

Evaluación de la información científica

Herschel S. Horowitz

Presentación

*Considerando la importancia de saber interpretar apropiadamente los reportes de investigación, el Comité Editorial se permite presentar a continuación la traducción y adaptación del Capítulo 10 del libro *Dental Practice and the Community*, de D. F. Striffler, Editor, publicado en su tercera edición en 1983 por W. B. Saunders, Filadelfia, al que considera de gran utilidad por los conceptos emitidos.*

Introducción

Los estudiantes, los clínicos y los salubristas deben leer reportes científicos continuamente para desarrollar tanto el conocimiento fundamental en sus áreas de trabajo como para mantenerse al día. Es pues muy importante que sean capaces de evaluar la calidad y validez de esos reportes que pueden afectar sus actitudes, acciones, o su práctica. El contenido de este artículo pretende ayudar a evaluar críticamente la información científica, y mostrar que un artículo no es necesariamente válido sólo porque fue publicado.

Con frecuencia se citan y se usan como argumento reportes de investigaciones de calidad cuestionable. A veces es difícil detectar las investigaciones científicamente pobres. Sin embargo, los lectores serán confundidos con menos

frecuencia si buscan errores graves de diseño, muestras inapropiadas, el uso inadecuado de métodos estadísticos, fallas en la ejecución de detalles de procedimiento, inexactitudes en los datos reportados, conclusiones no sustentadas, y citas inapropiadas o fuera de contexto. Además, la habilidad para diferenciar las investigaciones válidas le ayuda a estudiantes y clínicos a mejorar la calidad de sus propios reportes.

Este artículo no estudia fórmulas estadísticas o procedimientos para hacer pruebas estadísticas. Esa clase de información puede encontrarse en los textos de estadística. El propósito es más bien el de examinar algunos factores que ayudan al lector crítico a interpretar, entender, y evaluar la literatura científica. Los ejemplos de estadística citados en este artículo fueron escogidos para ilustrar las bases del razonamiento estadístico.

Muchos lectores pasan por dos etapas al evaluar el contenido de los reportes científicos. Inicialmente tienden a aceptar la veracidad de los reportes incuestionablemente. Posteriormente, después de fallar algunas veces, se hacen progresivamente más escépticos. Los lectores deben cuidarse sin embargo de no ser excesivamente críticos, o podrían permanecer desinformados respecto de los adelantos científicos. El propósito de desarrollar la habilidad para

evaluar es el de evitar extremos tales como la aceptación o la desconfianza ciegas respecto de los reportes científicos.

Tipos de información científica

Una característica de los estudios que contrasta con un editorial o el reporte de un caso es la de llegar a conclusiones sin contaminación por sesgos. El sesgo es una preferencia inconsciente o una inclinación que inhibe el juicio imparcial. La diferencia entre sesgo y prejuicio (un prejuicio u opinión preconcebida sin conocimiento o examen de los hechos) yace en lo inconsciente del sesgo. En el contexto de la conducción de una investigación, un sesgo es cualquier proceso en cualquier etapa de un estudio que tiende a producir resultados o conclusiones que difieren sistemáticamente de la verdad. Los sesgos suelen provenir de diseños inadecuados del estudio, de errores sistemáticos pero honestos de observación, optimismo infundado, o evidencia insuficiente. Los buenos estudios tienden a remover estas y otras fuentes de sesgos.

En la investigación terapéutica o en la preventiva la prueba de la efectividad clínica debe basarse en estudios controlados en humanos, o en estudios que comparen uno o más regímenes con algún otro procedimiento bien establecido, o con la no intervención. Las anécdotas y los reportes de casos no constituyen investigación, aunque pueden ofrecer claves para responder a problemas de investigación. Muchas formas de tratamiento y prevención populares pero inefectivas han sido adoptadas con entusiasmo en el pasado basándose en una suposición no verificada de su valor. De otro lado, hay casos en que la curiosidad o las impresiones de un clínico han llevado a la adquisición de importantes conocimientos. Piénsese sólo en la curiosidad de McKay acerca de la pigmentación marrón del Colorado, que llevó a la investigación posterior acerca de fluorosis, caries dental y fluorización comunitaria, y se apreciará la importancia potencial de los reportes de observaciones personales. Debe enfatizarse, sin embargo, que los reportes de casos no son investigación.

El término estadística tiene dos significados cuando se habla de investigación en salud. Las

siguientes definiciones son representativas de estos dos contextos. Usado en plural, estadísticas —como las que muestran la incidencia de una enfermedad— denota el grupo de figuras que describen un fenómeno. Cuando se usa como sustantivo singular, estadística significa el cuerpo de métodos o técnicas usado para tomar decisiones sobre la relación, estabilidad o competencia de la información recopilada.

Un estudio clínico es un proyecto de investigación prospectivo que involucra sujetos humanos. Está diseñado para evaluar qué tan bien trabaja un régimen particular bajo condiciones dadas, y para permitir un juicio científicamente sano del valor de la adopción (o de la continuación del uso) de una práctica biomédica aceptada.

Componentes de un reporte científico

Al leer un reporte científico debe evaluarse qué tan bien fueron preparadas las diferentes secciones que lo componen, comenzando con el título.

Título

El título de un artículo debe ser claro y conciso, aunque describiendo el contenido del reporte de la manera más completa posible. El título no es sólo importante por su capacidad para llamar la atención del lector dejándolo saber si le interesará el contenido del artículo, sino porque permite que las diferentes bibliotecas y servicios de archivo, o el propio lector clasifiquen el documento en categorías adecuadas que faciliten la recuperación del documento. Es por eso que la inclusión de palabras claves "key words" es importante.

Resumen

Un buen resumen de un reporte científico permite identificar rápidamente el contenido básico del artículo, y así valorar su importancia para los intereses particulares del lector. El resumen determina con frecuencia si un artículo es leído o descartado. Debe, por tanto, revisarse con cuidado. Un resumen debe 1) establecer los principales objetivos y el alcance de la investigación, 2) describir brevemente los métodos usados, 3) resumir los hallazgos, y 4) listar las

principales conclusiones. No debe contener información o conclusiones no mencionadas en el cuerpo del artículo. Debe ser directo, objetivo y veraz, sin campo para opiniones o para la especulación.

Introducción

La introducción o sección de antecedentes de un reporte debe señalar claramente las razones para adelantar el estudio o la revisión. El lector deberá preguntarse si es necesaria esa investigación en particular; esto es ¿Tiene la investigación importancia para las necesidades de la sociedad en general o para las de la comunidad odontológica en particular?

Revisión de la literatura

El artículo deberá revisar en forma completa pero concisa la literatura pertinente que resume la información disponible sobre el problema investigado. Los autores del reporte deben revisar independientemente la información disponible, y no depender o modificar la revisión de otro reporte. El estado actual de la investigación pertinente debe también ser resumido. Algunos reportes se refieren indiscriminada e inapropiadamente al trabajo de otros investigadores. Los escritores de buenos reportes científicos no deben depender únicamente de las afirmaciones de la literatura que sostienen un punto de vista particular, ignorando las afirmaciones que arrojen dudas sobre ese punto de vista. Tampoco incluyen afirmaciones o conclusiones fuera de contexto para dar la impresión de que los reportes originales decían algo diferente de su intención original. Los artículos citados deben ser pertinentes al reporte y no deben usarse para acumular nombres respetables en la bibliografía o para mostrar qué tanto ha leído el autor.

La aproximación efectiva y crítica a un reporte depende del entendimiento inicial del lector del objetivo de la investigación. Es por eso que los objetivos deben aparecer explícitamente. Deben ser específicos, y medibles. Si la investigación es en principio descriptiva, así debe constar. Si es probatoria, las hipótesis a probar deben estar enunciadas. De ser posible, deben ser presentadas a modo de hipótesis nulas, es

decir que se pretende que los procedimientos utilizados en el estudio no afecten la naturaleza de las observaciones, o los resultados obtenidos. El objetivo de un estudio puede ser refutar una hipótesis nula y entonces proponer una hipótesis alterna que sea la base para adelantar otro estudio. El lector deberá sin embargo estar alerta respecto de investigaciones que buscan probar hipótesis alternas sin razones adecuadas para rechazar la hipótesis nula.

Materiales y métodos

Los materiales y métodos usados en un estudio deben presentarse en detalle. El tratamiento dado a cada grupo deberá describirse claramente. La descripción de los criterios utilizados para determinar si una condición estaba presente o ausente deben ser específicos. Las variables medidas deben delinearse cuidadosamente. Las mediciones deberán ser, en lo posible, directas, y tan libres como sea posible su interpretación subjetiva. Deben darse suficientes detalles como para que un trabajador competente pueda repetir la investigación.

Existen esencialmente cuatro tipos de escala de medidas que pueden aplicarse en la investigación clínica: nominal, ordinal, de intervalo, y de relación. El uso del tipo apropiado de escala es crítico para la recolección y análisis de datos, por lo que se discuten a continuación algunos aspectos importantes de cada tipo.

Medidas nominales: son aquellas que agrupan a los sujetos en categorías separadas, mutuamente excluyentes, de tal manera que todos los sujetos en cada categoría son equivalentes respecto de algún atributo particular. Un ejemplo de medida nominal es la agrupación de hombres y mujeres en categorías diferentes, y la asignación de los números 1 y 2 a cada grupo, respectivamente. La asignación de niños a las diferentes categorías de la clasificación de Angle para la maloclusión es otro ejemplo de medición nominal. Como la medición nominal consta de nombres y números asignados a categorías, el adelantar pruebas matemáticas sobre esa medición es obviamente inapropiado, aunque puede ser útil determinar estadísticamente si las frecuencias observadas en cada clase difieren de la frecuencia esperada.

Medidas ordinales: las escalas en las que el aumento o disminución del número asignado implica un aumento o disminución de un atributo o propiedad son llamadas medidas ordinales. Aunque los valores pueden estar escalafonados, los intervalos entre categorías no son necesariamente iguales. La escala de dureza de los minerales es una escala ordinal. Un ejemplo odontológico podría ser la categorización de la higiene oral en excelente, buena, regular, o pobre, con la asignación de los números 4, 3, 2 o 1 a esas categorías, respectivamente. Es inapropiado afirmar, por ejemplo, que las bocas con higiene oral excelente son cuatro veces más limpias que las de la categoría pobre. Sólo se sabe que si los sujetos son clasificados apropiadamente, las personas con un grado excelente (4) tienen bocas más limpias que aquellos con un grado bueno (3), y así sucesivamente. Las poblaciones estudiadas pueden compararse con base en la distribución por categorías ordinales. En muchos reportes, sin embargo, los datos ordinales son sumados, restados, multiplicados o comparados de cualquier otra manera. Nada prohíbe el que esto se haga. Sin embargo, esta manipulación puede confundir acerca de la magnitud de la propiedad en cuestión. El lector de reportes en que se ejecutan ejercicios matemáticos sobre medidas ordinales deberá preguntarse: ¿Tiene algún significado real el resultado de estas medidas? ¿Importa que un grupo tenga un promedio inicial de 3.4 y después de alguna intervención ese promedio haya disminuido a 2.9? Cuando los valores proceden de medidas ordinales, el cambio numérico observado puede no ser importante.

Medidas de intervalo: en ellas las diferencias entre la cantidad de una característica puede ser distinguida, pero además, la diferencia entre categorías es igual en magnitud. La temperatura es un ejemplo de medida de intervalo. El intervalo entre 60 y 70 grados Fahrenheit es el mismo que hay entre 30°F y 40°F. El punto cero de una escala de intervalo es arbitrario. Cero grados Fahrenheit no es un valor absoluto en el que no hay temperatura; fue escogido arbitrariamente. El número de años es también una medida de intervalo. La cantidad de tiempo transcurrido entre 1776 y 1800 es la misma que pasó entre

1920 y 1924. La sobremordida horizontal (Overjet) es un ejemplo odontológico de medida de intervalo, con una relación borde a borde designada arbitrariamente como sobremordida horizontal O.

Escala de relación: solamente difieren de las medidas de intervalo en que el punto cero no es arbitrario sino que indica la ausencia de la propiedad medida. La estatura y el peso son ejemplos de escalas de relación. Un valor de peso de cero significa que no hay peso en lo absoluto, y una persona que pesa 150 kg pesa el doble de otra que pesa 75 kg. El índice CPO –caries, perdidos y operados– es un ejemplo odontológico de escala de relación.

Otro punto de interés para los lectores de reportes científicos es si el aspecto escogido por los investigadores para la evaluación es apropiado para probar la hipótesis planteada. El propósito final de un estudio clínico es el de tratar o prevenir la enfermedad en la población evaluada. Un estudio bien diseñado debe mostrar, por ejemplo, qué tanta caries dental fue prevenida mediante cierto régimen, o el grado al cual redujo la gingivitis cierta campaña educacional. Los estudios que solamente miden la concentración de flúor del esmalte externo después de la fluorización, o que determinan si ha aumentado el conocimiento de los sujetos sobre la salud oral, o si los niveles de placa han disminuido no responden a la pregunta última acerca de la reducción en la incidencia, prevalencia e intensidad de la enfermedad. Las variables indirectas pueden o no estar altamente relacionadas con la efectividad terapéutica, pero aun si lo estuvieran, no testifican la reducción de la enfermedad.

Selección de la muestra

La población a estudiar que se reporta en una investigación científica debe ser claramente definida. Si el grupo estudiado es una muestra de una población mayor, debe describirse brevemente esa población, así como los procedimientos utilizados para escoger la muestra.

El lector de literatura científica debe preguntarse si la muestra escogida para el estudio es representativa de un grupo mayor del que los datos fueron tomados. Si no es así, los autores no deben generalizar o extrapolar sus hallazgos

concluyendo que, por ejemplo, su tratamiento exitoso será también efectivo entre otros grupos. Si un estudio muestra que cierta dieta reduce efectivamente la incidencia de caries dental en niños institucionalizados cuyo estilo de vida es cuidadosamente controlado, ese resultado puede no ser cierto para niños no institucionalizados. Un método de higiene oral probado en estudiantes de odontología que seguramente ya tienen un grado relativamente alto de salud gingival, así como destrezas y conocimientos superiores a los de la población promedio, puede producir resultados totalmente diferentes en personas de la misma edad, pero de la población general. Si el autor ha ignorado el efecto de la naturaleza de la muestra sobre los resultados, el lector deberá decidir si los hallazgos son generalizables a otras poblaciones.

El lector debe también fijarse en cómo fueron seleccionados los sujetos del estudio. En caso de que fueran excluidos algunos participantes potenciales, las razones para la exclusión deben ser válidas. Ejemplos de razones válidas pueden ser: ausencia de suficientes dientes sanos evaluables, presencia de bandas ortodónticas, o de enfermedad periodontal. En algunos casos las exclusiones constituyen sesgos que prohíben la generalización de los hallazgos a otras poblaciones.

Hay muchas maneras de sesgar los resultados de estudios al seleccionar la población: por ejemplo, los sujetos de un estudio sobre enfermedad periodontal pueden ser pacientes que hayan buscado tratamiento para la enfermedad. Pueden haber tenido síntomas, o conocer personas que los sufran. Quizás se presten para el estudio por incentivos económicos, o por la oportunidad de una evaluación clínica. A veces los sujetos son reclutados mediante cuestionarios, sean directos o enviados por el correo, o por llamados a través de los medios masivos de comunicación. Los investigadores pueden tender a excluir cuestionarios confusos o incompletos, o rechazar participantes potenciales poco atractivos o de personalidad desagradable. Todas estas autoselecciones de los sujetos, así como las exclusiones de los investigadores pueden introducir sesgos en la muestra. Es obvio que la selección puede influenciar los resultados.

Asignación de sujetos a grupos de estudios

Siempre que sea posible los sujetos deben ser asignados por azar a los grupos de estudio. Una asignación fortuita o sistemática puede producir grupos con diferentes características que afecten la variable estudiada. La asignación por azar de los participantes y el fiel seguimiento del método son signos importantes de qué tanta credibilidad puede tener un estudio.

El lector debe tratar de establecer si la distribución de los sujetos es imparcial y equitativa.

La principal función de distribuir los sujetos en grupo experimental y grupo control es la de conformar grupos que dentro de lo posible sean idénticos en todos los aspectos salvo la condición evaluada. Si los grupos son realmente comparables al comenzar el estudio, cualquier diferencia observada en la variable estudiada al final del estudio puede atribuirse al tratamiento dado a cada grupo, una vez excluidas las diferencias debidas al azar.

Idealmente todo participante debe tener las mismas posibilidades de ser escogido para el grupo experimental o para el grupo control. El método de asignación no debe seguir sesgos sistemáticos que tiendan a favorecer a uno de los grupos al iniciar la investigación. La asignación por azar simple de los sujetos en grupos es una de las mejores formas de conformar grupos comparables respecto de factores conocidos y desconocidos que pudieran influenciar el resultado de la investigación clínica. Si en un estudio sobre prevención de caries hay factores que puedan influir el comportamiento futuro de la enfermedad —raza, sexo o experiencia cariosa previa, por ejemplo— los investigadores pueden utilizar la estratificación o conformación de bloques basados en esos factores en combinación con la asignación al azar: antes de la asignación por azar a los grupos, los participantes con características comunes son agrupados en bloques que determinan estratos. Por ejemplo, un grupo podría incluir hombres de diez a doce años de edad que tienen un índice CPO —Caries, perdidos, operados— de cuatro a siete. Dentro de cada bloque los sujetos son distribuidos por azar al grupo experimental o al de control. La conformación de bloques permite que una población heterogénea sea agrupada en estratos, cada uno

de ellos homogéneo dentro de sus límites. La creación de bloques basados en las variables apropiadas seguida de la distribución al azar mejora las posibilidades de que los grupos sean comparables al comienzo del estudio, lo que es menos probable que ocurra si se utiliza sólo la distribución al azar. Ningún método garantiza que todos los grupos son exactamente iguales en todos los aspectos. Los métodos de asignación, sin embargo, incrementan la posibilidad de que sólo existan pequeñas diferencias entre grupos, usualmente detectables y medibles.

Los investigadores deberán evitar la asignación de sujetos por cualquier mecanismo que impida la utilización del azar. El uso de un grupo control escogido por sistemas diferentes al azar requiere la suposición de que el grupo experimental y el grupo control son idénticos en todos los aspectos importantes salvo la variable analizada, o que los investigadores pueden compensar toda diferencia relevante. Esa suposición es muy raramente cierta. Un ejemplo extremo enfatiza este punto: para evaluar la efectividad de un dentífrico con fluoruro en la prevención de las caries es obvio que el dentífrico fluorado no debe asignarse a un grupo de estudiantes de odontología y el dentífrico control a obreros de la misma edad. Otros sesgos pueden ser más sutiles, pero igual deben ser detectados y neutralizados. Algunos ejemplos de asignación de grupo que pueden causar serios problemas para la interpretación de los resultados y que deben evitarse cuando sea posible son:

1. Por razones altruistas, asignar todos los sujetos con caries rampante al grupo experimental, suponiendo que se beneficiarán más.
2. Asignar el grupo control a los sujetos cuyo consentimiento no se pudo conseguir, y colocar en el grupo experimental a los que aceptaron participar.
3. Usar los sujetos de una institución como control para compararlos con los de otra.
4. En un estudio de enfermedad periodontal, usar pacientes de un periodoncista como sujetos de un régimen de tratamiento y los de otro periodoncista para recibir otra terapia a comparar.
5. Tratar de hacer grupos comparables colocando conscientemente sujetos en un grupo, ba-

sándose en factores que supuestamente influencian la futura enfermedad o el estado de salud.

6. Permitir que los sujetos escojan el grupo al que pertenecerán.

Estos y otros métodos cuestionables de asignación a grupos pueden influenciar el desarrollo del estudio.

El autor debe reportar las razones por las que potenciales candidatos rehusaron participar, y el número de rechazos. Esta información puede ofrecer claves para la racionalización de los procedimientos del estudio y quizás ayude a definir qué tanto pueden extrapolarse los resultados a la población general.

Características de la muestra

La edad de la población a estudiar debe ser apropiada para el tipo de estudio. Por ejemplo, sería inútil tratar de medir el efecto de un régimen anticaries en los dientes primarios de una población entre ocho y once años. A esa edad muchos de los dientes deciduos ya han desaparecido de la boca, y en una población susceptible la mayoría de las superficies temporales restantes ya estarán cariadas. Así mismo, la selección de una población de ocho años para una evaluación de retención de sellantes no es apropiada. A los ocho años muchas de las superficies de fosas y fisuras de los molares permanentes ya están cariadas, mientras que los premolares y segundos molares permanentes no han erupcionado. La determinación del número de dientes cariados, obturados y perdidos en un grupo de sujetos entre 60 y 80 años de edad puede prestarse a malas interpretaciones ya que seguramente muchos de los pacientes habrán perdido dientes por razones periodontales. Un estudio de las relaciones oclusales que utiliza pacientes con dentición mixta es otro posible ejemplo de selección inadecuada de edad de la muestra, así como un estudio de pérdida de inserción en adultos jóvenes. Estos ejemplos enfatizan que el lector debe preguntarse a sí mismo si los hallazgos pueden aceptarse como válidos o estaban comprometidos de antemano por la selección de una muestra inapropiada para el propósito de la investigación.

En estudios que comparen una población con otra las poblaciones deben ser realmente compa-

rables respecto de las características esenciales relacionadas con la variable a estudiar. Un ejemplo ilustra este punto: a finales de los años setenta algunos opositores a la fluorización comunitaria afirmaron que la tasa de mortalidad por cáncer en diez ciudades de los Estados Unidos con agua fluorizada estaba aumentando comparándola con la de diez ciudades no fluorizadas. Sus conclusiones se basaban en las tasas de mortalidad crudas, producto de dividir el número total de muertes entre la población total. No puede sacarse tal conclusión de las tasas crudas de mortalidad porque las muertes por cáncer varían ampliamente con la edad, y en menor grado con el sexo, grupo étnico y la raza. Cuando los datos fueron analizados independientemente corrigiendo las diferencias de edad, sexo y raza entre comunidades, la supuesta relación desapareció completamente. En otras palabras, las diferencias demográficas entre los dos grupos de poblaciones eran enteramente responsables de la diferencia en la tasa de mortalidad cruda por cáncer. No podía distinguirse el efecto del flúor. El lector de artículos científicos debe vigilar cuidadosamente los reportes buscando posibles sesgos que resulten de datos no ajustados o estandarizados.

Tamaño de la muestra

El tamaño de la muestra es otro factor importante a revisar. El número de personas necesarias para un estudio clínico puede determinarse con anticipación si la magnitud de la nueva enfermedad y la variación entre sujetos puede estimarse con un estudio piloto, o con datos similares reportados en la literatura. Cuando se esperan grandes diferencias entre el grupo control y experimental, las muestras pueden ser más pequeñas que cuando se esperan diferencias pequeñas. La selección de muestras de tamaño adecuado es esencial si se quieren hacer afirmaciones acerca del valor de un régimen al final del estudio. Las muestras deben ser lo suficientemente grandes como para que las diferencias entre grupos que se consideren clínicamente importantes sean estadísticamente significativas a un nivel del 5%, o lo que es lo mismo, al 0.05 de nivel de probabilidad ($p < 0.05$). Esto significa simplemente que el riesgo de aceptar una diferencia entre grupos cuando ésta en realidad

no existe es solamente del 5%. En ese caso, el 5% de probabilidades de aceptar falsamente un régimen como efectivo es llamado error de tipo I. Se le llama α a la probabilidad de cometer un error de tipo I. Para ofrecer una mayor seguridad de que la diferencia medida es real los investigadores pueden fijar el nivel α al 1%, o $p = 0.01$.

Si la muestra es demasiado pequeña es posible que se concluya erróneamente que el tratamiento no es efectivo, cuando realmente lo es. Este último tipo de error es conocido como error de tipo II, y a la probabilidad de cometerlo se le expresa con la letra griega β . Los investigadores solían cuidarse sólo de los errores de tipo I. Recientemente, sin embargo, han comenzado a cuidarse de los errores de tipo II, principalmente porque hay preocupación por muchos estudios que dieron resultados negativos en poblaciones demasiado pequeñas como para detectar un resultado positivo si existiera. La selección de un nivel α bajo no protege a los investigadores de la posibilidad de fallar en la identificación de un agente o un régimen efectivo. No debe descartarse un régimen efectivo sólo porque la posibilidad de detectar su efecto es pequeña. La capacidad para detectar un efecto específico verdadero, si existe es llamada poder de una prueba, y se expresa como $1 - \beta$. En esencia, si los resultados de un estudio son negativos (es decir, si la hipótesis nula es aceptada), el error β , y en consecuencia el poder de la prueba son los indicadores de qué tan confiable es el resultado negativo obtenido. El propósito de considerar el poder cuando se calcula el tamaño de la muestra es mejorar las posibilidades de que se encuentren efectos reales e importantes, si es que existen.

El poder de una prueba se hace especialmente importante en estudios con diferentes grupos, en los que más de un tratamiento efectivo es evaluado. Un estudio así puede comparar diferentes concentraciones y frecuencias de uso de enjuagues de fluoruro entre grupos cuya diferencia en el aumento de caries tiende a ser pequeña. Si los investigadores juzgan que diferencias tan pequeñas como del 15% entre grupos son clínicamente significativas y desean demostrar que diferencias de ese tamaño son estadísti-

camente significativas, el número de sujetos por grupo deberá ser extremadamente grande. Si los grupos son pequeños, puede concluirse falsamente que los regímenes son equivalentes porque no mostraron diferencias estadísticamente significativas, cuando el hecho es, que las posibilidades de demostrar que eran diferentes (el poder de la prueba) eran pequeñas —menores, por ejemplo, del 50%—.

La discusión anterior no quiere decir que la mejor forma de superar los problemas de tamaño de muestra es hacer grupos de estudio tan grandes como sea posible en todas las investigaciones. *Aun si el tamaño de la muestra es adecuado, una diferencia trivial entre dos grupos estudiados puede ser estadísticamente significativa pero clínicamente poco importante.* Muy a menudo se reportan estudios en los que se resaltan diferencias estadísticamente significativas, pero poco importantes clínicamente, debido al uso de una muestra excesivamente grande. Es importante evitar la conclusión de que existe un efecto terapéutico porque se encontró una diferencia estadísticamente significativa. El tamaño de muestra de cada estudio debe determinarse individualmente, dependiendo del número de grupos, la naturaleza y aceptabilidad de los regímenes usados en cada uno de ellos, la variación de respuesta esperada entre participantes, y la magnitud anticipada de la diferencia de comportamiento entre grupos. Como se mencionó previamente, un estadístico debe participar en la determinación del tamaño de muestra.

En estudios de múltiples grupos, los investigadores deberán reportar el poder ($1-\beta$) de la investigación. Deben describir la prueba estadística usada y afirmar, por ejemplo, que hay un 80% de probabilidad ($1-\beta$) de rechazar correctamente la hipótesis de nulidad con un nivel de significancia estadística del 5% (α), si existía realmente una diferencia del 15% entre grupos.

Grupos control y diseños alternos

Un componente esencial de los estudios clínicos es el grupo control, una colección de sujetos que provee respuestas con las que se puede comparar el efecto del régimen evaluado. El uso de grupos control es básico en la investigación científica. Los controles pueden ser sujetos que

no reciben tratamiento, que reciben un placebo (control negativo), sujetos que reciben un tratamiento reconocido contra el que se compara una nueva terapia, o que reciben el mismo tratamiento, pero con una dosis o frecuencia diferente (control positivo). Por ejemplo, si se va a comparar un nuevo enjuague bucal para la prevención de las caries con una solución de sal de cocina, el grupo de sujetos que reciben la solución de sal constituyen el grupo de control negativo. Si se va a comparar el nuevo enjuague con una solución de fluoruro de sodio 0.2%, el grupo que usa esta última sería el control positivo. Algunos estudios se diseñan con controles positivos y negativos.

Los estudios con controles negativos son más útiles y apropiados cuando el valor de el nuevo régimen es desconocido o incierto. La comparación con un control negativo es la única forma de establecer la magnitud real de la protección ofrecida por el nuevo régimen. En los últimos años los controles negativos han sido menos utilizados por consideraciones éticas sobre negar el acceso de algunos participantes a una terapia benéfica. Sin embargo, sin un control negativo poco puede hacerse por determinar la magnitud real del beneficio de algún tratamiento estudiado. Aun después de terminar el estudio pueden existir dudas acerca de la efectividad de los tratamientos adelantados.

Un diseño utilizado con frecuencia últimamente para superar las dificultades de negar el tratamiento a una parte de la muestra es el establecimiento de un estado inicial de condiciones existentes. El estado inicial sirve como base para comparaciones futuras. Se comienzan los regímenes de tratamiento, y los resultados periódicos se comparan con el hallazgo inicial. Este diseño retrospectivo no utiliza controles no tratados, y una de sus debilidades radica en los cambios naturales que pueden producirse con el tiempo en la condición investigada, y que no pueden medirse. Además, con este diseño pueden presentarse variaciones no detectadas en la aplicación de los criterios diagnósticos en los diferentes exámenes. Los lectores de un reporte de este tipo deben ser conscientes de sus debilidades, ya que no permiten sacar conclusiones de causa y efecto. Las ventajas éticas de estos dise-

ños pueden, sin embargo, contrarrestar sus inconvenientes. Es indudablemente preferible obtener controles de reportes de literatura. De hecho, el valor de muchas de las drogas más aceptadas actualmente, incluyendo la morfina, digital, aspirina, insulina, penicilina y corticosteroides fue establecido en estudios sin comparación concurrente. La mayor parte de los estudios sobre efecto cariostático del agua fluorizada han usado el diseño retrospectivo. Estos estudios han mostrado invariablemente que el beneficio aumenta con el tiempo de consumo. Aunque otros factores en la comunidad pueden haber afectado la salud dental durante las evaluaciones, la influencia del consumo de flúor ha sido tan profunda y los hallazgos tan consistentemente repetidos, que la conclusión del beneficio producido es inescapable.

Observaciones y tratamientos

Un reporte debe especificar qué enfermedad está estudiando y mostrar los criterios diagnósticos específicos que se usaron para determinar si la enfermedad estaba presente o no. Afirmar, por ejemplo, que un estudio involucraba pacientes con enfermedad periodontal es inadecuado, pues la composición de ese grupo de pacientes está abierta a todo un espectro de especulaciones entre los investigadores. La confiabilidad de las medidas y del diagnóstico es indispensable para la exactitud de los resultados.

Un buen índice debe ser específico, es decir que no debe medir un atributo en demasiadas personas que no tengan ese atributo. Además debe ser sensible, para que no pase por alto el atributo en los sujetos que lo tengan. Si los autores usan un índice bien establecido en el gremio, como el índice CPO para medir caries dental, o el índice de Russell de enfermedad periodontal, tal y como fueron descritos, o de acuerdo con modificaciones disponibles en las memorias de reuniones científicas de importancia, no es necesario repetir los detalles. Por ejemplo, si los criterios de caries dental son los mismos establecidos en 1968 por la Conferencia sobre Evaluaciones Clínicas de Agentes Cariostáticos, organizada por la Asociación Dental Americana, una afirmación de ese tipo puede ser suficiente. Si se realizaron modificaciones al

método o a los criterios, o si se ha aplicado un nuevo índice, los detalles son esenciales. Si se ha usado un nuevo índice, el lector crítico deberá preguntarse si el índice es válido, es decir, si mide lo que pretende medir. Deben presentarse los resultados de estudios piloto usando el nuevo índice para probar su validez.

Los reportes deben contener una descripción clara de los cuestionarios aplicados y de los tratamientos empleados y cómo fueron supervisados. El lector no deberá asumir qué le hicieron, a quién, cuándo, dónde y cómo.

La recolección de datos deberá ser hecha por tan pocos examinadores como sea posible para reducir las variaciones interexaminador. Los autores deberán aclarar qué tan de acuerdo estaban los diferentes examinadores al recoger la información, así como qué tan consistentemente repetía cada examinador la misma medida en diferentes momentos. Dicho de otra forma, un reporte completo debe incluir la variación interexaminador y la variación intraexaminador, para apreciar la confiabilidad y reproducibilidad del estudio. No es suficiente la simple afirmación de que estas mediciones se hicieron. Debe incluirse por lo menos un resumen de los resultados de las pruebas. El lector deberá sospechar de los resultados que carecen de medidas de reproducibilidad y confiabilidad, especialmente cuando el estudio contiene elementos subjetivos en el diagnóstico, en la observación, o en la determinación del pronóstico.

El lector crítico debe establecer si los sujetos fueron examinados al azar o no. No deberán examinarse primero todos los sujetos de un grupo y luego los de otro. Al intercalar sujetos de los diferentes grupos se evita la tendencia inconsciente a desviar la aplicación de los criterios diagnósticos, pudiéndose afectar los resultados de algún grupo en particular.

Técnicas ciegas

Las técnicas ciegas son otra forma importante de reducir los sesgos en los estudios clínicos. Al igual que las de selección de muestra, las técnicas para conseguir evaluaciones ciegas ayudan al lector a determinar la objetividad de un estudio. El protocolo debe asegurar que los examinadores tomen los datos sin saber a qué grupo

pertenece el sujeto (examen ciego). Siempre que sea posible, un estudio debe diseñarse con la modalidad doble ciego, consiste en que no sólo los examinadores desconocen la asignación a grupos de los sujetos, sino que los propios sujetos tampoco la saben. Un diseño en el que los sujetos no saben qué tratamiento reciben neutraliza cualquier efecto placebo que un nuevo tratamiento pueda tener. Este diseño es a menudo importante porque muchos sujetos pueden sentirse mejor cuando reciben un determinado tratamiento. Por supuesto, la ignorancia de los sujetos respecto del tratamiento recibido no es posible si los regímenes son obviamente diferentes, o si se usa un grupo no tratado como control. Sin embargo, cuando se usa un placebo como control, éste deberá ser tan parecido como sea posible al tratamiento experimental en términos de sabor, olor, color y apariencia física. Idealmente la única variación deberá ser la ausencia del ingrediente activo a evaluar. Uno de los elementos críticos en el diseño de un estudio es el de que los evaluadores no se den cuenta a qué grupo corresponden los sujetos, para prevenir que cualquier prejuicio o sesgo modifique los resultados. Otra precaución que debe tomarse siempre es que no estén disponibles los resultados de exámenes anteriores al momento del examen actual, para que los resultados previos no afecten los diagnósticos actuales.

El uso de técnicas ciegas no corrige, sin embargo, los errores en la calibración de los examinadores o en la selección del índice o de los criterios apropiados. Si los índices tienen elementos no especificados, si las escalas de medida no son claras, o si no se usaron criterios precisos para los índices y escalas, los hallazgos serán sospechosos aunque se use la técnica doble ciego. El lector de reportes no debe aceptar resultados irreproducibles sólo porque fueron recolectados con una técnica ciega.

El lector deberá también revisar las posibles maneras como se compromete la técnica ciega. Por ejemplo, si un agente tiñe los tejidos orales, como en el caso de la clorhexidina o las concentraciones altas de fluoruro de estaño, una técnica realmente ciega puede ser imposible.

Si se están registrando diferentes condiciones, por ejemplo el estado de los tejidos periodonta-

les, la presencia de placa y la cantidad de caries, sería ideal que un examinador diferente aplicara cada índice en vez de permitir que un solo examinador haga múltiples medidas. Cuando se usa un examinador diferente en cada índice es menos probable que el resultado de una medida influya sobre los resultados de las demás. Por ejemplo, como la mayoría de los investigadores asumirían que la gingivitis se relaciona con la presencia de placa bacteriana, pueden tener la tendencia a observar más enfermedad gingival en sujetos con mala higiene que en sujetos con buena higiene oral.

El lector debe evaluar otros aspectos de la técnica ciega. Si el artículo reporta un estudio en el que las medidas preventivas fueron aplicadas por profesionales, debe especificar si los profesionales sabían o no cuál era el agente terapéutico, y cuál el placebo. Ese conocimiento puede haber afectado inconscientemente el cuidado puesto durante la realización del procedimiento. La misma observación es aplicable a los supervisores de tratamientos autoaplicados. Siempre que sea posible, el personal de supervisión debe desconocer el contenido de los agentes. Todos los agentes utilizados en el estudio deben, por tanto tener etiquetas codificadas. Los estudios en que ni los examinadores, ni los participantes, ni el personal que aplica la terapia conoce qué agente se usa, se conocen como estudios triple ciego. Cuando las condiciones lo permiten, es el método de elección.

Diseños cruzados

El diseño cruzado es usado ocasionalmente en estudios en los que los sujetos conforman su propio grupo control. Cada sujeto recibe un tratamiento activo durante algún tiempo, y un placebo, o la ausencia de tratamiento durante el período control. Los diseños cruzados no son aplicables a estudios de prevención de caries porque el período de tiempo involucrado en el desarrollo de las lesiones es demasiado largo. Sin embargo, estos diseños pueden ser útiles en estudios sobre el efecto de la acumulación de placa o sobre la salud gingival por parte de enjuagues. La principal ventaja del diseño cruzado está en que evita diferencias en el pronóstico y variaciones en la respuesta entre los participan-

tes. El lector debe asumir que en estudios de diseño cruzado las características de los sujetos permanecieron iguales durante todo el estudio. Este diseño no es apropiado para tratamientos con efectos duraderos, pues la fase de tratamiento podría influenciar la respuesta durante la fase control. Si existe la posibilidad de un efecto terapéutico prolongado, los investigadores deberán incluir un intervalo de tiempo suficiente entre fases para que desaparezca ese efecto. Este tipo de diseño no es apropiado en caso que la terapia pueda producir efectos permanentes, o curación, o para agentes que permanezcan por períodos prolongados en el lugar de acción, ni cuando las condiciones puedan sufrir cambios naturales rápidos de prevalencia, incidencia o morbilidad.

El estudio clínico deberá continuar por suficiente tiempo para comparar exactamente la efectividad de dos o más tratamientos. A veces los primeros hallazgos resultan equivocados, o el efecto inicial no es sostenido. La publicación de resultados tempranos de un estudio puede llevar a una cadena de resultados errados.

Presentación de los datos

Los resultados de una investigación deben presentarse de manera clara, objetiva y con suficiente detalle para que el lector pueda examinarlos independientemente. Los datos reportados deben permitir un escrutinio cuidadoso. Las tablas y gráficas deben ser fáciles de interpretar y contener suficiente información para entenderse

por sí mismas, aunque no tanta como para que confundan al lector. Los datos deberán ser representativos más que exhaustivos o repetitivos. Los hallazgos importantes requieren documentación apropiada. No es suficiente, por ejemplo, afirmar que las mujeres le temían más que los hombres a un régimen. Deberán mostrarse las estadísticas que permitan a los lectores juzgar la validez de esa afirmación.

Las medidas de tendencia central, como promedios, medias y modos, deben darse, junto con las medidas de dispersión. La desviación estándar y el rango son medidas de dispersión usadas corrientemente. Las columnas y renglones de las tablas deben totalizarse. Los porcentajes de categorías mutuamente excluyentes deben sumar 100%, y los valores de las tablas y figuras deben corresponder con los del texto. Las tablas y gráficas estarán tituladas de manera completa pero concisa. El lector deberá ser capaz de entender la información contenida en ellas sin recurrir al texto.

Los datos de cada tabla deben organizarse para facilitar las comparaciones y mostrar las relaciones y tendencias. El título de cada columna debe describir exactamente la información contenida en esa columna, así como el de cada renglón debe indicar la naturaleza de los datos de esa línea horizontal. La Tabla uno es un ejemplo simple que muestra los hallazgos de un estudio sobre el efecto preventivo de un enjuague de fluoruro.

Grupo de estudio	No. de niños	Promedio de incremento del CPO	Diferencia (Porcentaje)	p*
Después de 12 meses				
Grupo control	210	1.46 (0.27)**		
Grupo experimental	198	1.05 (0.24)	28.1	0.12
Después de 24 meses				
Grupo control	210	2.92 (0.40)		
Grupo experimental	198	1.65 (0.28)	43.5	0.01

* Prueba t de dos colas ** Las cifras entre paréntesis son el error estándar de los promedios

Tabla 1. Incremento promedio del índice CPO-superficie después de doce y veinticuatro meses de enjuagues con fluoruro en niños de sexto grado que asistieron a todos los exámenes, según el grupo al que se asignaron. Portland, Oregon, 1971.

La representación gráfica es especialmente útil en el caso de complicados juegos de datos, o para resaltar relaciones interesantes entre resultados. La ilustración uno muestra la representación gráfica de los valores promedio de gingivitis medidos en niños y niñas que participaron en un estudio de tres años sobre la supervisión de la remoción diaria de placa dental en un colegio. El texto acompañante no debe simplemente repetir la información de las tablas o gráficas, sino que debe resumir el material y destacar tendencias interesantes o relaciones entre los resultados.

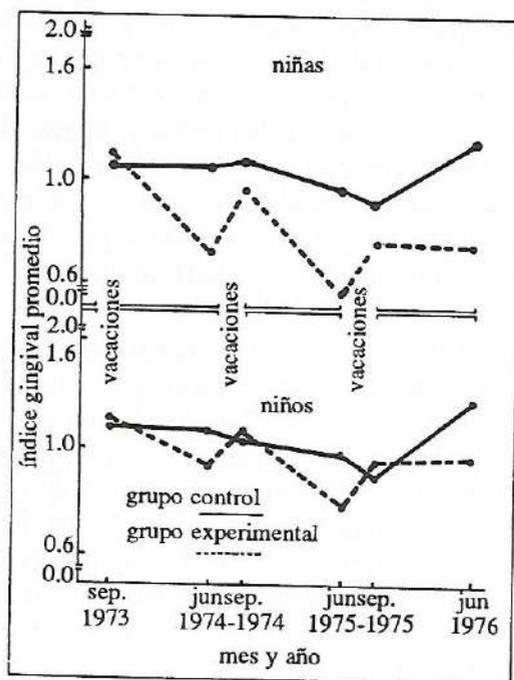


Ilustración 1. Índice promedio de gingivitis en cada examen de niños y niñas. East Hampton, Connecticut, septiembre de 1973 a junio de 1976.

En un reporte científico deben explicarse las razones para la pérdida de sujetos en el transcurso de la investigación. Si la atrición de los sujetos de un grupo es mayor que la de otros, las razones deben ser discutidas. Los datos deben demostrar la comparabilidad de los grupos, tanto en el caso de sujetos que comenzaron la investigación, como en el de aquellos que participaron de ella más tarde. Por ejemplo, debe ser mostrado el índice CPO-superficie de todos los niños que inician un estudio (para indicar la comparabilidad inicial), así como el de aquellos que terminen el estudio a los tres años (para in-

dicar si los sujetos que permanecieron eran similares a los participantes iniciales).

Cuando los datos han sido recogidos durante un período largo los lectores deberán verificar si hay lógica o consistencia en los hallazgos. Así, como las caries de los dientes permanentes tienden a aumentar su prevalencia linealmente entre los 6 y los 17 años, el incremento anual del índice CPO-superficie reportado en estudios de tres o más años de duración deberá aumentar linealmente. Las desviaciones grandes respecto de ese patrón podrían deberse al inadecuado entrenamiento de los examinadores o a cambios en los criterios diagnósticos.

Análisis de los datos

La respuesta a los problemas científicos viene de preguntas cuidadosamente planteadas y de datos bien recogidos, no de análisis esotéricos o abundantes. El análisis de los datos no es un proceso para generar la información, sino para ayudar en la interpretación y análisis de los resultados. Los métodos de análisis estadístico son simples herramientas del estudio científico, como un microscopio o una sonda para medir la concentración de fluoruro en una solución. El análisis estadístico aplica las matemáticas de probabilidades para mostrar con diferentes grados de seguridad la posible ocurrencia de un evento o efecto observado. Para aprender si un efecto observado es real, debe determinarse su variabilidad, es decir cómo fluctúa dentro de la muestra. Si el efecto claramente excede la fluctuación, es posiblemente real, o estadísticamente significativo. Si el efecto es relativamente pequeño comparado con la fluctuación o variabilidad entre observaciones, quizás no sea real o quizás haya ocurrido por azar con un grado de probabilidad mayor de lo aceptable.

El lector debe preguntarse si el diseño del estudio o el método de observación permite el análisis estadístico de los resultados. Un defecto grande de diseño o un sesgo importante del método de observación puede impedir un análisis válido de los resultados. Cualquier juego de números puede ser sometido a análisis estadístico, pero la cuestión es, si se justifica hacer cálculos necesarios para establecer inferencias.

Limitaciones de espacio impiden una discusión exhaustiva de los métodos estadísticos inapropiados. Los siguientes son, sin embargo, algunos errores comunes:

1. Aplicación de pruebas de significancia estadística a datos sesgados.

2. Realización de ciertas pruebas estadísticas a datos ordinales.

3. Realización de *pruebas t* y otras pruebas paramétricas a datos dimensionales con distribuciones excesivamente amplias (no normales) o a pequeños grupos con varianzas diferentes.

4. Uso de series de *pruebas t* en múltiples comparaciones sin considerar el hecho de que la posibilidad de encontrar diferencias estadísticamente significativas aumenta al aumentar el número de pruebas aplicadas.

5. Concluir que no hay efecto, o que no hay significación estadística a partir de una muestra inapropiadamente pequeña, sin determinar el error de tipo II.

6. Considerar importante una conclusión realmente trivial sólo porque una prueba la muestra altamente significativa.

7. Concluir que el valor de una variable medida es causa de la modificación de otra variable cuando una prueba produce un coeficiente de correlación (r) estadísticamente significativo. La correlación de dos variables puede ser grande y significativa, pero no necesariamente una relación de causa a efecto.

En algunos artículos publicados parece haber una relación inversa entre la cantidad y complejidad de los análisis y la validez de la investigación. Los datos que no muestran resultados significativos al ser sometidos a pruebas simples pero sí cuando se usan análisis refinados o complejos deben examinarse críticamente. Es posible mostrar significancia estadística como resultado de analizar datos transformados, o de procesar otra información mediante métodos complejos, pero la significancia puede reflejar diferencias sin importancia práctica, si por ejemplo, una evaluación no muestra beneficio al utilizar un enjuague anticaries, no debe hacerse el subanálisis de los datos hasta un punto tal que se concluya que es efectivo en las superficies distales de los segundos premolares superiores y

en mesial de los incisivos laterales superiores, pero no en las demás superficies. Conclusiones como esa no tienen valor práctico. Si se hacen suficientes análisis, alguno producirá hallazgos significativos por simple azar.

La costumbre de presentar los resultados sólo como porcentaje de reducción puede también llevar a conclusiones erróneas. En la interpretación de valores proporcionales de caries es esencial poner atención tanto al numerador como al denominador de los valores. Por ejemplo, si el incremento promedio de nuevas superficies cariadas en el grupo control de un estudio es de 0.5 por niño, y si en el grupo control el valor correspondiente es de 0.25 por niño, esa diferencia corresponde a un 50% de reducción. Sin embargo, sólo se han ahorrado 0.25 superficies cariadas por niño. Al evaluar reducciones porcentuales debe observarse críticamente la diferencia absoluta involucrada.

Los lectores deben estar pendientes de estudios pobremente concebidos o descuidadamente conducidos que muestran resultados "estadísticamente significativos". La significancia estadística no dice nada respecto del significado biológico o clínico de la diferencia entre números. No debe pensarse que el valor o significado de un estudio es auténtico sólo porque los resultados son estadísticamente significativos.

En este sentido quizás lo más importante tanto para lectores como para investigadores es establecer la diferencia entre significancia estadística y significancia clínica. La primera es la expresión matemática del grado de confianza en que una diferencia observada entre grupos sea real. Es la confianza en que una diferencia cero no se produciría si el estudio se repitiera y de que la diferencia observada no es debida al azar. En contraste, la significancia clínica es un juicio hecho por el investigador o el lector sobre si las diferencias observadas son importantes para la salud. La significancia clínica es una determinación subjetiva basada en la experiencia clínica y en el entendimiento de las características de la condición medida. Es una estimación sobre el valor práctico de las diferencias entre el grupo control y el grupo experimental. Son muchas las condiciones que determinan los juicios sobre la significancia clínica. Una de las más comunes

es la relación costo-beneficio, es decir el costo relativo para conseguir una cierta protección o efecto. Por ejemplo, un régimen que cuesta 250 pesos por año (por niño) para reducir media superficie CPO se considera costo-efectivo, mientras que otro régimen que cuesta 20.000 pesos por año (por niño) para prevenir una superficie CPO no es costo-efectivo.

El uso apropiado de varias pruebas estadísticas como las pruebas t y las relaciones F requiere que los datos tengan una distribución normal. Muchos de los datos recopilados en la investigación biomédica no están distribuidos normalmente, y por tanto la presunción de normalidad es frecuentemente violada. Afortunadamente muchas pruebas paramétricas usadas para analizar datos odontológicos son lo suficientemente robustas como para ser aplicadas legítimamente cuando el número de observaciones es grande, aun cuando la distribución de los resultados no sea normal. Existen también pruebas paramétricas basadas en distribuciones diferentes de la normal, así como pruebas que no implican la suposición de ningún tipo especial de distribución de resultados. Son las pruebas libres de distribución o no paramétricas.

Los datos que no están distribuidos normalmente son ocasionalmente normalizados usando logaritmos, raíces cuadradas u otras transformaciones de los datos originales para legitimar el análisis estadístico. La lógica científica y la justificación de estas transformaciones no es clara. Por ejemplo, ¿qué significa realmente decir que la raíz cúbica de alguna variable medida es significativamente más baja en un grupo que en otro?

Muchos reportes fallan al reportar el método utilizado para analizar los datos. Al lector debe advertírsele específicamente qué prueba se usó, si era de una o dos colas, y cuál fue el valor de probabilidad estadística (p), en vez de sólo mostrarle una expresión de significancia estadística del tipo $p < 0.05$. Las pruebas estadísticas de dos colas se utilizan generalmente cuando no hay certeza respecto de la efectividad de un régimen sobre otro. Si existe información previa respecto de si un régimen es superior a otro, deberá usarse una *prueba de una cola* al evaluar la superioridad del primero de los regímenes. La

decisión de si se harán pruebas de una o de dos colas debe hacerse durante la fase de planeación del estudio y no después de la observación.

El *valor p* es la probabilidad de obtener sólo por azar una diferencia tan grande como la observada en un estudio si la hipótesis de nulidad de que no hay efecto resulta cierta. Resultados de investigaciones que arrojan un *valor p* de 0.05 o menos son generalmente considerados estadísticamente significativos. *Valores p* menores de 0.01 son altamente significativos. Estas demarcaciones son arbitrarias. El resultado del cálculo de p no debe determinar estrictamente la decisión sobre la validez de una terapia. Existe la tendencia infortunada a descartar o descontar los hallazgos de estudios bien diseñados científicamente importantes; si los hallazgos se acercan pero no consiguen determinado valor de significancia estadística.

Las decisiones biológicamente importantes son raramente tomadas con base en un solo estudio. La tendencia a dicotomizar los resultados de un estudio a un sí (estadísticamente significativo, o $p < 0.05$) o no (no estadísticamente significativo, o $p > 0.05$) puede ser contraproducente y llevar a errores de interpretación. *Valores p* de por ejemplo, 0.06 y 0.04 no deben llevar a una conclusión opuesta, sino a una muy parecida a la original. Además, el análisis del estudio no debe limitarse al enfoque estadístico. Un análisis más descriptivo o informativo de los datos cuantificados puede ser apropiado. Este tipo de análisis puede tomar la forma de intervalos de confianza. Aunque el intervalo de confianza calculado a partir de los datos de un estudio puede incluir una diferencia cero entre grupos, el intervalo puede caer bien dentro de un territorio que indique el efecto benéfico del tratamiento. Puede no ser correcto clasificar como negativos hallazgos que producen un intervalo de confianza que incluya al cero entre sus límites, así como dar por positivos los que muestran valores que indican un fuerte efecto del tratamiento. Las pruebas tradicionales de significancia estadística enfocan todo su interés en un punto preciso: el límite inferior del intervalo de confianza.

Las condiciones de un estudio puede correlacionarse. Esto significa que puede haber una

asociación causal, complementaria, paralela, o recíproca entre dos o más variables medidas. Un ejemplo es la relación positiva que ha sido repetidamente mostrada entre el índice de placa dental y la gingivitis. Un *coeficiente de correlación* es una medida de la interdependencia entre dos variables al azar, cuyo valor fluctúa entre -1 y +1, indicando la perfecta correlación negativa y la perfecta correlación positiva, respectivamente. El coeficiente de correlación se expresa con la letra *r*. La magnitud de variación en una variable explicable por otra se determina con el cuadrado de *r*, y se expresa r^2 . Así, un coeficiente de correlación de +0.30, que puede ser estadísticamente significativo, explica sólo el 9% de la variación entre dos variables; el 91% restante no queda explicado. Además, un estimado de la correlación no prueba que un factor sea la causa de otro. La relación entre factores puede ser casual o fortuita.

Los análisis de correlación múltiple, o análisis de regresión permiten el cómputo del efecto máximo total de la combinación de factores independientes sobre la variación de la variable dependiente. Estos análisis determinan la influencia de esos factores en orden decreciente de influencia sobre la variable dependiente. Un análisis de regresión puede mostrar, por ejemplo, que la enfermedad periodontal de una determinada población puede explicarse en orden decreciente de importancia por la edad, el índice de placa dental, las prácticas de higiene oral, el sexo, la educación, y los ingresos, y explicar en total el 38% de la variación de los resultados.

Discusión

El propósito principal de la discusión de los resultados es el de resaltar la relación entre los hallazgos observados. Un error común de muchos reportes es el que la discusión en su mayor parte repite los resultados. En vez de eso, la discusión debe presentar los elementos principales, las relaciones y las generalizaciones mostradas por los resultados. Debe también señalar cualquier excepción, inconsistencia o falta de correlación entre los hallazgos, y destacar los puntos no sustentados. Incluirá también el juicio crítico de los autores respecto de su propio trabajo. Esta parte de un reporte debe también mostrar cómo

los resultados de la investigación coinciden o contrastan con artículos previamente publicados, o con preceptos establecidos. La especulación sobre las implicaciones o la aplicabilidad de los resultados pertenece también a esta sección. Resumiendo, la discusión le da al escritor la oportunidad de expresar sus opiniones e ideas respecto del estudio.

Conclusiones

La mayor parte de los reportes debe contener una o varias conclusiones claramente expresadas, que se basan en los datos recolectados y en su análisis. Las conclusiones deben ser relevantes respecto de la pregunta original propuesta por los investigadores, y deben expresar claramente si cada hipótesis de nulidad fue aceptada o rechazada. Por supuesto, las conclusiones deben estar de acuerdo con los resultados del análisis estadístico. El lector debe preguntarse si las conclusiones generalizan más allá de los límites de la población estudiada. Las conclusiones deben tener sentido biológico, y ser de importancia práctica. Los autores deben señalar si sus hallazgos y la demás información existente sobre el tema sugieren la necesidad de investigación adicional. El lector deberá también preguntarse si las conclusiones lo llevarán a cambiar su forma de pensar, sus actitudes, su práctica o sus recomendaciones.

Referencias

Los reportes científicos, así como los de revisión deben contener referencias que apoyen las afirmaciones del texto. La mayor parte de las referencias documentan las afirmaciones hechas en la revisión de literatura, y las de la discusión de los resultados del estudio. Deben dirigir al lector a otras fuentes de información adicional. La citación de referencias deberá por tanto dar suficiente información para que cada fuente pueda ser localizada y verificada rápidamente. Se prefiere la citación de fuentes originales de información más que afirmaciones atribuidas a terceros, para reducir la posibilidad de reproducir errores de la documentación. Una *bibliografía* es una lista de trabajos de un autor específico, o en el contexto de un reporte científico, una lista de fuentes de información impresa sobre el tema del reporte. Algunas veces la biblio-

grafía al final del reporte se titula "Lecturas recomendadas". Los documentos de la bibliografía deben ser pertinentes al tema del artículo. No deben incluirse documentos irrelevantes escritos por investigadores bien conocidos. Toda la información pertinente debe ser citada en la bibliografía, y no solamente aquellos artículos que apoyan la premisa o los hallazgos del autor. El lector debe desconfiar de los reportes que se apoyan mucho en comunicaciones personales en la lista de referencias, así como de los artículos que citan desordenadamente el trabajo del propio autor, omitiendo referencias importantes de otros trabajos sobre el tema.

Consideraciones éticas en estudios con sujetos humanos

Durante los últimos años ha aumentado la preocupación por las consideraciones éticas en la conducción de estudios biomédicos con seres humanos. Aunque muchas consideraciones éticas de la investigación con humanos no pertenecen propiamente a un artículo sobre la evaluación de la literatura científica, otras sí. Por esta razón se discuten varios aspectos de la ética de la investigación biomédica.

En estudios éticos las preguntas hechas deberán responderse válidamente manteniendo los derechos y el bienestar de los participantes totalmente protegidos. No es ético adelantar un mal experimento científico en humanos. No se le debe pedir a los sujetos que asuman riesgos de posible daño, incomodidad o inconvenientes, si no se incluyen suficientes sujetos para que los resultados sean válidos. Tampoco utilizar un número innecesariamente grande de sujetos para probar un punto clínicamente importante, o usar diseños de asignación al azar, o de "doble ciego" inapropiados. Las deficiencias de diseño no pueden ser justificadas o corregidas posteriormente. Por definición, un diseño pobre no es ético, y no puede ejecutarse. El diseño de la investigación científica tiene profundas implicaciones éticas que requieren la atención del lector de reportes científicos.

Los estudios que utilizan seres humanos deben seguir principios científicos que dan racionalidad a la investigación. La justificación para la investigación dependerá de los resultados de

estudios de laboratorio, de otros estudios clínicos, o del profundo conocimiento de la literatura científica pertinente. Casi todas las instituciones de investigación han establecido comités permanentes para que revisen las investigaciones propuestas con humanos. Los comités deben ser competentes para revisar los estudios propuestos científicamente, y para determinar su aceptabilidad en términos de la ley local, normas de conducta profesional y práctica comunitaria.

Los derechos y el bienestar de los participantes deben ser adecuadamente protegidos. Cualquier riesgo predecible debe compararse con los beneficios que aparentemente se recibirán, o con la importancia para el público del conocimiento que se obtendrá. Los intereses de la ciencia y de la sociedad deben valorarse respecto de los intereses de los participantes. Por estas razones los reportes deben contener una breve descripción de cómo fueron reclutados los participantes, y cómo se les informó del estudio.

Consentimiento escrito

El consentimiento escrito de los pacientes o participantes debe obtenerse por medios adecuados y justos, después de que se ha explicado totalmente el estudio y en la medida de lo posible, cuando haya sido entendido por los candidatos. La participación debe ser realmente voluntaria y no presionada en forma alguna. Debe explicarse que los hallazgos individuales serán manejados confidencialmente. Los riesgos y beneficios razonables deben ser explicados y puestos en perspectiva y se debe autorizar a todos los sujetos para retirarse del estudio en cualquier momento, sin ningún perjuicio para ellos.

En la investigación de métodos preventivos es particularmente importante, cuando se usa un placebo, que los sujetos sepan de la posibilidad de recibir un tratamiento inefectivo. En investigaciones sociales o de comportamiento, la discreción respecto de las creencias, prácticas o hábitos de los participantes son especialmente importantes.

Las consideraciones éticas entran a menudo en conflicto con el rigor científico del diseño de los estudios clínicos. Cuando los sujetos son asignados a los grupos por azar, aquellos que pertenezcan al grupo control no recibirán trata-

miento efectivo o prevención, si la intervención es efectiva. Aun si no se usa grupo control tratado con placebo o no tratado, puede existir la duda de que el régimen probado sí sea efectivo. Si no se incluye un grupo control en el diseño del estudio en que se comparen dos regímenes, los sujetos en uno de los dos grupos corren el riesgo de recibir el régimen menos efectivo. Si los investigadores piensan que uno de los regímenes es más efectivo que los demás, éticamente todos los sujetos deberían ser ubicados en ese grupo. Sin embargo, a menos que los sujetos sean asignados por azar a los grupos, el estudio no arrojará resultados válidos desde el punto de vista científico. Dado el conflicto entre las demandas éticas y el rigor científico, está aumentando la posibilidad de que los estudios científicamente comprometidos lleven a resultados no concluyentes. El buen diseño experimental es elemento esencial de la investigación ética en humanos. No es ético colocar personas innecesariamente en riesgo, aunque éste sea pequeño, dentro de estudios que no serán capaces de dar resultados útiles.

Son tres las preguntas básicas que deberán hacerse cuando se está diseñando un estudio: ¿Deberá haber un grupo control negativo?, ¿deberán ser asignados los sujetos por azar?, ¿es éticamente aceptable asignar a los sujetos por azar dentro de los grupos? Desde el punto de vista científico la respuesta a la segunda y tercera pregunta es rutinariamente sí, pero personas de la comunidad con diferentes puntos de vista pueden dar diferentes respuestas a estas preguntas. Puede no haber respuestas correctas que sean aceptadas por todos. El lector de reportes científicos debe pensar en estas consideraciones cuando lee un reporte y hacer un juicio ético del diseño del estudio después de leerlo.

Revisiones

La literatura científica incluye artículos diferentes de los reportes de estudios clínicos. Tal es el caso de las revisiones de la información actualizada sobre temas particulares. La proliferación de información científica en los últimos diez a veinte años ha hecho que la mayor parte del personal de la salud estreche el rango de tópicos sobre los que puede mantenerse actualiza-

do. Pocos son capaces de leer y absorber el contenido de todos los reportes científicos en tantos temas. Las revisiones de temas particulares son usualmente escritas por expertos con conocimiento de primera mano respecto del tema, que han digerido y evaluado críticamente la información disponible, y la condensan en paquetes coherentes comprensibles para otros científicos o para los practicantes.

No debe subestimarse la importancia de los artículos de revisión para el avance de la ciencia y la práctica de la profesión. Existe una deficiencia crónica de gente capaz de escribir revisiones valiosas. Recoger la literatura disponible sobre cualquier tema y colocarla en un único trabajo conciso, relativamente libre de sesgos, e inteligible es un trabajo que requiere mucho tiempo, conocimiento y habilidad para escribir.

Recuperación de la información científica

La cantidad de información científica está creciendo de manera explosiva. Varias revistas que contienen información odontológica importante han comenzado su publicación en los últimos años. Enfrentados con el siempre creciente cuerpo de la literatura odontológica, los profesionales deben limitar su lectura en razón de limitaciones económicas o de tiempo, a una proporción cada vez más pequeña de las publicaciones disponibles.

Uno de los mayores problemas que se presentan con el creciente campo de la literatura es el almacenamiento y recuperación de la información. Técnicas nuevas y económicas para clasificar y recuperar la información por computador, como en el caso de los sistemas "Medline" y "Medlards" de la Biblioteca Nacional de Medicina de los Estados Unidos han simplificado estas tareas. El acceso al servicio "Medlards" es posible en nuestro medio a través de la Facultad de Medicina de la Universidad de Antioquia. Sin embargo, aún así, la persona que busca información a través de estos sistemas es abrumada por la gran cantidad de literatura marginal aparentemente importante que es recuperada automáticamente cuando se usa el servicio. Los sistemas de recuperación por computador tienen como limitación que su eficacia depende de la forma como se haya clasificado la información

dentro de ellos. Si un artículo ha sido codificado inapropiadamente en el sistema, no aparecerá cuando se le solicite por su palabra clave. Aunque parece deseable la eliminación de entradas cuestionables a estos sistemas de recuperación, esto impediría la posibilidad de encontrar un artículo no reconocido previamente, pero pertinente, que haya sido clasificado erróneamente.

El *Index Medicus* y el *Index to Dental Literature* son fuentes invaluable de referencias científicas. Ambas publicaciones se encuentran disponibles en la Universidad de Antioquia.

Tipos de revistas

Algunas revistas dedican su contenido casi exclusivamente a la investigación, dirigidas básicamente a investigadores, especialistas o personal con intereses muy específicos. Otras son publicadas para grandes grupos profesionales, sean o no de especialistas. Contienen artículos de investigación y de revisión, así como artículos de amplio interés para los profesionales de cada campo específico. Estas revistas generalmente revisan rigurosamente los artículos a publicar.

Otro tipo de publicaciones es difícil de categorizar, pero puede ser llamado comercial o de marca, como *Dental Survey*, *Dental Management* o *Dental Products*. Aunque estas publicaciones contienen artículos, rara vez son revisados por jueces, o reciben edición cuidadosa. Puede decirse que las investigaciones importantes no se publican en estas revistas, aunque ocasionalmente se encuentren artículos útiles.

Un artículo publicado puede influir la práctica dental, los programas comunitarios, y la investigación futura de los científicos. El comité editorial de una revista y sus jueces tienen una responsabilidad ética importante al buscar que los artículos publicados sean científicamente sanos.

Las revistas no deberían rechazar artículos válidos estadísticamente porque sus hallazgos sean negativos. Los hallazgos negativos parecerían poco excitantes, pero pueden ser tan importantes en la planeación de la salud como los que muestran resultados positivos. Investigadores y editores tienen la responsabilidad ética de publi-

car artículos con resultados negativos si se basan en investigación sana.

Para mantenerse informados del desarrollo científico, los profesionales deben concentrarse en las principales publicaciones de su campo de interés. Es muy posible que menos de ocho a seis revistas contengan los artículos más útiles de cada disciplina. La revisión periódica de ese bien escogido grupo de publicaciones es probablemente la mejor manera de mantener un nivel razonablemente actualizado de conocimiento del área. Los medios de recuperación ya citados pueden usarse con propósitos específicos. El lector puede complementar el grupo básico de publicaciones con artículos de revisión y las memorias de congresos y conferencias. Es también útil la revisión de compilaciones como el *Dental Abstracts* y los números especiales del *Journal of Dental Research*, que contienen resúmenes de artículos presentados en reuniones anuales de la Asociación Internacional de Investigación Dental, y sus divisiones componentes. El principal valor de los resúmenes de artículos ya publicados es el de informar a los lectores sobre los hallazgos globales, y ayudarlos a mantener actualizados respecto del desarrollo de la investigación. El lector que requiera detalles de un estudio deberá buscar la publicación original, especialmente si está escribiendo un reporte propio. No debe usarse como referencia un resumen si existe una publicación completa del mismo artículo. El resumen preparado por el investigador para presentaciones orales puede ser usado como cita mientras aparece el respectivo artículo. Si se solicita, los investigadores usualmente envían manuscritos completos de los artículos que presentan en congresos y reuniones. El manuscrito invariablemente contiene información no hallada en el resumen.

Muchos artículos rechazados en revistas exigentes serán eventualmente publicados en otras revistas, pero este hecho no justifica la relajación de las normas de los comités editoriales para mantener publicaciones de alta calidad científica. Los autores escogen la revista en que publican su trabajo según la importancia para su investigación y según la calidad de la revista. Estos son los mismos criterios que debe usar el lector de una revista en particular.

Evaluación de los autores de reportes científicos

La reputación científica del autor de un reporte es a menudo útil para predecir la calidad de un reporte antes de leerlo. El trabajo de autores con reputación de conocedores del tema, con aptitud hacia la investigación, y que han demostrado previamente su habilidad para conducir y reportar cuidadosa y exactamente las investigaciones, seguramente será válido. Por supuesto, el lector novato de literatura científica puede no estar familiarizado con la reputación de los investigadores, o un artículo puede ser el trabajo de un autor relativamente nuevo. Hay, sin embargo, claves para la aparente calidad de un reporte. La afiliación de los autores, el lugar donde trabajan, las fuentes que utilizan (bibliotecas, consultores y equipos necesarios), sus colegas y colaboradores, y la fuente de soporte económico para la investigación son todas guías para predecir la calidad de un artículo.

Estilo de la literatura científica

La mayor parte de la literatura científica es comunicada en escritos que describen fenómenos y conceptos. La forma en que estos fenómenos y conceptos son expresados en lenguaje es el estilo, que coloca las palabras adecuadas en el sitio correcto.

Los mejores escritos científicos son organizados, precisos, claros y breves. Infortunadamente muchos científicos concluyen que para ser tomados en serio deben escribir tan científicamente como sea posible, usando jergas en lugar de lenguaje llano. Los que defienden el uso de términos científicos afirman que son más breves y exactos que el lenguaje corriente. Algunos científicos pueden hacer inconscientemente que sus términos hagan parecer su campo como algo incomprendible para el lector, lo que los convierte en expertos en su tema. Quizás la siquiatria y la sociología han contribuido más que otras áreas a la creación de jergas, pues muchos de los términos, tomados por otros profesionales, o por el público, ya no tienen ese significado preciso. Considérense términos como ambivalente, fobia, subconsciente, depresión y área socioeconómicamente deprimida.

La ciencia se construye con ladrillos de hechos, y el estilo determina la arquitectura con la que los hechos se expresan. El propósito de un buen escrito científico es comunicar hechos científicos de manera honesta y clara. Los estudiantes deben comprender que no vale lo que dicen hasta que pueden decirlo en términos tales que otros pueden entenderlo. Deben aplicar esta norma a los reportes que leen, así como a su propio trabajo.