry J. Editores. Estrategias de Investigación en Medicina Clínica. Editorial El Manual Moderno; Bogotá; 2000; Capítulo: 13.
7. Micceri T. The unicorn, the normal curve and other improbable creatures. Psychol Bul 1989; 105: 156-66.
8. Kleibaum DG, Kupper LL, Muller KE. Straight-Line Regression Analysis. In: Kleibaum DG, Kupper LL, Muller KE. Eds. Applied Regression Analysis and Other Multivariable Methods. PWS-Kent Publishing Company. Boston; 1988; Chap:5.
9. Echeverry J. Cómo estimar el rendimiento de una nueva prueba diagnóstica?. En: Echeverry J. Haciendo un diagnóstico clínico desde la perspectiva de la Medicina basada en evidencia. In: Capítulo 5. Editorial El Manual Moderno (en prensa). Bogotá. 2003.