

CEREMONIAS ESTADÍSTICO-METODOLÓGICAS EN LA INVESTIGACIÓN MÉDICA CONTEMPORÁNEA



Instituto Nacional de Higiene Epidemiología y Microbiología (INHEM)
Ciudad de la Habana. 2007

CEREMONIAS ESTADÍSTICO-METODOLÓGICAS EN LA INVESTIGACIÓN MÉDICA CONTEMPORÁNEA

Estudio para aspirar al Título de Especialista en Primer Grado de Bioestadística

Autor: Pedro Luis Hernández González

Tutor: Luis Carlos Silva Ayçaguer

Todas las teorías son legítimas y ninguna tiene importancia.
Lo que importa es lo que se hace con ellas.

Jorge Luis Borges

AGRADECIMIENTOS

Varias han sido las personas que, de una manera u otra, ayudaron a llevar adelante esta investigación merecedora de profuso esfuerzo, exigencia y dedicación. Sin su auxilio mi denuedo por concluir la hubiera resultado en vano. No he querido dejar de ofrecerles, como una pequeña muestra de afecto y gratitud, mi más profundo y sentido agradecimiento.

Principalmente al doctor Luis Carlos Silva, por su inigualable desempeño como tutor que, en todo momento, sobrepasó los límites de esta tesis. Su actitud ante la vida y su consagración y seriedad en el trabajo me han servido como máximo ejemplo, digno de imitar. De él he obtenido un caudal de conocimientos valiosos, desprovistos de cualquier dogma e impregnados principalmente de lo que considero imprescindible en un investigador: la intuición y el razonamiento lógico. En verdad ha sido un gran honor trabajar junto a alguien de tan alta estatura intelectual y calidad humana.

A mis profesores de la especialidad. Todos han sabido transmitir, inmejorablemente, útiles enseñanzas en mi formación como bioestadístico.

A mis padres, porque todos mis asuntos son para ellos causa de permanente desvelo.

A mis abuelas Luisa y Teresa y mis tías Norma, Benita, Basilisa y Silvia por su cariño, interés y apoyo. A mis tíos Guillermo, Lucio y Ramón.

A Yeni, por su entrega y preocupación constantes.

A mis compañeros de cuarto: Alberto, Aldo, Alfredo, Michel, Guillermo, Bacallao, Miguel y Ernesto. Verdaderos amigos que han hecho más llevaderos todos estos años. De cada debate sostenido con ellos siempre surgieron muy buenas ideas para este estudio.

A mis compañeras de aula; Miday, Zeida y Yasenia. Por su solidaridad en todo momento.

A Elena y Pepe por su sincera amistad y ayuda incondicional. Sus aportes fueron de absoluto beneficio.

PALABRAS CLAVES

- Tamaño de muestra
- Factores de riesgo
- Regresión logística
- Regresión paso a paso (stepwise regresión)
- Ensayo clínico controlado
- Pruebas de significación estadística
- Pruebas de hipótesis
- Encuestas
- Fichas técnicas

RESUMEN

Paralelo al desarrollo de la estadística aplicada a la investigación médica, han ido apareciendo perniciosas desviaciones ajenas a la disciplina propiamente dicha: el uso ceremonial que se hace de algunos procedimientos estadístico-metodológicos. Definidos como técnicas y/o procedimientos estadísticos que se aplican de manera adocenada e inercial en aras de cumplimentar requerimientos puramente formales, tales rituales representan un serio peligro en la labor investigativa contemporánea por carecer de utilidad práctica real y, ocasionalmente, conducir a resultados contraproducentes. El presente estudio, de corte teórico-metodológico, se orientó a caracterizar algunas de las referidas ceremonias, describir su naturaleza espuria y poner de manifiesto sus eventuales efectos perjudiciales. Se trazó asimismo la encomienda de ilustrar, acudiendo a producciones reales realizadas en un período reciente, su presencia en la literatura especializada actual. Para cumplir con tal propósito se eligieron cuatro procedimientos que suelen ser manejados de esta manera ritual: la determinación del tamaño de muestra a través de fórmulas, la aplicación algorítmica de la regresión logística paso a paso para la identificación de factores de riesgo, el uso de pruebas de hipótesis para corroborar homogeneidad entre los grupos de tratamiento en ensayos clínicos controlados y el empleo formalista e informativo de fichas técnicas para dar a conocer los rasgos de las muestras empleadas para llevar adelante encuestas. Para cada uno de estos procesos se hizo una discusión teórica y se caracterizó su posible empleo ritual. En cada caso se eligieron cuatro ejemplos tomados de revistas y sitios Web asociados a la investigación de salud, los cuales se analizaron de manera exhaustiva. El trabajo, fruto de un intenso pesquizaje y un profundo examen de la literatura actual, demuestra de manera razonada e incontrovertible el riesgo inherente al uso acrítico de recursos que, empleados de manera racional, pudieran ser útiles.

ÍNDICE

INTRODUCCIÓN	1
OBJETIVOS	4
METODOLOGÍA	5
DESARROLLO	7
1. Procedimiento estadístico-metodológico: cálculo del tamaño de muestra	
1.1. Bases teóricas generales.....	7
1.2. Empleo ceremonial del procedimiento.....	8
1.3. Análisis de su improcedencia.....	9
1.4 Ejemplos del empleo ceremonial.....	12
• Ejemplo 1.4.a.	12
• Ejemplo 1.4.b.	14
• Ejemplo 1.4.c.	15
• Ejemplo 1.4.d.	16
1.5 Sobre la frecuencia con que se emplea el ritual.....	18
2. Procedimiento estadístico-metodológico: uso de la regresión logística para identificar factores de riesgo.	
2.1 Bases teóricas generales.....	18
2.2 Empleo ceremonial del procedimiento.....	19
2.3 Análisis de su improcedencia.....	20
2.4 Ejemplos del empleo ceremonial.....	22
• Ejemplo 2.4.a.	22
• Ejemplo 2.4.b.	24
• Ejemplo 2.4.c.	27
• Ejemplo 2.4.d.	29
3. Procedimiento estadístico-metodológico: uso de pruebas de significación para demostrar homogeneidad entre los grupos de comparación de ensayos clínicos	
3.1 Bases teóricas generales.....	30
3.2 Empleo ceremonial del procedimiento.....	31
3.3 Análisis de su improcedencia.....	31
3.4 Ejemplos del empleo ceremonial.....	32
• Ejemplo 3.4.a.	32
• Ejemplo 3.4.b.	34
• Ejemplo 3.4.c.	35
• Ejemplo 3.4.d.	37
4. Procedimiento estadístico-metodológico: Confección de fichas técnicas para comunicar resultados de encuestas	
4.1 Bases teóricas generales.....	38
4.2 Empleo ceremonial del procedimiento.....	38
4.3 Análisis de su improcedencia.....	39
4.4 Ejemplos del empleo ceremonial.....	39

• Ejemplo 4.4.a.	39
• Ejemplo 4.4.b.	42
• Ejemplo 4.4.c.	42
• Ejemplo 4.4.d.	43
4.5 Una última observación.....	43
CONSIDERACIONES GENERALES Y RECOMENDACIONES.....	46
BIBLIOGRAFÍA.....	47

INTRODUCCIÓN

La estadística, sin dudas, se ha agenciado un espacio de gran importancia en casi todas las ramas del quehacer investigativo. En este trabajo nos interesa su aplicación particular en el campo de la medicina y la salud, donde la materia ha dado en llamarse *Bioestadística*.

Dentro de este ambiente se ha consolidado, sobre todo en los últimos decenios, como una disciplina íntimamente ligada a la investigación clínica y epidemiológica, campos en que ha demostrado con creces ser una herramienta de inestimable valor.

Sin embargo, hace tan solo cuatro décadas, la gran mayoría de los investigadores carecían de entrenamiento profesional y asesoría para la correcta aplicación de procedimientos metodológicos, principalmente de índole estadística; de ahí que las evidencias de sus trabajos eran informadas, con frecuencia, tan solo como simples hechos anecdóticos. Se manejaba incluso el criterio que apelar a tales recursos era totalmente injustificado en virtud de la dudosa calidad de los datos a los que habría de aplicársele^{1,2,3}.

Con el paso del tiempo esta situación se ha ido invirtiendo gradualmente, en la misma medida que se ha ido comprendiendo que el enfrentamiento a un problema, ya sea biomédico, clínico o de salud pública, requiere de una sistematización estricta, y que la perspectiva y análisis estadístico, sobre todo inferencial, es sumamente útil siempre y cuando la naturaleza del asunto lo exija.

Actualmente se observa la tendencia por parte de muchos investigadores de la salud a incorporar a sus medios de análisis, y de hecho no pocas veces lo hacen indiscriminadamente, alguna que otra técnica o procedimiento estadístico. Persuadidos de que lo ideal sería recurrir a los más intrincados y complejos, algunos los emplean irreflexivamente; de una manera inercial y carente de un razonamiento profundo, son insertados dentro de algoritmos metodológicos a manera de rituales o ceremonias, con la ingenua convicción de que proveerán poco menos que mecánicamente la respuesta adecuada a sus preguntas.⁴

En el ámbito científico contemporáneo parece olvidarse que los resultados obtenidos mediante análisis estadísticos son solo un elemento más que debe integrarse al marco teórico de conocimientos e información aportada por estudios anteriores.^{2,5} En consecuencia, se cometen cotidianamente graves errores dentro de los cuales uno de los más repetidos es la interpretación, en calidad de conclusión, de algo que no pasa de ser simplemente un resultado. Se ha perdido en estos casos la percepción elemental de que la estadística, como disciplina, está única y exclusivamente basada en las matemáticas; por tanto, su aplicación no puede ir más allá del mero examen cuantitativo de los datos y las inferencias numéricas, límite donde comienza la acción juiciosa e irremplazable del intelecto humano.⁶

Resulta llamativo que en la actualidad, para dar respuesta a disímiles problemas de investigación, se

recurra de manera algorítmica y mimética a puros formalismos metodológicos.

A lo largo del camino de un estudio, desde su diseño hasta la exposición de los resultados se aplican procedimientos, muchos de ellos cimentados en técnicas y/o *tests* estadísticos, que a pesar de ser en sí mismos utilizados a la perfección, según describe la teoría, no dan cumplimiento real al objetivo del investigador, pero éste actúa como si así hubiera ocurrido. Ocasionalmente, tales recursos se aplican siguiendo una costumbre o moda y no desempeñan papel efectivo alguno. Semejantes usos de procedimientos metodológicos son expresiones de lo que se ha decidido llamar en este trabajo *ceremonial* o *ritual* estadístico.

El empleo ceremonial de un procedimiento estadístico-metodológico se produce cuando su aplicación se inserta de manera inercial como parte de la metodología de un estudio, pero más en virtud de una tradición o de una expectativa externa, que como un recurso adecuado o necesario para resolver el problema encarado.

Las condicionantes históricas que favorecieron la aparición de este fenómeno en la investigación médica tienen un origen complejo y predominantemente exógeno, es decir, ajeno a la estadística propiamente dicha.¹ Pueden mencionarse, entre otras:³

- 1) **La falta de cultura investigativa de muchos profesionales:** esto, unido a la convicción propia de que cualquiera de ellos puede desarrollar una investigación, conduce, en no pocos casos, a aplicar a ciegas y esquemáticamente metodologías que ni siquiera entienden por qué y para qué usan. En el marco específico de la estadística, dada su gran complejidad teórica y difícil interpretación, este fenómeno se recrudece aun más.
- 2) **La enseñanza masiva, generalizada y despersonalizada de la estadística y la metodología de la investigación (cursos a distancias, teleconferencias, etc.):** Hoy en día se considera que dominar la Bioestadística junto a otros elementos necesarios en la investigación científica es una necesidad para el adecuado desempeño del profesional de la salud. Recibir cursos caracterizados por el mecanicismo da lugar a que el proceso educativo se oriente hacia el aprendizaje secuencial de códigos operativos, técnicas estadísticas e interpretaciones adocenadas con la inconveniencia aparejada de bloquear en el alumno el imprescindible razonamiento lógico y el pensamiento crítico.
- 3) **Costumbre:** La aplicación de los mencionados ceremoniales metodológicos es transmitida de generación en generación de investigadores, con lo que pasa a ser una costumbre supuestamente legitimada por la práctica.
- 4) **La frecuente exigencia de su uso por parte de los editores:** Una de las principales finalidades de la mayoría de las investigaciones en salud es la publicación en medios especializados; la demanda de que figuren ciertos requisitos metodológicos por parte de los editores puede hacer que se incurra en la

tentación de utilizar algunos recursos a toda costa como si dieran un aval de calidad y credibilidad a los trabajos.

- 5) La gran popularidad de que gozan:** Esta realidad es debida fundamentalmente a la imagen que exhiben de alto cientificismo y rigor; y sobre todo al hecho de ser fácilmente aplicables, gracias a los numerosos programas informáticos de procesamiento estadístico, hoy accesibles a la mayoría. Resulta contradictorio que el vertiginoso avance de los recursos computacionales e informáticos conspiren, de manera indirecta y solapada, en contra de la racionalidad investigativa.

Procede también mencionar el enmascarado trasfondo comercial y publicitario que no pocas veces alienta el empleo de estos productos¹. Los mercaderes del *software* intentan convencernos de que sus programas, dotados de las técnicas estadísticas más avanzadas y modernas, son instrumentos infalibles e indispensables para el desarrollo de investigaciones. Se crea un espejismo de posibilidades que ayuda a pensar que lo mejor y más importante es aprender el programa más sofisticado y reciente. Adquirirlo se torna entonces una necesidad para todos, incluso para aquellos que carecen de los conocimientos más elementales en materia de investigación pero que, de una manera u otra, anhelan incursionar en dicho escenario.

Infortunadamente, el empleo ritual de procedimientos estadístico-metodológicos es frecuente. Algunos ejemplos de ellos son:

- El cálculo del tamaño de muestra mínimo requerido y la exposición detallada en los informes y artículos científicos de las consideraciones para su cómputo.
- La implementación de estrictos y complejos diseños muestrales de tipo probabilístico (frecuentemente buscando la equiprobabilidad que evita realizar ponderaciones) sin la posterior determinación o exhibición de los errores de muestreo, es decir sin el cálculo de los intervalos de confianza.
- La segregación de los objetivos de un estudio en generales y específicos.
- El uso de la regresión logística y la regresión paso a paso para identificar factores de riesgo de enfermedades.
- La corroboración de homogeneidad entre los grupos de tratamiento de un ensayo clínico controlado a través de pruebas de significación.
- El empleo de pruebas de hipótesis para corroborar bondad de ajuste
- La evaluación de la fiabilidad de un instrumento de medición mediante coeficientes de correlación entre resultados obtenidos en dos momentos de aplicación.
- Empleo de fichas técnicas en la comunicación de los resultados de encuestas.

Una simple revisión webométrica en Internet permitirá hacer abundante acopio de investigaciones biomédicas, clínicas o sanitarias que incurren cotidianamente en ellos.

Es bueno aclarar que se ha estado haciendo referencia a un esquematismo de tipo conceptual que no

debe ser confundido con el de índole formal. Este último es aquel que define determinado conjunto de pasos o requisitos en aras de estandarizar y viabilizar ciertas acciones sin comprometer su calidad.¹

Tal es el caso, dentro del sector experimental, de los ensayos clínicos controlados. Para la validación de los mismos se exige el cumplimiento de una serie de requisitos estrictos dentro de los cuales, por poner un ejemplo, la asignación aleatoria de los pacientes ocupa un lugar metodológicamente prominente y esencial en el diseño. Definir los grupos de tratamientos, escoger las unidades de análisis y asignarlas al azar a cada grupo es crucial para eliminar los sesgos determinados por la subjetividad del investigador, lo que permite considerar que las muestras sean independientes y susceptibles de ser comparadas fructuosamente. Aquí se sigue formalmente un orden secuencial y lógico que garantiza el cumplimiento de estándares de calidad ya consagrados.

La aplicación repetida de recursos estadísticos a toda costa, constituye un serio peligro en la práctica de investigación, no solo por el marcado estancamiento intelectual que provoca en aquellos que se acomodan a usarlos, sino porque muchas veces son estériles y en gran medida contraproducentes.^{2,3,6-14} Conllevan de esta manera a un gasto innecesario de recursos y esfuerzos, y en el peor y no menos corriente de los casos, a publicaciones destinadas a engrosar el caudal de conocimientos erróneos sobre una materia. Al no contarse actualmente con un examen acucioso y serio de este fenómeno, raramente mencionados en los libros de texto y cursos sobre Bioestadística y Metodología de la Investigación, resultaría sumamente beneficioso desarrollar un estudio que alerte acerca de las mencionadas conductas espurias y las distinga de las verdaderamente útiles. Se contribuirá así a advertir acerca del negativo efecto que cotidianamente producen en detrimento de la buena práctica investigativa y de la estadística en sí misma.

No nos proponemos hacer un inventario exhaustivo de las susodichas ceremonias, ni tampoco cuantificar el grado en que aparecen aquellas que se seleccionen para ser examinadas. A los efectos de ilustrar el problema basta con exponer las bases teóricas de algunas de ellas así como discutir sus consecuencias. Llamar la atención sobre su presencia en la producción científica actual ayudaría a evitarlos y a facilitar la comprensión más generalizada de la estadística como herramienta fundamental pero solo complementaria del método científico.

A tenor de todo lo anterior, se propone debatir de modo sistematizado y racional algunos aspectos y expresiones de esta problemática. En esa dirección se orientan los siguientes objetivos:

OBJETIVOS

- 1) Caracterizar la naturaleza ceremonial con que pueden emplearse recursos estadísticos en la investigación biomédica, clínica y sanitaria.
- 2) Valorar los potenciales efectos perniciosos de tales ceremonias estadístico-metodológicas.
- 3) Ilustrar este fenómeno a través de ejemplos reales identificados en la literatura especializada actual.

METODOLOGÍA

A partir de una apreciación recurrente aunque informal acerca del modo inercial y acrítico con que ocasionalmente se emplean determinados recursos estadísticos en la investigación médica contemporánea, se concibió un estudio de índole predominantemente teórico-metodológica vertebrado en torno a un copioso examen bibliográfico. Primeramente se eligieron varios procedimientos estadísticos particularmente susceptibles de ser utilizados de la manera descrita, es decir, de forma ceremonial. Para la selección de los procedimientos que serían objeto de escrutinio detallado se procuró que ellos se vincularan a diferentes etapas por las que transita un estudio:

- 1) **Diseño:** Cómputo del tamaño de muestra mínimo requerido (n).
- 2) **Diseño:** Uso de pruebas de significación estadística para corroborar homogeneidad entre los grupos de comparación en los ensayos clínicos controlados.
- 3) **Análisis:** Empleo de la regresión logística para identificar factores de riesgo de enfermedades.
- 4) **Informe de los resultados:** Uso de las llamadas “fichas técnicas” en la exposición de resultados en encuestas.

La búsqueda de información se dividió en dos etapas o fases: una primera donde se hizo una compilación de aspectos teóricos que permitieran valorar el carácter potencialmente ceremonial de cada uno de los cuatro procedimientos seleccionados. La recopilación del material de referencia estuvo centrada en libros de texto y otras publicaciones de carácter científico que trataran elementos teóricos relacionados con la metodología de investigación y/o técnicas y procedimientos estadísticos, o incursionaran en análisis críticos sobre su uso ceremonial y las consecuencias nocivas a que esto conlleva. Sobre estas bases, se realizó una exhaustiva reflexión teórico-metodológica para cada procedimiento.

En la segunda fase, la revisión bibliográfica se encaminó a ilustrar, utilizando ejemplos de estudios reales, el empleo en forma ritual de cada uno de los procedimientos estadístico-metodológicos seleccionados. Ésta exploración se llevó a cabo utilizando las siguientes fuentes:

- 1) **Internet.** Se accedió a artículos científicos *on line* adoptando como motor de búsqueda “Google Académico” (*Scholar Google*). Los criterios de selección estuvieron definidos por las siguientes palabras claves y descriptores según el procedimiento estadístico-metodológico:
 - Para el cálculo de n : *cálculo del tamaño de muestra mínimo requerido y fórmula*.
 - Para las pruebas de significación para corroborar homogeneidad intergrupo en el marco de ensayos clínicos controlados: *características basales y ensayo clínico controlado*.

- Para la regresión logística en la identificación de factores de riesgo: *regresión logística y factores de riesgo; regresión paso a paso y factores de riesgo*.
- Para las “fichas técnicas” en la exposición de resultados en encuestas: *fichas técnicas y encuestas*.

2) **Revistas médicas cubanas.** Se trabajó con publicaciones tanto de corte biomédico, clínico como sanitario, procedentes de los últimos cinco años. La obtención de los números en formato electrónico (formato *html*) fue gestionada en la empresa editorial de la red informática *INFOMED*. Se seleccionaron aquellos en los que aparecieran las palabras claves arriba mencionadas.

3) **Tesis de doctorado.** Para la revisión de estos estudios se accedió al banco de tesis de doctorado de la Biblioteca de Ciencia y Técnica de la Academia de Ciencias de Cuba. Primeramente se solicitó el registro en formato digital (formato de texto *rtf*) de las tesis biomédicas (categoría que incluye todas las tesis de doctorado relacionadas con medicina), el cual contiene los datos generales y un resumen de cada investigación en particular. Se identificaron, y posteriormente se solicitaron en la Biblioteca, los trabajos donde aparecían las palabras claves.

4) **Revistas médicas españolas:** Se apeló a Medicina Clínica, Atención Primaria, y Revista de Salud Pública Española. Los artículos, en formato de texto *pdf* fueron revisados desde respectivos *CD-ROM* comercializados en el país de procedencia bajo los siguientes rótulos:

- *Medicina Clínica 1992-1998*, emitido en el año 1999 por *EDICIONES DOYMA*.
- *Medicina Clínica 2000-2004*, expedido en el año 2006 también por la editorial *DOYMA*
- *Atención Primaria 1994-2001*, expedido en el 2002 por la misma editorial.
- *Revista Española de Salud Pública 1991-2000*, emitido en el 2001 por el Ministerio de Sanidad y Consumo español.

Igualmente, se eligieron para su revisión aquellos trabajos donde figuraran las referidas palabras claves.

De cada estudio seleccionado, se precisaron sus objetivos, los resultados alcanzados y la discusión de éstos, así como las conclusiones emitidas. Se realizó un análisis crítico y exhaustivo de la metodología utilizada con el propósito de valorar:

- **Para el cálculo de n :** si en el artículo se brindaba información sobre el cómputo realizado para la determinación del tamaño muestral, las fórmulas y datos para aplicarlas. En los casos donde se daba la información necesaria se comprobaron los valores de n comunicados. Se valoraron además las inconsecuencias que pudieran concurrir en esta materia.
- **En relación con el uso de las pruebas de significación para corroborar homogeneidad intergrupos en el marco de ensayos clínicos controlados:** si en el artículo se hacía referencia a la realización de tales pruebas o si se exponía en los resultados las diferencias entre las “características

basales” de los grupos de tratamiento con los valores p de las pruebas de significación para cada variable analizada.

- **Para la regresión logística con vistas a la identificación de factores de riesgo:** cómo se consideraron y analizaron las variables ajustadas a la regresión, la realización de un cribado inicial de variables a través de pruebas de hipótesis, el criterio empleado para la determinación de asociación estadística de las covariables con la variable dependiente (valor p o intervalo de confianza para los *odds ratio*) y la elaboración de conclusiones sobre el hallazgo de factores de riesgo basándose única y exclusivamente en evidencias estadísticas. También se reparó en el uso de la regresión paso a paso como método seleccionador de variables.
- **Para el análisis del uso de fichas técnicas en los resultados de encuestas:** si la información incluida en ellas resultaba verdaderamente informativa.

Conviene reiterar que la búsqueda de ejemplos de aplicaciones ceremoniales no fue exhaustiva, pues el interés, según se señaló en los objetivos de esta investigación, no radicó en cuantificar el fenómeno sino en aportar evidencias reales de su existencia, de ahí que las fuentes hayan sido seleccionadas usando criterios razonados sin empleo de técnicas formales de muestreo o afines a éstas.

Para la exposición de los ejemplos que ilustraran cada ritual se consideraron aquellos artículos que, incurriendo en alguna de las aplicaciones ceremoniales de interés, presentaran, además, características que facilitarían su papel ilustrativo, así como que la metodología descrita fuera susceptible de comprobación y evaluación.

Con el propósito de viabilizar la exposición se construyeron escenarios hipotéticos con ocasionales modificaciones de los datos en algunos de los artículos originales.

La estructura elegida para el desarrollo del trabajo comprendió, para cada procedimiento estadístico-metodológico, principalmente los siguientes acápites o secciones:

- **Bases teóricas generales:** Comprende una exposición, a manera de introducción, de elementos teóricos generales del procedimiento estadístico-metodológico en cuestión y de los escenarios investigativos donde usualmente suele ser insertado.
- **Empleo ceremonial del procedimiento:** Donde se describe y explica detalladamente la ceremonia.
- **Análisis de su improcedencia:** En este epígrafe se realiza una deconstrucción teórico-metodológica de cada procedimiento estadístico estudiado. Incluye el análisis y esclarecimiento minuciosos de los elementos que condicionan, en determinados contextos, el carácter ceremonial y su naturaleza estéril o contraproducente.
- **Ejemplos del empleo ceremonial:** Se ilustra, mediante ejemplos reales obtenidos de la literatura científica actual, la inconsistencia del procedimiento.

DESARROLLO

1. Procedimiento estadístico-metodológico: Cálculo del tamaño de muestra

1.1 Bases teóricas generales:

En la mayoría de las investigaciones que se desarrollan en el campo de las ciencias biomédicas y de la salud resulta conveniente, generalmente por motivos económicos, recurrir al estudio de la población a través de una muestra.^{1,7,15} Con frecuencia el muestreo se asume incluso como única alternativa en situaciones donde es imposible acceder a cada uno de los integrantes del universo de estudio por ser éste en teoría infinito, o también cuando se realiza una intervención que implica riesgo para los participantes, situación en la que se impone involucrar, por razones éticas, la menor cantidad de personas posibles.

Lógicamente, mientras más pequeño sea el número de efectivos estudiados, la inversión de tiempo, recursos financieros, materiales y humanos será proporcionalmente menor. Sin embargo, el hecho de que todo proceso de muestreo lleva implícito futuros errores inferenciales (precisamente porque la información disponible se obtiene a partir de datos que son solo una fracción de todos aquellos que reflejan exactamente la realidad que se observa) determina que optar por un número pequeño de unidades de análisis produzca imprecisiones en los resultados de la investigación, conocidas como errores de muestreo. En ocasiones tales errores son de tal magnitud que pueden cancelar su valor a los efectos de aportar conocimientos provechosos.¹⁵⁻¹⁷

De lo anterior se deriva un razonamiento tan importante como intuitivo: el tamaño de muestra también tiene un papel determinante sobre la dimensión del error muestral y su relación es inversa a éste, de manera que a menor número de efectivos estudiados, mayor será el tamaño de los errores dependientes del proceso de selección. En síntesis, es necesario seleccionar la menor cantidad posible de unidades de análisis, que supongan errores inferenciales aceptables dentro de una exigencia económica asumible.

En auxilio de tal razonamiento se han desarrollado numerosas fórmulas que teóricamente permiten calcular el tamaño mínimo necesario. Cada una de estas expresiones matemáticas conciernen específicamente a algún parámetro poblacional y a la acción básica que se vaya a realizar en relación a éste: bien estimarlo, bien valorar mediante *tests* de significación si su comportamiento difiere en cierta magnitud para dos o más poblaciones.^{1,7,15-17}

Usualmente, las estimaciones se realizan en el marco de los llamados estudios descriptivos donde el objetivo consiste en caracterizar determinada población a través de diversos parámetros. Generalmente se trata de medidas de resumen (proporciones, porcentajes, razones, índices, medias y medianas), de posición (deciles, cuartiles y, en general, percentiles) y totales absolutos, aunque también con propósito descriptivo pueden realizarse estimaciones sobre coeficientes de correlación y regresión. La fórmula específica para calcular el tamaño muestral en base a la mayoría de estos estadígrafos suele ser en extremo compleja. Por lo general son poco tratadas en la literatura especializada y, menos aun, aplicadas

en la práctica. Dada su sencillez, aquella a la que más se acude es la relacionada con la estimación de un porcentaje y, en menor grado, de una media aritmética.^{1,7}

Lo mismo sucede en el ámbito analítico, donde usualmente se aborda la problemática para calcular el tamaño muestral cuando se desea contrastar, mediante respectivas pruebas de significación, porcentajes o medias aritméticas pertenecientes a sendas poblaciones.

Con mucha frecuencia se aprecia en los informes de investigaciones la exposición detallada de la metodología que se ha seguido para determinar el tamaño de muestra utilizado. La tendencia general es a exhibir la expresión matemática que conduce a este valor y/o a argumentar las consideraciones tenidas en cuenta para establecer todos o la mayoría de los valores que intervienen en esta.

Las fórmulas que con mayor regularidad se implementan son las siguientes:^{1,7}

- Para la estimación de un porcentaje o media poblacional:

$$n = \frac{n_o}{1 + \frac{n_o}{N}} ; \quad n_o = \frac{Z_{1-\alpha/2}^2 P(100 - P)}{E_o^2} \quad [4.1]$$

Donde: n representa el tamaño de muestra mínimo requerido, $Z_{1-\alpha/2}^2$ el percentil $1-\alpha/2$ de la distribución normal estándar; P es el porcentaje que se quiere estimar; N el tamaño de la población muestreada y E el error absoluto mínimo admisible.

- Para la comparación mediante pruebas de hipótesis de una media o un porcentaje en dos poblaciones mediante estudios observacionales analíticos (casos y controles, estudios de cohorte) e incluso experimentales.

$$n = \frac{[\sqrt{(Z_{1-\alpha/2})(2\delta^*)} + \sqrt{(Z_{1-\beta})(\delta_1 + \delta_2)}]^2}{(\delta_1 + \delta_2)} \quad [4.2]$$

Donde α representa la probabilidad máxima admisible de rechazar indebidamente la hipótesis nula (error de tipo I), β la probabilidad máxima admisible de no rechazar la hipótesis nula cuando debe hacerse (error de tipo II); δ_1 y δ_2 son las desviaciones estándar correspondientes a los parámetros de cada población involucradas en el estudio y que son calculadas de la siguiente manera si la variable es cualitativa dicotómica:

$$\delta_1 = P_1(100 - P_1) \quad y \quad \delta_2 = P_2(100 - P_2) \quad [4.3]$$

Donde P_1 es el valor estimado del parámetro de mayor interés en la muestra 1 y P_2 , el valor del mismo parámetro en la muestra 2.

- Para la comparación, mediante pruebas de hipótesis de porcentajes o medias entre dos grupos de tratamiento en el marco de un ensayo clínico:

$$n = \frac{M}{(DS)^2} \quad ; \quad M = 2(Z_{1-\alpha} + Z_{1-\beta})^2 \quad ; \quad DS = \frac{\phi}{\delta} \quad [4.4]$$

Donde ϕ representa el beneficio mínimo que resulte clínicamente interesante con el nuevo tratamiento. Si el resultado del tratamiento es cuantitativo, el beneficio se cuantifica a través de la diferencia entre los resultados medios con el nuevo tratamiento y con el tratamiento estándar. Si el resultado del tratamiento es cualitativo dicotómico, por ejemplo éxito o fracaso, el beneficio será la diferencia entre los porcentajes de éxito con el nuevo tratamiento y con el tratamiento habitual. δ es la desviación estándar de la respuesta entre sujetos.

1.2. Empleo ceremonial del procedimiento.

Independientemente de los objetivos de un estudio y su diseño, la determinación del tamaño de muestra mínimo requerido contiene, entre otros aspectos generalmente escamoteados en los textos, una enorme e inevitable carga de subjetividad.^{1,7,15}

Justificar, basándose en expresiones matemáticas, el tamaño de muestra utilizado no confiere objetividad a tal decisión ni aporta elementos de rigor científico pues todo dependerá de los valores que serán sustituidos en cada fórmula y que fueron determinados por la percepción individual del fenómeno, la intuición y la experiencia del investigador, adscritas al desarrollo de los conocimientos alcanzados en la materia que se estudia y fundamentalmente de los recursos disponibles. Tales realidades conducen a asegurar que no existe para cada investigación un óptimo tamaño de muestra único. Calcularlo algebraicamente se diferencia muy poco de establecerlo usando el sentido común.

Por esta y otras razones que se exponen debajo, consideramos que operar con fórmulas, y exponer luego parcial o totalmente los datos tomados en cuenta para arribar al tamaño de muestra mínimo entraña una conducta meramente ceremonial.

Consideremos en detalle el problema:

1.3 Análisis de su improcedencia:

1) La determinación de los parámetros que demanda la fórmula es de total carácter subjetivo:

Para argumentar la subjetividad implícita en este procedimiento tomaremos como ejemplo la determinación de cada uno de los parámetros que exige la fórmula del tamaño de muestra para estimar un porcentaje poblacional (fórmula [4.1]). Cabe enfatizar, sin embargo, que las consideraciones para esta situación particular no se diferencian de aquellas a tener en cuenta para las fórmulas que puedan ser utilizadas ([4.2], [4.3], [4.4]) u otras.

- **El tamaño poblacional (N):**

Como puede apreciarse para llevar a cabo el cálculo del tamaño de muestra se requiere de un valor que represente el total de personas que componen la población de la cual va a ser extraída la muestra de estudio (N) a pesar de que es bien sabido que la calidad de una estimación depende mínimamente del tamaño poblacional. Consecuentemente, el valor del dato no tiene que ser exacto. Ocurre incluso que si se aumentara el valor de N mientras se mantienen el resto de los parámetros de la fórmula constantes, a partir de cierto umbral ya no se producen incrementos en el tamaño de muestra.

No obstante esta observación, en la mayoría de los casos, incluso tratándose de poblaciones finitas, es extremadamente difícil y con frecuencia imposible, conocer exactamente cuántas unidades componen una población, a menos que ésta sea realmente tan pequeña como para que sus integrantes puedan ser fácilmente identificados. Esto determina que cuando se trabaja sobre poblaciones relativamente grandes su tamaño se obtiene mediante estimados que brindan con mejor o peor calidad una aproximación.

Sin embargo, el valor de N se suele conocer con bastante precisión, por lo cual no es este un elemento que aporte sensiblemente subjetividad a la determinación del tamaño muestral.

- **El porcentaje para la variable “de mayor importancia” y su preestimación (P):**

La fórmula para el cálculo del tamaño muestral en el caso que nos ocupa requiere, paradójicamente, que se anticipe el propio valor P que se desea estimar o comparar. Puesto que en cualquier estudio se manejan un nutrido grupo de porcentajes poblacionales y como el tamaño de muestra debe ser único para todo el estudio, la solución usualmente sugerida consiste en seleccionar aquel que se considere “de mayor relevancia” y hacer el análisis correspondiente al tamaño de muestra para este.

Esta discriminación impone una clara carga de subjetividad. Basta reparar en que usualmente una tabla con propósito descriptivo exige una estimación por cada una de sus celdas y en ocasiones suelen construirse muchas tablas con numerosas casillas, donde todas contienen un parámetro de interés. Resulta obvio que la decisión sobre cuál ha de elegirse inevitablemente entraña una decisión de fuerte corte subjetivo.

Por otra parte, suponiendo que un porcentaje se distinga de otros, por su importancia para la investigación, su selección implica que el error máximo que se acepta cometer (E_0) sea válido solo para éste mientras que para los otros tal consideración queda en un limbo.

Pero mucho más serio es el conflicto que crea la demanda de comunicar de antemano el valor del porcentaje que se desea estimar. En principio se trata desde luego, de un problema sin solución real por obvias razones lógicas. Para escapar de ese laberinto (hay que anticipar el valor cuyo

desconocimiento es la razón del estudio) se sugiere usar un valor de P reportado por estudios similares, o el que dimane de un estudio piloto que aporte una cifra aproximada. Resulta, además de paradójica, extremadamente dependiente de la subjetividad del investigador determinar de dónde escoger su preestimación, pues pueden existir diferentes estudios con disímiles resultados de P . La paradoja se resume en que el valor preestimado de P , sin ser conocido a ciencia cierta, debe ser cercano al verdadero parámetro poblacional pues si su estimación previa fuera muy errónea influiría de manera negativa en el resultado del cálculo del tamaño muestral; al mismo tiempo si se conociera verdaderamente con bastante precisión el parámetro, carecería de toda utilidad estimarlo ya que la finalidad del proceso es justamente obtener una buena aproximación del parámetro en cuestión.

En la fórmula para el cálculo de muestra en relación a una media aritmética el proceso se agrava pues la preestimación no corresponde a la media poblacional sino a su varianza, parámetro aún más esquivo.

Una de las soluciones que ha sido sugerida con éxito, pues su aplicación en la práctica resulta bastante frecuente, consiste en asumir, caso de que la magnitud de P en la población sea totalmente desconocida, que ésta es igual a 50%. Con este artificio matemático se consigue que la cantidad $P(100-P)$ adopte su mayor valor produciendo el mismo efecto en el tamaño muestral, dada en la relación directamente proporcional que guarda con él.¹⁸⁻²²

Pero el argumento es falaz, pues esto solo sería válido siempre que E_o esté fijado de antemano. Sin embargo, el máximo error que se puede admitir al estimar P no puede establecerse razonablemente hasta que no se suponga su magnitud, y cuando ésta se desconoce totalmente no existen criterios para considerar aceptable un error en términos absolutos. Lo que sí pudiera fijarse de antemano es un error relativo máximo admisible cuyo cálculo se realiza mediante la siguiente expresión:

$$E_r = \frac{E_o}{P} \quad [4.5]$$

La fórmula correspondiente para el cálculo del tamaño muestral puesta en términos del error relativo en vez del error absoluto (sustituyendo [4.5] en [4.1]) sería:

$$n = \frac{n_o}{1 + \frac{n_o}{N}} ; \quad n_o = \frac{Z_{1-\alpha/2}^2 \frac{100-P}{P}}{E_r} \quad [4.6]$$

Pero al fijar el error en términos relativos, el tamaño de n_o (y por ende, de n) crece en la medida que disminuye, un resultado coherente con la intuición: si el rasgo cuya prevalencia se quiere estimar es muy poco frecuente, entonces el tamaño de muestra necesario ha de ser muy alto.

Siendo así, la regla del 50% pierde todo sentido, ya que si se trabaja con errores absolutos, éstos no se pueden establecer racionalmente bajo el supuesto de que no se tiene idea alguna de cuál es el

valor de P ; y si se trabaja con errores absolutos, deja de ser válida la afirmación de que n está acotada superiormente en virtud de que lo está el producto $P(100-P)$.

- **El nivel de confiabilidad:**

No existe criterio que determine que, dada determinadas circunstancias, la confiabilidad planificada para las inferencias de una investigación requiera de una magnitud específica. Se suele elegir 95%, pero este número no tiene más méritos que 90% o 99%, o 93,7%. La subjetividad, por tanto, también aparece en este caso.

- **La determinación del error mínimo admisible, E_0 :**

Cuando se calcula el tamaño de muestra mínimo requerido se debe realizar una preestimación de P , no solo porque aparece explícitamente en la fórmula, como ya vimos, sino porque sin este conocimiento es imposible decidir el valor del error absoluto que también aparece en ella.

En la identificación a priori del error máximo admisible se deberá considerar qué valor de este resulta lo suficientemente pequeño como para considerarlo admisible, decisión que no deja de ser esencialmente subjetiva pues, tal y como sucede con la determinación del nivel de confianza, no se puede establecer un juicio carente de subjetividad que precise, según las particularidades de la investigación, si determinado error es más adecuado que otro.

Existen además dos elementos que en determinadas ocasiones culminan el bloque de fórmulas para el cálculo del tamaño muestral. Estos no se deben pasar por alto, sobre todo debido a que la determinación de sus valores también entraña decisiones subjetivas. Su participación generalmente tiene lugar en estudios observacionales donde la aleatorización solo se garantiza mediante el muestreo, comúnmente aplicado antes de obtener el consentimiento de los efectivos y que en ocasiones debe adoptar formas complejas mediante la estratificación de la población, a veces en varias etapas.

- **El efecto de diseño (*deff*):**

Todas las expresiones matemáticas que permiten calcular el tamaño de muestra se basan en el muestreo simple aleatorio, aunque con frecuencia los diseños muestrales que se implementan en la mayoría de las investigaciones son de naturaleza compleja, con etapas de selección que incluyen conglomerados y/o estratos.^{1,7,22-25} En estos casos los errores no solo dependen del tamaño total de la muestra sino también del número de conglomerados en cada etapa y de sus tamaños cuando se usa un muestreo polietápico.

Para alcanzar bajo estas circunstancias un grado de precisión previamente establecido (E_0) es necesario seleccionar una muestra de mayor tamaño que el que demanda el muestreo simple aleatorio. De ahí que resulte necesario, tras obtener el valor aportado por la fórmula inicial,

multiplicarlo por un número (el llamado efecto de diseño, *deff*) que contrarreste el efecto desfavorable del diseño muestral sobre la exactitud de las inferencias.

La cuantificación de este valor es totalmente subjetiva. Algunos autores plantean como regla asumir un *deff* igual a 2,^{1,7} mientras que otros, con un poco más de lógica, sugieren emplear, teniendo en cuenta el grado de complejidad del diseño muestral, valores entre 1,5 y 2, o entre 1,5 y 4. Cálculos en algunos estudios orientados a precisarlo han evidenciado que su valor cambia radicalmente de un diseño muestral a otro, de una muestra a otra, y que dentro de un mismo diseño inclusive suele modificarse notablemente en dependencia del parámetro elegido.¹

El valor empleado como *deff* es quizás lo que con más subjetividad se anticipa.

- **La tasa de no respuesta:**

La muestra sufre generalmente una reducción de su tamaño cuando en la práctica los efectivos que al inicio fueron seleccionados en el proceso de muestreo no consienten participar en el estudio o no pueden ser interrogados o medidos por alguna razón. La magnitud en que se presenta este fenómeno, conocida como “tasa de no respuesta”, debe ser tenida en cuenta por el investigador quien, para atenuar su efecto, deberá incrementar el valor resultante de la fórmula en una cantidad que compense por las personas que no podrán ser estudiadas.

Establecer previamente la cantidad de personas que no van a poder ser encuestadas es literalmente imposible. Solo puede especularse atendiendo a características muy particulares de cada estudio y de la población sobre la que se trabaja. El valor que se determine finalmente como tasa de no respuesta, incluso cuando se tomen como ejemplo estudios de iguales características, será en cualquier caso un acto con importantes “cuotas de adivinación”.^{1,7}

Todas estas fuentes de subjetividad en el proceso de decidir un tamaño de muestra son ocasionalmente reconocidas en la literatura. No obstante, también pueden encontrarse explicaciones doctrinarias como, por ejemplo la que se expone a continuación, procedente de un artículo publicado en México:²⁶ “... podría parecer que la suposición de estos valores es extremadamente arbitraria; sin embargo, es mejor intentar esta aproximación a llevar a cabo el estudio sin intentarlo...” Los autores no explican, sin embargo, los elementos por los cuales consideran mejor hacer suposiciones arbitrarias.

2) La cantidad de recursos disponibles es lo que rige la determinación del tamaño de muestra:

A pesar de que se han realizado numerosas propuestas teóricas para incluir en las fórmulas parámetros que responden a aspectos de orden económico, éstas son por lo general desconocidas, o bien desdeñadas por ser expresiones extremadamente complejas y abigarradas cuya aplicación

exige especificaciones de costos y variabilidades que escapan del campo del investigador en salud sin dejar de estar también sujetas a un alto grado de subjetividad.²⁶⁻²⁷

Sin embargo, lo realmente concluyente para decidir el tamaño de muestra son los recursos; en consecuencia, la metodología frecuentemente se encamina a determinarlo basándose en la disponibilidad de estos y posteriormente ajustar los datos de las fórmulas en correspondencia a este valor.^{1,7,15} Esta manipulación ha sido descrita en varios artículos especializados con el nombre de reacondicionamiento o *retrofitting*.⁹ Su aplicación constituye la más clara expresión del empleo ceremonial de las fórmulas.

Utilizar, como elemento orientativo general, fórmulas para calcular el tamaño de muestra requerido en un estudio no lo consideramos censurable en sí mismo; pero sí consideramos espurio, en virtud de todo lo desarrollado hasta el momento, ofrecer justificaciones formales con el fin de avalar científicamente el tamaño seleccionado como si fuera el único recurso empleado.

Antes de examinar algunos ejemplos, procede resaltar que la exposición en un artículo de los elementos considerados para elegir el tamaño de muestra, ya sea mediante cálculos matemáticos o análisis de disponibilidad de recursos, en realidad no resulta pertinente, carece de toda utilidad. En rigor, al lector no le interesa para nada conocer las motivaciones de los autores para trabajar con un tamaño de muestra u otro. Es cierto que algunas revistas lo exigen o sugieren, pero se trata de una información absoluta y totalmente superflua a los efectos de valorar la calidad del trabajo o de entender sus resultados o sus interpretaciones. El tamaño empleado influirá en la calidad de las estimaciones y en la potencia de las pruebas de hipótesis, pero conocer o no por qué se eligió el tamaño en cuestión no cambia ni una ni otra. El tamaño es algo ya consumado y lo que interesa al lector es conocer qué se puede sacar en limpio de lo que objetivamente se ha hecho, y no las motivaciones que pudieran haber existido para hacer una cosa u otra en materia de tamaño muestral. Conocer esas razones es tan útil como saber por qué el número de autores del trabajo fue el que fue, o la calidad de la computadora que emplearon.

Debe quedar claro que lo dicho no es válido cuando hablamos de un proyecto o protocolo: el financiador sí tiene que recibir información acerca de las razones para planificar determinado tamaño.

A continuación se exponen algunos ejemplos reales que permiten apreciar diferentes expresiones ceremoniales relacionadas con el cálculo del tamaño de muestra.

1.4. Ejemplos del empleo ceremonial

Ejemplo 1.4.a.

El estudio “Características de la primera relación sexual en adolescentes escolares de Ciudad de La Habana”,²⁸ desarrollado por el Departamento de Salud Sexual y Reproductiva del Instituto Nacional de

Endocrinología, se trazó el objetivo de determinar las características de la primera relación sexual en un grupo de jóvenes de la capital. Con ese fin se llevó a cabo una investigación de carácter descriptivo, que exigió la aplicación de un cuestionario autoadministrado a una muestra de adolescentes entre 15 y 18 años de edad, residentes en Ciudad de la Habana durante los meses de enero a mayo del 2002.

La elección del número necesario de jóvenes para encuestar y la metodología del estudio en general, se explica textualmente en el artículo del modo siguiente:

El cálculo del tamaño muestral se realizó sobre la base del total de estudiantes de esta provincia vinculados al Sistema Nacional de Educación de la Enseñanza Regular (N=56 699) que se encontraban estudiando en preuniversitarios e institutos tecnológicos. Según referencia se estimó que alrededor del 45% de la población ha iniciado las relaciones sexuales en la adolescencia y se consideró un error de 2% entre la frecuencia real y la esperada y un nivel de significación* del 99%, por lo que la población a encuestar sería de 1581, pero teniendo en cuenta la posibilidad de tener el 5% de no respuesta, la muestra final a encuestar era de 1780 adolescentes y finalmente quedó constituida por 1 675, con una pérdida del 4,9%, tal como se había previsto. Para lograr una mayor representabilidad de la muestra, se estratificó de acuerdo al tipo de enseñanza y edad y se organizaron un total de 6 estratos. Tipo de enseñanza: preuniversitario urbano (15-16 años y 17-18 años), preuniversitario rural (15-16 años y 17-18 años) e Instituto Politécnico (15-16 años y 17-18 años). En la selección de los politécnicos se tuvo en cuenta que existiera una representación de cada una de las especialidades.

Como puede observarse, se comunican todos los datos teóricamente requeridos para el cómputo del tamaño muestral en el supuesto de que se tratara de una muestra simple aleatoria, los cuales serán sustituidos en la fórmula específica para estudios descriptivos.

Según el artículo los valores en cuestión serían:

- $N = 56\ 699$
- $P = 45\%$
- $E_o = 2\%$
- $Z(1 - \alpha/2) = 2,575$ que resulta de un $\alpha = 0,01$ (es decir, confiabilidad de 99%)
- Tasa de no respuesta = 5%

Puesto que ya se expuso con detalle la naturaleza subjetiva de la determinación de cada uno de estos valores; nos centraremos en mostrar, entre otros aspectos, la naturaleza ritual de su presentación con el fin de justificar el tamaño de muestra asumido.

Para comenzar, nótese el valor tan exacto con que los autores han asumido el parámetro N (56 699).

Evidentemente se ha desdeñado la aproximación numérica de esta cifra (por ejemplo, a 57 700) creándose una ilusión de extrema rigurosidad y exactitud en el cálculo.

Tal precisión contrasta con los valores de otros parámetros como P y la tasa de no respuesta cuyas estimaciones no dejan de ser, por definición, un tanto especulativas. Por lo demás, es obvio que, independientemente del método o la fuente con que se obtuvo la cantidad de estudiantes vinculados al Sistema Nacional de Educación de la Enseñanza Regular en Ciudad de la Habana, dicho número está sometido a constantes y dinámicas variaciones que impiden precisarlo a ciencia cierta. Pero cabe destacar otros elementos de mayor importancia a nuestros propósitos.

El procesamiento estadístico de los datos se llevó adelante teniendo en cuenta solo una parte de la población muestreada; en efecto, las estimaciones de los porcentajes según las respuestas brindadas a cada pregunta del cuestionario se diferenciaron posteriormente atendiendo a las revelaciones emitidas por varones y hembras, y solo para aquellos jóvenes que ya habían experimentado su primera relación sexual, como se recoge en dos de las tres tablas de resultados que aparecen en el artículo. La Tabla 1 es una de ellas.

Tabla 1. Motivos para tener la primera relación sexual

Motivos	Femenino		Masculino		Significación
	#	%	#	%	
Deseos de probar	167	27,1	449	72,9	$\chi^2 = 189,4$ p = 0,001
La pareja lo deseaba mucho	79	28,0	203	72,0	
Todos los compañeros(as) lo hacían	30	32,6	62	67,4	
Consideraba que tenía edad para hacerlo	68	25,8	196	74,2	
Estaba enamorado(a)	378	58,0	274	42,0	
Le obligaron	11	44,0	14	56,0	
Otros	32	69,6	14	30,4	

Fuente: Revista Cubana de Salud Pública. 2006. Volumen 32, Número 3.

Según se plantea textualmente en la publicación:

1 195 (71,3 %) adolescentes ya habían iniciado las relaciones sexuales, de ellos el 42,6 % correspondieron al sexo femenino y el 57,4 % al masculino.

Vale decir, solo 1 195 podían ser interrogados sobre ciertos temas, pues es esta la cantidad que ya ha experimentado relaciones sexuales. Esto se traduce a que serán desechados del estudio 480 de los jóvenes que inicialmente fueron escogidos por muestreo. Ha ocurrido que, en el fragor del estudio, el tamaño de muestra originalmente calculado descendió de 1 675 a 1 195. Tal efecto reductor sobre *n* conlleva a un inexorable aumento del error de muestreo inherente a las estimaciones.^{1,7}

Dicho efecto no se consideró en lo absoluto para determinar el tamaño de muestra. Recuérdese que su cómputo fue definido para la estimación de un solo parámetro que incluye a toda la población bajo estudio

(porcentaje de adolescentes que ya han tenido relaciones sexuales). Notablemente, la teoría a la que una y otra vez se apela para el cálculo de n tampoco contempla estas reducciones, que aparecen casi siempre en la práctica, como efectivamente ocurre aquí.

Por otra parte, en el artículo se informa que, una vez procesados los datos, el porcentaje de adolescentes que ya sostenía una relación sexual resultó ser 71,3%. La comparación de este valor con el que se ha prefijado para P (45%), arroja que la preestimación fue errónea. El hecho de realizar preestimaciones de P poco certeras no solo es lógico sino que confiere inevitablemente un carácter errático al tamaño de muestra calculado sobre su base.^{1,7}

Adiciónese a todo esto lo que consideramos la mayor evidencia del carácter ceremonial de todo este procedimiento:

Puede comprobarse fácilmente que calculando n mediante la fórmula descrita para estudios descriptivos y luego considerando un d_{eff} igual 1 (en el estudio no se informa que el diseño muestral implementado haya incluido la selección de conglomerados), se obtiene como resultado 3 826, que posteriormente habría que incrementar en 191 unidades al considerar la supuesta tasa de no respuesta de 5% (191,3 representa el 5% de 3 826), lo que determina que el número ascendería a 4017, cifra muy distante de la calculada para la investigación (1 780).

Los párrafos precedentes ponen de manifiesto que en este estudio el cálculo realizado para determinar n fue erróneo desde el punto de vista matemático. El interés de señalar estos equívocos no reside, sin embargo, en la posible crítica a autores o editores (válida, pero irrelevante ahora) sino en lo siguiente:

A pesar de que los autores han utilizado, aparentemente a pesar de sus propios deseos, un tamaño de muestra muy inferior al que pudiera sugerir la teoría, esto no ha determinado la invalidez científica del estudio ni la ausencia en éste de conocimientos que puedan ser de utilidad.

A lo sumo, las consecuencias se centrarán en la obtención (en no todos los casos) de errores de muestreo mayores e intervalos de confianza más amplios con la consiguiente pérdida de precisión. En este punto caben las siguientes preguntas. ¿Pudiera haberse llevado adelante la investigación si en vez de encuestar solo 1 675 adolescentes los autores consideraran obligatorio aplicar el cuestionario a 4 017? ¿Habrían sido suficientes los recursos para encuestar esa mayor cantidad de jóvenes? Quizás no. Pero lo cierto es que el estudio se hizo y que los errores cometidos en el andamiaje formal para el cómputo de n no produjeron ninguna catástrofe visible. En este ejemplo, sin embargo, no podemos evaluarlo claramente, ya que se ha omitido, verdadera prueba del carácter ceremonial de todo el procedimiento, el cómputo de los errores. Las estimaciones, principalmente de porcentajes poblacionales, se han mostrado solo de manera puntual (obsérvese la Tabla 1), modo que en definitiva impide la apreciación de la inexactitud cometida.

Ejemplo 1.4.b.

Hernández, García, Pérez y Ramos en su artículo “Información sobre prevención de accidentes que poseen adultos y niños a su cuidado”,²⁹ publicado en el año 2001 por la Revista Cubana de Higiene y Epidemiología desarrollaron “un estudio descriptivo transversal sobre la información recibida en prevención de accidentes”. Allí puede leerse:

...el universo estuvo constituido por las 5 370 personas responsables del cuidado de los niños menores que 10 años en el hogar, en el municipio Regla, de la provincia de Ciudad de La Habana, durante los meses de diciembre de 1996 a marzo de 1997.

Se calculó un tamaño muestral de 1 156 de estos adultos, tomando en consideración que se estuvo dispuesto a aceptar un error de muestreo no mayor del 10,0 % del parámetro a estimar (porcentaje de adultos que cuidaban niños y que estuvieran desinformados sobre la prevención de accidentes) y un efecto de diseño de 1,5. Como no se encontraron estudios en la literatura que trataran este aspecto, para el cálculo del tamaño muestral se tomó como cifra de desinformados el 30,0 % hallado en una prueba piloto realizada previamente en el municipio Habana Vieja.

Se utilizó el muestreo por conglomerados menoteápico en la selección de la muestra. Para ello, se consideraron como conglomerados los 64 consultorios del médico de la familia existentes en el municipio. De estos conglomerados se seleccionaron 17, en los que se estudió a toda la población de personas que cuidaban a los niños menores que 10 años, que constituyeron las unidades de análisis”.

Como puede apreciarse, en este trabajo también se da cuenta detallada de los valores seleccionados para los parámetros que demanda la fórmula para el cálculo de n en el marco de estudios descriptivos.

En resumen se informa que se trabajó con:

- $N = 5\,370$
- E_R : no mayor del 10%
- $d_{eff} = 1,5$
- $P = 30\%$, a partir de un estudio piloto realizado.

Suponiendo que se empleó el sacralizado nivel de confiabilidad de 95% y un error del 3% (el 10% de $P = 30\%$ es precisamente 3%) se puede corroborar, mediante sustitución de los valores anteriores en la fórmula [4.1] que el valor resultante de n es, en efecto, igual a 1 156.

Sin embargo a lo largo de la metodología descrita en el artículo sobresalen algunos aspectos que llevan a pensar que los autores arribaron primero a n y luego manipulado mediante *retrofitting* el resto de los parámetros de la fórmula para que se ajustaran a este valor.

En primer lugar llama la atención que en este trabajo, donde se ha seguido al pie de la letra lo que expone la teoría oficial para el cálculo de n , no se tiene en consideración la posible tasa de no respuesta. Los autores deciden seleccionar del universo exactamente la cantidad de unidades de análisis que ha arrojado el cómputo, sin considerar la posibilidad de que alguna cantidad de personas seleccionadas por el muestreo se niegan a ser encuestados o no puedan ser contactados. Resulta simplemente insólito que se haya logrado estudiar al 100% de los 1 156 individuos que arrojó la fórmula (tasa de no respuesta igual a cero). A nuestro juicio esto constituye una prueba irrefutable del reacondicionamiento de n , ya que tal tasa es imposible en la práctica.

Otro aspecto no menos indicativo concierne al hecho de que el diseño muestral seleccionado ha sido el de conglomerados monoetápico. Este tipo de muestreo teóricamente produce muestras de tamaño aleatorio, que dependen del número de integrantes contenidos en los conglomerados que resulten seleccionados.¹ A pesar de esto, una vez elegidos al azar los 17 consultorios (dentro de un total de 64), según el diseño muestral del estudio, se encuentra el inaudito resultado de que la cantidad total de individuos incluidos coincide exactamente con la previamente calculada para el tamaño de muestra (1 156).

Parece no quedar dudas de que lo que realmente se ha hecho en la práctica es identificar primero los conglomerados a elegir, seleccionarlos aleatoriamente, realizada la encuesta y una vez conocida la cifra de los que hayan respondido a la encuesta, maniobrar con la fórmula para el cálculo de n de manera que su resultado coincida con la cifra de personas realmente estudiada.

Resulta difícil hallar una expresión más clara de la dictadura que ejerce la ceremonia de la determinación de tamaños muestrales mediante fórmulas.

Ejemplo 1.4.c.

En ocasiones la exposición del cálculo de n toma un carácter ceremonial desmesurado cuando se exhibe todo el respaldo matemático para la determinación de una cifra supuestamente acorde a los objetivos del estudio y posteriormente, por la razón que sea, se toma otra muy superior a la calculada.

Así ocurrió en el informe final del estudio llevado a cabo en el Instituto Superior de Medicina Militar "Dr. Luis Díaz Soto" de Ciudad de la Habana, bajo el título "Hilo quirúrgico *Agasut QE* como una alternativa para el cierre de aponeurosis y piel en incisiones de laparotomía",³⁰ presentado en calidad de tesis de doctorado en el año 2004.

Como objetivo general la investigación se trazaba el siguiente:

Evaluar la eficacia del *Agasut* en el cierre de aponeurosis y piel mediante un ensayo clínico (fase III) y compararla con los hilos convencionales.

Con ese fin se realizó en el servicio de cirugía general del referido instituto un ensayo clínico fase III prospectivo, aleatorio, simple ciego y controlado en el cual se incluyeron a todos los pacientes que llegaron al

centro por enfermedades no traumáticas que requerían tratamiento quirúrgico con urgencia mediante incisiones de laparotomía.

El autor, refiriéndose al tamaño de muestra empleado y al método que a su juicio avala su pertinencia, escribe en la sección correspondiente al análisis estadístico, lo que sigue:

El cálculo del tamaño de muestra se realizó de forma automatizada mediante la hoja de cálculo de Excel de Windows 95 con el nombre de archivo *t-muestra*. En nuestro caso se tomó $\alpha = 0,05$; $P_o = 60\%$; $\beta = 0,4$ (tamaño del intervalo de confianza) y un 5% de pérdida. Se determinó la necesidad de 24 pacientes como mínimo por grupo; pero se logró llevar hasta 60 personas.

La información que se brinda es bastante confusa (y en cualquier caso insuficiente para corroborar los cálculos). Esto de por sí, transparenta la naturaleza ritual de lo que se comunica.

Por otra parte, la inclusión en el ensayo de una mayor cantidad de pacientes que la calculada otorga una clara invalidez a la exposición de los parámetros que determinaron el tamaño de muestra. La cifra calculada (24) parece enarbolarse como la cantidad óptima que garantiza la mayor credibilidad del estudio con el menor número de casos posibles. Por supuesto que se deja entrever que dicha credibilidad será ampliamente superada si se estudian 36 efectivos más. Pero esto es más bien trivial, pues lo que en verdad importa es lo que hayan podido concluir los autores respecto al tamaño de muestra realmente empleado.

Ejemplo 1.4.d.

Un último ejemplo pone de manifiesto varias nociones que enfatizan de manera general el carácter ceremonial y predominantemente subjetivo del cálculo de n .

Se trata de un artículo correspondiente a una investigación realizada en España que permitió conocer la prevalencia de las formas clínicas de macroangiopatía diabética y sus factores de riesgo en el área de atención *Raval Sud*.

Dicho estudio se desarrolló entre septiembre de 1993 y marzo de 1994 y sus resultados fueron publicados en la revista *Atención Primaria* al año siguiente de su culminación bajo el título *Macroangiopatía en la diabetes tipo II. El estudio Raval Sud*.³¹

Allí figuran los valores necesarios para calcular el tamaño de muestra que supuestamente amerita el estudio:

- $N = 18\ 088$
- $P = 6,6\%$
- $Z = 1,96$
- $E = 0,025$ que equivale porcentualmente a 2,5%
- Tasa de no respuesta = 10%

Estos datos son expresados de la siguiente manera:

En las 18 088 historias clínicas de adultos actualmente abiertas en nuestro centro, se ha observado una prevalencia de diabetes mellitus registrada del 6,6% (1 193 casos), que es una de las más altas de España. Para cumplir el objetivo de nuestro trabajo, se diseñó un estudio descriptivo transversal entre todos los casos de diabetes mellitus tipo II registrados en nuestra área. Se predeterminó un tamaño muestral de 379 casos ($p=6,6\%$; $Z=1,96$; $e=0,025$). En previsión de los posibles casos perdidos se incrementó el tamaño muestral en un 10%, con lo que la muestra finalmente estudiada fue de 387 casos.

En otra zona del artículo se lee:

Pacientes u otros participantes. Muestreo aleatorio ($n=387$) de los pacientes diabéticos tipo II registrados en el área básica (prevalencia del 6,6%).

Se habría tratado entonces de un muestreo simple aleatorio sobre los 1 088 pacientes que recibían servicio en la referida área de atención. Consecuentemente; el *deff* es igual a la unidad.

Al sustituir cada uno de los valores expuestos en la fórmula [4.1], se obtiene que n se aproxima bastante a 379^{*}.

Por otra parte, resalta el marcado contraste que existe entre la exactitud con que ha sido declarado N , como elemento que sugiere extrema precisión y confiabilidad, y los grandes cambios de magnitud que puede experimentar el tamaño de muestra ante pequeñas y verosímiles fluctuaciones en los parámetros involucrados en el cómputo.

Es de notar la cercanía de N a 1 090 o incluso a 1 100. Cualquiera de estos valores utilizados, atendiendo al lugar que ocupa el parámetro en la fórmula producirían variaciones inapreciables en el tamaño final de la muestra. Sin embargo la influencia que ejerce la subjetividad de los datos seleccionados es bastante marcada, lo cual puede apreciarse con todo detalle realizando el siguiente “experimento”.

Supongamos que en vez de determinar previamente un error mínimo admisible igual a 2,5%, los autores hubieran decidido, con el resto de los parámetros constantes, utilizar un valor de E diferente aunque bastante cercano al original (por ejemplo $E=3,5\%$).

* El lector puede comprobar que el verdadero valor que arroja la fórmula es aproximadamente 371,2. Para obtener n igual a 379 se necesita redondear, al menos a dos espacios decimales, todos los resultados de los cálculos intermedios implicados en la fórmula, pero considerar cero la cantidad resultante de dividir n_0 entre N . Esta licencia matemática permite que n_0 y n sean iguales ($n_0 = 378,74 \approx 379 = n$). Al parecer algo semejante fue realizado por los autores de la investigación.

En tal caso, puede comprobarse fácilmente que la cantidad de individuos necesarios para el estudio, según la fórmula, descendería a casi la mitad de los incluidos originalmente (193 efectivos). Si por el contrario el grado de precisión deseado hubiese sido un poco menos de 2,5%; digamos 2, se necesitaría seleccionar entonces 213 individuos más (un total de 406).

De forma similar, con una confianza del 99%, en lugar del 95% (ahora el parámetro Z adoptaría un valor de 2,575); siempre que se mantengan las cifras originales para el resto de los datos, n se igualaría a 650,9 es decir a 651 efectivos, 272 menos que los que han sido estudiados.

Para el caso en que la confiabilidad adoptada sea de 99% y además el error del 2%, el tamaño de muestra resultante equivaldría entonces a 1 017,1 (1 017 pacientes). Si bajo esta última situación también se modifica el valor P y se toma 7% en vez del 6,6% preestablecido (ya sea porque dicho valor haya sido referido en otra literatura o estudio anterior; ya sea para no incurrir en el ridículo de anticipar una tasa al nivel de las décimas), entonces n se remontaría a 1 074 unidades de análisis.

Esta última cifra, resultante de insignificantes diferencias de apreciación respecto a tres de los parámetros primarios que definen la fórmula para n , representa casi 3 veces el valor inicial concebido para el estudio *Raval Sud*.

Es válido hacer notar que todo el ejercicio anterior se ha realizado sin involucrar en el cómputo el prefijado 10% de incremento muestral por el concepto de no respuesta. Si tal incremento hubiera sido considerado se necesitarían tamaños de muestra aun más grandes y por supuesto las diferencias serían mayores respecto al cálculo de *Mur* y colaboradores.

Esta reflexión ilustra la idea de que no existe un tamaño de muestra único y exclusivo para cada investigación, ni consideraciones que avalen una cantidad como la más pertinente para ser estudiada.

Como pudo observarse ligeras variaciones en los datos utilizados para el cómputo de n pueden llegar a producir drásticas diferencias en su valor práctico final. Dichas diferencias están sujetas, inexorablemente, a la visión individual y por demás subjetiva que cada investigador tenga del problema que estudia.

1.5. Sobre la frecuencia con que se emplea el ritual:

La metodología empleada en la presente tesis incluyó una revisión profunda de una zona de la literatura especializada en publicaciones científicas de carácter biomédico, clínico y sanitario donde con frecuencia ven la luz estudios basados en muestras. Si bien se han elegido los 4 ejemplos arriba analizados, corresponde señalar sin ánimo de aquilatar formalmente la magnitud del problema, que los equívocos concernientes al cómputo de n están muy extendidos. El objetivo de esta sección del trabajo se ha dirigido a revelar la existencia del conflicto. No obstante, si el lector estuviera interesado en revisar una buena cantidad de ejemplos adicionales, podrá hallarlos atendiendo a las siguientes acotaciones bibliográficas.³²⁻⁶²

Por último, queremos enfatizar que no solo la justificación que pueda hacerse en una investigación respecto al tamaño de muestra empleado resulta un procedimiento ceremonial, sino también la rigidez de pensamiento que supone considerar que la cantidad de unidades de análisis, para que aporte validez y carácter científico, tiene que ser determinada mediante malabares computacionales.

2. Procedimiento estadístico-metodológico: uso de la regresión logística para identificar factores de riesgo.

2.1. Bases teóricas generales:

Uno de los principales propósitos de la investigación epidemiológica consiste en descifrar o confirmar los mecanismos que rigen la complicada madeja del proceso salud-enfermedad en comunidades humanas. Por lo general, tal búsqueda se realiza no con intención de establecer las causas verdaderamente íntimas, en ocasiones insospechadas, de los fenómenos en estudio, sino más bien en un sentido funcional a descubrir asociaciones entre eventos donde los cambios implementados en aquel que se considera causante implique variaciones sobre su efecto.^{63,64}

Desde este enfoque, ha cobrado trascendental magnitud en los últimos años el estudio de los factores de riesgo; condiciones o características identificables en individuos y grupos poblacionales susceptibles de ser medidas y modificadas y que, a pesar de que pudieran no intervenir como responsables directas de una determinada afección, incrementan la posibilidad de que el verdadero agente causal de esta actúe y la desencadene.^{14,64,65}

La separación conceptual de las categorías “causa” y “factor de riesgo” resulta en extremo polémica. La tendencia más aceptada es considerar como causal la relación entre factores de riesgo y enfermedades;¹⁴ no obstante las exigencias metodológicas para la corroboración de una u otro en el marco de la epidemiología son en principio las mismas. Un factor de riesgo para ser catalogado como tal debe cumplir con todos los criterios necesarios para considerar que su relación con un daño a la salud es causal. Algunos de estos requisitos, como la plausibilidad biológica, constituyen premisas de elemental observancia; pero aquellos que resultan verdaderamente insoslayables desde el punto de vista metodológico son la congruencia temporal, la ausencia de sesgos y la presencia de asociación estadística de carácter intrínseco entre el factor y la variable que registra la enfermedad.

En general el empleo de adecuados diseños de investigación garantiza el cumplimiento de la coherencia temporal y la ausencia de sesgos, mientras que la comprobación de asociaciones exclusivas se consigue básicamente mediante procedimientos de índole estadística cuya finalidad es el control de aquellas variables que a juicio del investigador pudieran estar influyendo sobre la asociación de interés, distorsionándola. Estas variables son conocidas en el argot investigativo como factores de confusión.^{10,14,65-68}

Dentro del marco epidemiológico, donde la experimentación, paradigma explicativo de las relaciones causales, resulta limitada por aspectos de orden ético y económico, los estudios observacionales analíticos son frecuentemente implementados para identificar posibles causas y/o factores de riesgo de enfermedades. En estos casos los estudios de cohorte devienen como sucedáneos de los experimentos con el inconveniente de requerir, casi siempre, de un seguimiento muy extenso y ser además extremadamente caros. Los comúnmente llamados *estudios de casos y controles* y los transversales o de corte resultan por tanto los de mayor factibilidad práctica. Estas investigaciones, de tipo observacional, implican diseños que no permiten la manipulación de la exposición sino que determinan el estudio de los hechos según hayan ocurrido de manera natural en la población. Por su conducto los investigadores pueden evaluar varios factores y, aunque es importante señalar que no bastan para demostrar causalidad propiamente dicha, sí proporcionan evidencias de asociación causal que permiten, sobre todo, llevar a cabo intervenciones de salud para reducir la exposición al factor de riesgo con la consiguiente reducción de la incidencia de sus efectos.

Pasar por alto el posible efecto confusor de una variable malogra directamente los resultados de una investigación explicativa, de modo que un elemento singular en estos estudios radica en distinguir aquellos factores que son realmente influyentes sobre el desenlace de aquellos que parecen serlo solo por el hecho de estar simultáneamente asociados con la verdadera causa y el desenlace.

Precisamente el principal desafío de los estudios no experimentales para probar hipótesis causales radica en controlar el efecto de dichas variables o factores de confusión.⁶⁵⁻⁷⁰

Durante la fase de análisis el investigador puede recurrir a métodos de naturaleza estadística como la postestratificación y posterior aplicación de las técnicas propuestas por *Mantel* y *Haenszel* o el ajuste de los datos a modelos matemáticos multivariados dentro de los cuales la regresión logística tiene particular relevancia. Esta es una de las técnicas inferenciales actualmente más empleada debido a que propicia, entre otras bondades, una mayor eficiencia al estimar los efectos de varios factores simultáneamente.^{67,68,71} Sin embargo el empleo ritual de un procedimiento que implementa la regresión logística en estudios observacionales orientados a identificar factores de riesgo ha sido sumamente extendido en la producción científica contemporánea, hasta el punto de convertirse en una verdadera ceremonia metodológica. Esta discurre, a grandes rasgos, de la siguiente manera.¹⁴

2.2. Empleo ceremonial del procedimiento.

- 1)** Se comienza por identificar y seleccionar un conjunto de variables consideradas como factores de riesgo de la enfermedad que se estudia. Para ello se acude a criterios de expertos, estudios previos, consideraciones teóricas o búsquedas bibliográficas y en Internet.
- 2)** Se implementa un estudio bivariado donde se valora la asociación entre cada una de las variables seleccionadas y la variable dicotómica de respuesta (típicamente una enfermedad o una complicación). En este caso se utilizan pruebas de hipótesis específicas en dependencia del tipo de variables estudiadas.

- 3) Seguidamente se ajusta un modelo multivariado mediante regresión logística con aquellas variables independientes para las que se obtuvo significación estadística en el paso anterior.
- 4) Se Identifican las variables independientes en las que se determinó asociación con la variable de respuesta tras efectuar el ajuste. En este punto existen dos variantes básicas:
 - 4a) La identificación de aquellas variables que mantienen su carácter significativo tras haber realizado la modelación.
 - 4b) La aplicación de un procedimiento algorítmico iterativo de selección de variables paso a paso (*stepwise regression*) para identificar aquellas que habrán de permanecer en el modelo. Este procedimiento puede efectuarse ya sea incorporando consecutivamente variables (*forward selection*) o ir eliminándolas de un modelo inicial conformado por todas las que se incluyeron en el análisis (*backward elimination*).
- 5) Por último se califican como factores de riesgo a aquellas variables independientes que mantuvieron la significación estadística o las que se hayan conservado en el modelo final si se hubiera apelado a la regresión paso a paso.

Debe puntualizarse que en la práctica este proceder puede tener ciertas variaciones. En ocasiones puede omitirse el estudio bivariado y se pasa directamente al multivariable. A veces, al concluir, no se habla de “factores de riesgo” como tal sino que se menciona solamente la existencia o no de asociación entre las covariables y la variable dependiente.^{14,72,73}

2.3. Análisis de su improcedencia:

1) Identificación y selección de un conjunto de posibles factores de riesgo:

La tarea de reconocer la mayoría de los factores de riesgo para determinada afección, aunque en esencia no tiene nada de objetable es mucho más típica de ambientes de menor desarrollo que de espacios académicos consagrados.

Resulta relevante en relación a este aspecto señalar la importante distinción entre factores y marcadores de riesgo. Estos últimos constituyen características inherentes al individuo, por lo común estrechamente ligadas a su biología e inmodificables independientemente de cualquier voluntad (sexo, color de la piel, grupo sanguíneo, etc.). Su invariabilidad anula que se puedan tomar conductas directas para contrarrestar sus efectos; solo sirven para señalar grupos más vulnerables. Su presencia puede ser determinada como confusora en determinados casos, pero nunca deben ser interpretados como factores de riesgo.^{14,73} La irreflexiva selección de variables sin considerar esta diferenciación, verosímilmente traerá consigo la inclusión en el análisis de marcadores de riesgo para ser identificados como factores de tal condición.

2) Valorar la existencia de asociación entre cada variable independiente, presunto factor de riesgo y la enfermedad a través de un estudio bivariado mediante pruebas de significación:

La distorsión introducida por un factor de confusión puede conducir tanto a la observación de un efecto que en realidad no existe o a la apreciación de manera desmedida de cierta asociación real como, por el contrario, a que un efecto se observe indebidamente atenuado e incluso con un sentido inverso con el que realmente está actuando. Por tanto, determinar cuáles de las covariables están relacionadas o no con la variable dependiente mediante un análisis bivariado, que obvia la rica interacción multifactorial implícita en la realidad, puede conllevar a que sean desechadas aquellas sobre las que actúan factores ajenos, disminuyendo o invirtiendo su contribución a la ocurrencia del efecto.^{67,71} En otras palabras: variables que resultaron no significativas en el análisis bivariado pueden perfectamente ser contribuyentes a la manifestación del efecto y mostrarse significativas bajo análisis multivariado.

Por otro lado, en la clásica relación causal es mucho mayor el peso que ejerce la magnitud de la asociación entre las variables que el carácter significativo de la misma. La significación en este caso resulta irrelevante y esto se debe fundamentalmente a que puede alcanzarse con solo incrementar lo suficiente el número de efectivos estudiados, endeblez intrínseca de las pruebas de hipótesis que ha sido reiteradamente apuntada, especialmente en los últimos años.^{2,3,10,71-78}

La significación estadística de unos u otros factores involucrados en el estudio bivariado y por tanto su tránsito al próximo escaño del procedimiento depende, además de la magnitud de la asociación de estos con la variable de respuesta, de un elemento que resulta totalmente ajeno a la realidad: los recursos disponibles y, por tanto, del tamaño de muestra empleado.⁷⁹

Lo dicho supone que un cribado inicial de variables utilizando como elemento discriminante la significación de las pruebas de hipótesis es erróneo y muchas veces contraproducente.

3) La identificación de aquellas variables que mantienen su carácter significativo tras haber realizado la modelación.

El ajuste de una ecuación logística para controlar variables confusoras e identificar factores de riesgo es en esencia válido si se desarrolla bajo un análisis lógico bien estructurado. Sin embargo, en el método que se ha descrito, las inconsistencias de este paso vienen condicionadas, en primer lugar, por la inclusión en el modelo de todas las variables que son consideradas posibles factores de riesgo, cuando lo razonable sería haber procedido con cada una por separado de la siguiente manera:¹⁴

- Identificar con claridad la variable cuya condición de factor de riesgo se quiere evaluar.
- Incluir esta como una de las variables independientes del modelo.
- Valorar teóricamente uno por uno aquellos factores que pudieran ser candidatos como confusores corroborando además que cumplen con todas las premisas necesarias para ello.

- Incluir los que estén en ese caso como covariables del modelo con el fin de controlar su efecto confusor sobre la asociación de interés.

Se enfatiza que el control de variables en el contexto observacional no debe efectuarse si no es para evaluar hipótesis de investigación que hayan sido preconcebidas nítidamente y fundamentadas con anticipación, donde el potencial papel confusor de las variables que se controlan debe contar con un respaldo teórico y un fundamento lógico. Cuando se identifican todas aquellas variables cuyos coeficientes de regresión son significativamente distintos de cero se le asigna al algoritmo toda la responsabilidad de decidir cuáles son aquellos factores que verdaderamente implican riesgo.

El ajuste de los datos al modelo logístico se realiza indistintamente de cuáles variables el investigador considere como responsables del efecto y cuáles como factores de confusión por lo que esta acción ignora los siguientes conflictos:¹⁴

- El posible control de factores que no son confusores (un factor de riesgo no necesariamente es un confusor para la asociación de la enfermedad con otro factor de riesgo).
- La frecuente inclusión de marcadores de riesgo como si fueran factores de riesgo.
- La sobrevaloración de la significación estadística en la evaluación de asociaciones causales donde, como ya se explicó anteriormente, la verdadera valía recae sobre la fuerza de la misma entre las variables estudiadas.
- Las ya mencionadas limitaciones de las pruebas de hipótesis.

4) La aplicación de un procedimiento algorítmico de selección de variables paso a paso para identificar las variables que habrán de permanecer en el modelo:

La aplicación de la regresión paso a paso como vía para la detección de factores de riesgo es la situación en que este ritual se lleva a su máxima expresión.

En la práctica no es inusual que se aplique esta metodología para determinar qué variables han de “quedarse” en calidad de factores de riesgo detectados y cuáles ameritan ser desechadas. Autores de gran prestigio, incluso, sugieren que se ajuste un modelo de regresión múltiple y de inmediato se aplique un procedimiento algorítmico que elija por el investigador aquellas variables cuya asociación con el desenlace le atribuye el carácter de factor de riesgo.⁸⁰

Tal camino pasa por alto totalmente que la lógica subyacente en los procedimientos algorítmicos de selección consiste en conservar las variables independientes que contienen información relevante y prescindir de aquellas que resulten redundantes respecto a las que quedaron, elemento muy distante de dar una explicación de causalidad. El procesamiento transita simplemente siguiendo algoritmos estadísticos programables en los que, una vez elegido el conjunto inicial de variables, no intervienen los imprescindibles juicios teóricos de los investigadores.¹⁴ Como suele decirse, “los números no saben de dónde vienen”.

La regresión paso a paso, además, produce resultados discrepantes cuando se emplean sus diferentes alternativas. La bibliografía especializada en relación a las estrategias de selección de variables, no aporta ningún argumento convincente que demuestre que incorporar variables (*forward selection*) sea mejor que ir eliminándolas (*backward elimination*) consecutivamente del modelo o viceversa; y ya se ha dicho que los resultados con frecuencia no son los mismos.^{67,72,80,81}

Tal circunstancia constituye un primer indicio de la posible improcedencia de la regresión paso a paso para explicar la realidad, puesto que es legítimo sospechar que las variables que conforman el modelo final podrían ser diferentes en dependencia del procedimiento empleado y determinado por la subjetividad del investigador. En pocas palabras: la subselección algorítmica no podrá ser capaz de identificar las variables que expliquen el proceso modelado.

Merece ser aclarado el uso que pueda dársele a este procedimiento como instrumento de predicción, para vaticinar el comportamiento de cierto fenómeno (variable de respuesta) en función de un conjunto de variables predictivas.⁸¹ En este contexto, quizás la regresión paso a paso no es en principio objetable. Una variable puede hacer importante aportación a los efectos de predecir, aunque en sí misma no sea un factor de riesgo ni tenga papel causal y un factor puede servir para la predicción sin que eso lo convierta en un agente causal. Por ejemplo, los marcadores de riesgo suelen tener un valor predictivo, pero no pueden considerarse factores de riesgo.^{14,66,72}

5) Calificar como factores de riesgo a aquellas variables independientes que mantuvieron la significación estadística o las que se mantuvieron en el modelo final en el caso de aplicar la regresión paso a paso:

Esta interpretación con la que se cierra el procedimiento produce un franco divorcio de este con el objetivo que animaba su empleo.

Es frecuente que, para *curarse en salud*, se concluya especificando simplemente cuáles variables *están asociadas* con el efecto en estudio. Con este recurso, más semántico que sustantivo, se elude el compromiso de valorar si las mismas son o no causantes del desenlace. Transmutar a modo de conclusión este pronunciamiento carece de sentido, pues la asociación de causalidad debe ser el resultado de un pensamiento o razonamiento lógicos, procesos intelectuales cualitativamente superiores a la simple cuantificación de si hay o no asociación.

Es importante recalcar que la medición de una asociación nunca tiene interés en sí mismo y carece de un sentido claro, excepto que se inserte en el contexto de una conjetura causal.¹⁴

Para concluir, procede señalar que, a diferencia del resto de los procedimientos ceremoniales presentados en este estudio, cuya práctica a lo sumo es estéril, la aplicación de la regresión logística para identificar factores de riesgo puede llegar a ser profundamente contraproducente pues puede conducir a que se arraiguen convicciones falsas acerca de cómo es la realidad.

2.4. Ejemplos del empleo ceremonial:

Ejemplo 2.4.a.

En un estudio orientado a determinar la prevalencia de la enfermedad arterial periférica en pacientes con diabetes tipo 2 del Centro de Salud de Ronda Prim de Mataró; Cataluña, España, los investigadores decidieron, además, identificar factores de riesgo para dicha dolencia.⁸² Según la clasifican sus propios autores se trató de una investigación de índole descriptiva. Sin embargo, el enfoque con que fue abordado el segundo de los objetivos, resultó esencialmente explicativo.

La metodología empleada puede sintetizarse así:

Participaron en el estudio los 329 diabéticos tipo 2 registrados en el centro, a los que se le realizó, para documentar la presencia de daño vascular, una anamnesis que incluía: edad, sexo, tiempo de evolución y tratamiento de la diabetes, antecedentes personales de accidentes cerebrovasculares, cardiopatía isquémica, tabaquismo, medicamentos consumidos diariamente, presencia de precordalgias y dolor en las extremidades inferiores. Se efectuó, además, un examen físico general y de los pies, buscando signos de insuficiencia arterial. El examen físico se complementó con un electrocardiograma y una batería de estudios sanguíneos que incluía glicemia, hemoglobina glicosilada A1c, colesterol, creatinina y triglicéridos. Finalmente se investigaron los antecedentes de hipertensión arterial y dislipemia así como el grado de control de ambas y de la glucemia.

A todos los pacientes, excepto en los que ya constaba el diagnóstico de arteriopatía periférica en la historia clínica, se les practicó la determinación de la tensión arterial sistólica (TAS) en brazos y tobillos. Posteriormente se calculó el índice tobillo-brazo (ITB) de cada extremidad inferior, medida que determinaría la incidencia o no de una afección arterial periférica. Los datos se introdujeron en una base de datos informatizada para su posterior explotación. En el análisis correspondiente se utilizaron estadígrafos descriptivos; la comparación de proporciones se llevó a ejecución mediante la prueba de ji-cuadrado mientras que las medias se cotejaron utilizando la prueba *t* de *Student*. El análisis multivariante se realizó mediante regresión logística.

Una lectura detallada del texto anterior permite reconocer que, a los efectos de lo que nos interesa, lo expuesto no es susceptible de un análisis demasiado profundo, ya que no se han definido con claridad las variables independientes de la regresión logística ni la escala con que estas han sido operacionalizadas. Es evidente, sin embargo, que la variable de respuesta es aquella que registra de manera dicotómica la presencia o no de enfermedad arterial periférica.

Con relación a los resultados puede leerse de manera textual en el artículo:

Se estudiaron un total de 289 pacientes, 157 mujeres (54,3%) y 132 varones (45,7%), con una edad media de 65,3 años ($\pm 10,8$ DE).

El análisis mediante regresión logística mostró una relación significativa entre arteriopatía periférica y tabaquismo (analizando juntamente fumadores y ex fumadores), historia de hipertensión arterial, edad y tipo de tratamiento de la diabetes mellitus.

Los resultados se muestran en una tabla similar a la siguiente:

Tabla 2: Análisis multivariante de la RL para la variable dependiente arteriopatía periférica.

Variable	β	DE*	Wald	df	Sig	R	Exp(β)
Tabaco	1,5741	0,3991	15,5546	1	0,0001	0,2365	4,8263
Edad	0,0822	0,0199	17,0856	1	0,0000	0,2495	1,0857
HTA	0,8711	0,3773	5,3296	1	0,021	0,1173	2,3896
Tratamiento	0,5056	0,2007	6,3456	1	0,0118	0,1339	1,658
Constant.	-9,1866	1,6027	32,8561	1	0,0000		

Fuente: Revista Atención Primaria 1998. Volumen 22, Número 1

En primer lugar, es cuestionable que los autores de este trabajo se hayan dado a la tarea de identificar mecánicamente, dentro de un conjunto de cofactores, aquellos que guardan una asociación de tipo estadística con la variable dependiente para conferirle la condición de factor de riesgo. Como ya se ha explicado, lo verdaderamente útil y procedente desde el punto de vista metodológico es determinar en qué medida una variable con todos los requisitos necesarios para ser considerada como un factor de riesgo y cuyo comportamiento como tal no se conoce a ciencia cierta, se asocia a la ocurrencia de una enfermedad o afección en general, inclusive al ser controladas mediante la regresión otras variables cuyo papel confusor en la asociación de interés es reconocido.¹⁴

Puede apreciarse que muchas de las variables que supuestamente han sido incluidas como cofactores en la regresión (sirvan de ejemplo el hábito de fumar y el tiempo de evolución de la diabetes) son factores de riesgo cardiovascular bien establecidos y universalmente aceptado como tales.⁸³⁻⁸⁵ Es decir, tras este proceso se ha llegado a lo que ya es hartamente conocido.

Nótese también que en la tabla solo se muestran las variables que resultaron significativas tras la regresión logística. De aquí fácilmente se deriva que lo determinante en su categorización como factores de riesgo fue precisamente el carácter significativo y no la intensidad de su asociación con el desenlace, expresada en forma de *odds ratio* como es metodológicamente adecuado.¹⁴ Para reafirmar esta observación, véase como los *odds ratio* son expresados de manera puntual y no por intervalos de confianza como deben ser expresados cuando se trabaja con muestras.^{66,67}

Si leemos lo que concluyen los autores y examinamos la Tabla 2, observaremos que solo la hipertensión, el tabaquismo, el tiempo de evolución de la diabetes y el tratamiento llevado por el paciente son las variables que se consideran verdaderos factores de riesgo en la incidencia de arterioesclerosis periférica pues son precisamente las que presentan un coeficiente β significativamente distinto de cero.

Resulta asombroso que, por ejemplo, los trastornos lipídicos, no figuren en esta lista cuando hoy día se reconocen como los factores de mayor fuerza en la arterioesclerosis.⁸³⁻⁸⁹

Al comparar los valores p asociados a cada variable con su respectivo *odds ratio* puede observarse que para algunas de ellas a pesar de que p resultó muy pequeño, los correspondientes *odds ratio* son muy próximos a la unidad. Este es el caso de la variable-*tratamiento* con valor p igual a 0,0118 y *odds ratio* de solo 1,6580.

Como se ha planteado, se consideran valores de *odds ratio* superiores a 2 o incluso a 3 para atribuir importancia al grado de asociación entre una variable posible factor de riesgo y una enfermedad.¹⁴ En la tabla los valores de *odds ratio* son superiores a 2 solo para la variable *tabaco* e *HTA*.

Ejemplo 2.4.b.

En el artículo “Hipoparatiroidismo e hipocalcemia en el postoperatorio de la cirugía del bocio multinodular. Estudio multivariante de los factores de riesgo”, que apareció en la Revista Medicina Clínica en el año 2004,⁹⁰ se declara el propósito de analizar los factores de riesgo para el desarrollo de hipoparatiroidismo transitorio y definitivo tras la operación sobre la glándula tiroidea afectada.

Se revisaron de manera retrospectiva 672 casos de pacientes intervenidos quirúrgicamente de dicha entidad entre los años 1970 y 1999 en el servicio de Cirugía General y del Aparato Digestivo perteneciente al Hospital Universitario Virgen de la Arrixaca de la localidad de Murcia.

En relación al análisis estadístico se comunica en el trabajo:

Las variables analizadas para valorar los factores de riesgo para desarrollar hipoparatiroidismo fueron la edad, el sexo, la cirugía previa, la situación clínica asintomática, el hipertiroidismo, la presencia de síntomas compresivos, el componente intratorácico del bocio, la gradación del bocio, la duración de la cirugía, la experiencia del cirujano, la técnica quirúrgica, la identificación de las glándulas paratiroides y la presencia de un carcinoma tiroideo asociado.

Se aplicó la prueba de la X^2 cumplimentada con el análisis de residuos y el test de la *t de Student*. Para la determinación y evaluación de riesgos múltiples se realizó un análisis de regresión logística utilizando las variables que en el análisis bivalente dieron una asociación significativa. Las diferencias se consideraron significativas para unos valores de p inferiores a 0,05.

Puede apreciarse que en esta investigación, a diferencia de la anteriormente expuesta, los autores incorporan el filtrado de variables, previo a la implementación de la regresión logística.

Para el caso del hipoparatiroidismo desarrollado durante el postoperatorio del bocio, los resultados tras el cribado inicial son expuestos de esta manera:

Al estudiar a los pacientes que durante el postoperatorio presentaron hipoparatiroidismo, se observó, en el análisis bivalente, que las variables que se asociaban significativamente a su presencia eran el hipertiroidismo ($p = 0,0370$) y la técnica quirúrgica ($p < 0,00001$)...

Estos datos se expusieron en un cuadro que, además, reúne todas las variables analizadas y su significación estadística según el valor p encontrado en cada prueba de hipótesis.

Su contenido se muestra en la Tabla 3.

Posteriormente tras el análisis multivariado y refiriéndose a las variables que exhibían diferencias significativas en el primer análisis, se manifiesta:

En el análisis multivariante, las 2 variables persistieron como factores de riesgo independientes para el desarrollo de hipoparatiroidismo tras cirugía del bocio múltiple.

Los resultados fueron expuestos en un cuadro similar a la Tabla 4.

Como veremos, el procedimiento es totalmente inconsistente y poco confiable, pues en principio (y es ésta es una de las limitantes de mayor peso) el paso de los factores a través del “filtro bivariado” está influenciado por el tamaño de muestra del estudio.

Supongamos que en vez de 672 individuos se hayan escogido para el estudio 2 015, cantidad que indudablemente conferiría más potencia a las pruebas de hipótesis que serán implementadas.

Podemos encontrar que si se mantuvieran los porcentajes de hipoparatiroidismo antes hallados, variables que anteriormente mostraban independencia respecto al desarrollo o no de hipoparatiroidismo durante el postoperatorio, ahora estarán asociadas estadísticamente a este, o lo que es lo mismo, resultarán significativas.

En este escenario tomemos como ejemplo uno de los factores que se encontraba inicialmente en igual caso: *la situación clínica asintomática*.

Auxiliándose de la Tabla 5 el lector podrá comprobar que adoptando la siguiente distribución (nótese que se han mantenido los mismos porcentajes para cada una de las categorías según reflejaron Ríos y colaboradores en la Tabla 3), la variable en cuestión se asocia estadísticamente con la presencia o no de hipoparatiroidismo (*ji cuadrado con corrección de Yates* = 3,8955; con un valor p igual a 0,0484).

Se ha tomado de ejemplo esta variable para demostrar, entre otros aspectos, cómo el tamaño de muestra puede afectar directamente los resultados de las pruebas de hipótesis en general.

Es cierto que el tamaño muestral no fue 2 015 sino 672, pero el simple hecho de que pueden encontrarse datos hipotéticos que conducen a resultados discordantes a los encontrados sugiere el claro peligro que entraña dejar en manos de una prueba de significación estadística la apreciación sobre lo que puede o no estar influyendo.

Tabla 3: Variables asociadas con el hipoparatiroidismo postoperatorio. Análisis bivariante.

Variable	Hipoparatiroidismo postoperatorio (n = 75)	Sin hipoparatiroidismo postoperatorio (n = 596)	p
Edad media (DE), años	51 (12)	48 (15)	0,0829
Sexo			
Varones (n = 55)	5 (7%)	50 (8%)	0,6109
Mujeres (n = 617)	66 (93%)	547 (92%)	
Cirugía previa			
No (n=626)	74 (98,7%)	552 (92,5%)	0,0549
Sí (n = 46)	1 (1,3%)	45 (7,5%)	
Asintomático			
No (n=274)	35 (47%)	239 (40%)	0,2705
Sí (n =398)	40 (53%)	358 (60%)	
Hipertiroidismo			
No (n=560)	52 (69%)	23 (31%)	0,0370
Sí (n =112)	508 (85%)	89 (15%)	
Síndromes compresivos			
No (n=515)	59 (79%)	456 (76%)	0,6594
Sí (n =157)	16 (21%)	141 (24%)	
Componente intratorácico			
No (n=425)	44 (59%)	381 (64%)	0,383
Sí (n =247)	31 (41%)	216 (36%)	
Grado del bocio			
0-I (n = 105)	5 (7%)	100 (16%)	0,537
II (n = 410)	54 (72%)	356 (60%)	
III (n = 157)	16 (21%)	141 (24%)	
Duración de la cirugía (min)	107 (33)	103 (36)	0,9856
Experiencia del cirujano			
Endocrino (n = 532)	63 (84%)	465 (78%)	0,1214
No endocrino (n = 140)	8 (16%)	132 (22%)	
Técnica quirúrgica			
TT (n = 413)	60 (80%)	353 (59%)	< 0,00001
TS (n = 43)	3 (4%)	40 (6,7%)	
Dunhill (n = 54)	5 (6,7%)	49 (8,2%)	
Completar tiroidectomía (n = 24)	6 (8%)	18 (3%)	
Hemitiroidectomía (n = 113)	1 (1,3%)	112 (18,8%)	
HSU (n = 25)	0 (0%)	25 (4,2%)	
Identificación de paratiroides			
Ninguna (n = 176)	18 (24%)	158 (26%)	0,1501
1 (n = 15)	1 (1%)	14 (2%)	
2 (n = 160)	14 (19%)	146 (24%)	
3 (n = 57)	12 (16%)	45 (8%)	
4 (n = 264)	31 (41%)	233 (39%)	
Carcinoma tiroideo asociado			
No (n = 613)	70 (93%)	543 (91%)	0,5326
Sí (n = 59)	5 (7%)	54 (9%)	

DE: desviación estándar; TT: tiroidectomía total; TS: tiroidectomía subtotal bilateral; HSU: hemitiroidectomía subtotal bilateral.

Fuente: Revista Medicina Clínica 2004. Volumen 122. Número 10.

Tabla 4: Variables asociadas con el hipoparatiroidismo postoperatorio. Análisis multivariante.

Variable	Coefficiente de regresión β	Error estándar	Odds ratio (IC)	p
Hipertiroidismo				
No (n=560)			1	0,0401
Sí (n=112)	-0,28056	0,1425	1,76 (2,33-1,32)	
Técnica quirúrgica				
TT (n=413)			1	0,0001
TS (n=43)	-1,3962	1,144	0,77 (7,6-0,08)	
Dunhill (n = 54)	-1,7244	1,083	0,62 (5,4-0,07)	
Completar tiroidectomía (n=24)	0,67921	0,1408	1,67 (2,22-1,26)	
Hemitiroidectomía (n=113)	-0,69425	0,00001	0,20 (0,20-0,20)	
HSU (n=25)	-5,46135	0,04102	0,006 (0,007-0,006)	

IC: intervalo de confianza; TT: tiroidectomía total; TS: tiroidectomía subtotal bilateral; HSU: hemitiroidectomía subtotal unilateral.

Fuente: Revista Medicina Clínica 2004. Volumen 122. Número 10.

Tabla 5: Pacientes según situación clínica asintomática e hipoparatiroidismo postoperatorio.

Asintomático	Con hipoparatiroidismo postoperatorio		Sin hipoparatiroidismo postoperatorio		Total	
	#	%	#	%	#	%
	No	106	47	716	40	822
Sí	119	53	1074	60	1193	59
Total	225	100	1790	100	2015	100

Una vez que se hayan insertado otras variables en la regresión logística, posiblemente la técnica arrojará resultados diferentes en cuanto al nivel de asociación del resto de las covariables, máxime si la primera actúa como confusora en alguna o algunas de las restantes relaciones.

En una situación como ésta que hemos prefabricado, la variable *situación clínica asintomática* habría de ser incorporada como un cofactor más en la ecuación logística.

Consecuentemente los autores podrán, caso de que el carácter significativo persistiera una vez obtenido los resultados del análisis multivariado, considerar la presencia de síntomas de hipoparatiroidismo como un factor de riesgo individual de la enfermedad en el postoperatorio del bocio multinodular. Lo interesante de todo lo anterior es cómo un supuesto factor de riesgo que de inicio fue desechado en el análisis bivariado adopta nueva connotación con solo variar la cantidad de sujetos a estudiar.

Por otra parte, resulta obvio que en este estudio se ha prescindido totalmente del concepto "factor de riesgo" al incluir como posible candidato una variable que recoge dicotómicamente la condición de presentar o no síntomas específicos de un proceso mórbido. Los síntomas de una enfermedad no modifican en absoluto la probabilidad de su aparición, a lo sumo propiciarán el diagnóstico pero no su desarrollo; son consecuencias, no causas, de ahí que también el ejemplo ha servido para ilustrar la naturaleza meramente ceremonial de insertar determinadas variables en análisis que entrañen causalidad.

Finalmente el ritual se repite en este estudio, pues similar metodología se implementó para la detección de los factores de riesgo del hipoparatiroidismo, pero ahora de tipo definitivo, tras la resección quirúrgica del bocio multinodular.

En este caso la variable que resultó significativa en el tamizaje bivariante fue solamente el hipertiroidismo ($p=0,0109$), que se mantuvo como factor de riesgo independiente en el análisis multivariante ($RR=2,28$) según el texto.

De ahí que los autores planteen a manera de conclusión algo que puede ser fácilmente cuestionable:

En nuestra serie el principal factor de riesgo independiente de lesión paratiroidea es la cirugía por bocio multinodular tóxico, tanto para los hipoparatiroidismos postoperatorios ($RR = 1,7$) como para los definitivos ($RR = 2,3$). En conclusión, el principal factor de riesgo de hipoparatiroidismo definitivo en la cirugía del bocio multinodular es la presencia de hipertiroidismo.

La conclusión aparece de manera mecánica y, como ha podido observarse, se ha sustentado en un único pilar: el algoritmo matemático-estadístico por el que fluye el análisis de la regresión logística.

Ejemplo 2.4.c.

Como se había anticipado, una de las variantes posibles del ceremonial transita a través de un método algorítmico de selección de variables conocido como *regresión paso a paso* e implementado en base a la ecuación logística.

El trabajo de los investigadores del Servicio de Medicina Preventiva perteneciente al Hospital Universitario La Paz en Madrid en pos de conocer los factores de riesgo que influyen en la aparición de la infección intrahospitalaria ilustra este *modus operandi*.

Se llevó adelante un estudio de cohorte que incluyó los 551 pacientes ingresados durante 48 horas o más en la Unidad de Cuidados Intensivos a lo largo del año 1992. Bajo el título "Infección nosocomial en una unidad de cuidados intensivos: análisis multivariante de los factores de riesgo", los resultados del trabajo vieron la luz en 1997 en un número de Medicina Clínica.⁹¹

Según la metodología descrita, los autores, consideraron de utilidad estudiar una gran cantidad de variables como candidatas a ser consideradas factores de riesgo. La operacionalización fue realizada nada menos que para 68 variables algunas de las cuales se mencionan a continuación:

Las variables de seguimiento estudiadas fueron edad, sexo, presencia de infección nosocomial y tipo: neumonía nosocomial, bacteriemia nosocomial, infección del tracto urinario nosocomial u

otras, factores intrínsecos: desnutrición, obesidad, diabetes mellitus, APACHE II[•], insuficiencia renal, acidosis metabólica/respiratoria, neoplasia, alteración del nivel de conciencia, infección extrahospitalaria asociada, leucopenia, anemia, inmunodeficiencia, inmunodepresión, tabaquismo, alcoholismo, presencia de broncoaspiración, cirugía, tubo orotraqueal, traqueostomía, ventilación mecánica, sonda nasogástrica, número de catéteres intravenosos, antibióticos, antiácidos, descontaminación orofaríngea, traqueal o intestinal.

Conforme el texto, tal “cajón de sastre” fue manejado del modo siguiente:

Posteriormente hemos llevado a cabo un análisis simple o bivariante para demostrar asociaciones, empleando el test de la χ^2 para variables cualitativas; la magnitud de la asociación se determinó mediante el riesgo relativo (RR) y la precisión de la estimación del efecto se midió a través del intervalo de confianza (IC) del 95% del RR. Empleamos la prueba de la t de Student para demostrar diferencias entre medias de variables cuantitativas independientes.

Y luego:

Para controlar factores de confusión realizamos un análisis multivariante mediante regresión logística que nos permitió obtener un modelo matemático que relaciona la aparición de infección nosocomial con los distintos factores de riesgo y medidas preventivas. Realizamos una regresión logística stepwise, esto es, «paso a paso», en la que en cada etapa o «paso» se añaden o se sustraen variables según un criterio de significación estadística y buscando una mejora en el ajuste del modelo. La estrategia de entrada de variables en el modelo vino determinada por su significación en el análisis bivariante junto con la importancia o interés de una variable a pesar de no ser claramente significativa en dicho análisis.

Las variables que han resultado significativas tras el análisis bivariado se resumen en el artículo en sendos cuadros estadísticos, uno para las que han sido medidas mediante una escala cualitativa y otro para las de tipo cuantitativo continuo (escala de razón). De todas las variables *testeadas*, 44 resultaron significativas en el marco bivariado, lo que supuso su inclusión en la regresión.

El hallazgo final se enuncia así:

En conclusión, hemos encontrado que en el desarrollo de las infecciones nosocomiales

• Es un sistema de puntuación destinado a clasificar a los pacientes en función de su gravedad. APACHE II se encuadra dentro de los denominados índices de gravedad de tipo fisiológico, basado en que la gravedad de un proceso patológico agudo puede medirse cuantificando un determinado número de variables fisiológicas.

influyen factores de riesgo tanto intrínsecos (fundamentalmente la alteración del nivel de conciencia) como extrínsecos (instrumentaciones y medicaciones), así como el tiempo de permanencia en el hospital hasta la aparición de la primera infección nosocomial.

Los autores hacen esta inferencia basándose en el modelo obtenido mediante la regresión paso a paso (Tabla 6) que incluye precisamente las variables anteriormente declaradas.

Tabla 6: Resultados de la regresión logística.

<i>Variables</i>	<i>Mejora de la p</i>	<i>Coefficiente (β)</i>	<i>Error estándar de β</i>	<i>Test de Wald</i>	<i>Odds ratio</i>	<i>IC de los odds ratio</i>
Presencia de seis o más instrumentaciones	0,000	1,55	0,27	5,62	4,75	2,75-8,19
Estancia de más de 10 días desde el ingreso en el hospital hasta el desarrollo de la primera IN	0,000	1,42	0,24	5,92	4,17	2,60-6,70
Relajantes musculares	0,000	0,81	0,23	3,49	2,25	1,43-3,55
Sonda nasogástrica	0,005	0,78	0,28	2,75	2,19	1,25-3,84
Alteración del nivel de conciencia	0,001	0,7	0,23	3,05	2,02	1,28-3,17

IC: intervalo de confianza; IN: infección nosocomial.

Fuente: Revista Medicina Clínica. Volumen 108. Número 11

De esta suerte, informan:

Las variables independientes que entraron en el modelo de regresión logística construido para la infección nosocomial son las siguientes: número de instrumentaciones (cualificada como presencia de seis o más instrumentaciones frente al resto), número de días de hospitalización previos a la aparición de la primera IN (cualificada como la duración mayor de 10 días frente al resto), empleo de relajantes musculares, presencia de sonda nasogástrica y alteración del nivel de conciencia.

Resulta difícil imaginar un ejemplo más claro de absurda algoritmización del pensamiento médico.

Ejemplo 2.4.d.

Barcelona, marzo de 1997. Tras producirse de forma explosiva un brote epidémico de colitis en una unidad de geriatría donde se afectaron 12 de los 24 pacientes ingresados, un grupo de científicos pertenecientes al Hospital General Vall d'Hebron de la universidad autónoma de la misma ciudad se lanzó a la caza de los principales factores de riesgo de infección por *Clostridium difficile*, el agente patógeno que con mayor frecuencia se asocia a la diarrea nosocomial en el medio español.

Los resultados del estudio fueron divulgados en el año 2000 a través de la revista Medicina Clínica bajo el título "Factores de riesgo de infección por *Clostridium difficile* en pacientes ancianos. Estudio de casos y controles".⁹²

El diseño implementado, como bien sugiere el título, correspondió a un clásico estudio de casos y testigos que se extendió desde la fecha del brote hasta diciembre del siguiente año. El grupo de enfermos estuvo conformado por 15 pacientes que ingresaron en la institución a lo largo del período de estudio y que presentaron diarreas asociadas a la presencia específica de la toxina A bacilar determinada por un método de enzimoimmunoanálisis. Por cada caso se incluyeron de forma retrospectiva 3 controles emparejados según fecha del ingreso, edad y sexo.

Posteriormente, a partir de la revisión de historias clínicas, se seleccionó un conjunto de variables de interés para ser probadas como posibles factores de riesgo.

El análisis estadístico correspondiente quedó enunciado en el artículo de esta manera:

El estudio estadístico se realizó con el programa SPSS. Para el estudio de las variables categóricas se aplicó la prueba de la χ^2 con la corrección de Yates, o la prueba exacta de Fisher cuando los valores esperados en una o más casillas eran inferiores a cinco. Para estudiar la asociación entre variables numéricas y categóricas se utilizó la prueba de la t de Student o la prueba no paramétrica de la U de Mann-Whitney cuando las variables no seguían una distribución normal. Se consideró que existía significación estadística con una $p < 0,05$. Se realizó un análisis multivariante mediante un modelo de regresión logística paso a paso que incluía las variables que resultaron significativas en el análisis univariado.

Según palabras de los propios autores, el análisis univariado permitió identificar como factores de riesgo para la infección por *Clostridium difficile* a:

- Utilización de sonda nasogástrica para nutrición enteral ($p = 0,035$; OR = 5,13; IC del 95%, 0,95-29,07).
- Número de antibióticos administrados por paciente durante su ingreso hospitalario (1,7 frente a 0,9; $p < 0,01$; OR = 5,69; IC del 95%, 1,26-26,8 en pacientes que recibieron 2 antibióticos o más).
- Duración del tratamiento antibiótico (14,6 frente a 8,8 días; $p = 0,035$).
- Valor sérico de albúmina (2,87 frente a 3,23 g/l; $p = 0,029$).
- Valor plasmático de la hemoglobina (10,8 frente a 12,2 g/l; $p < 0,007$).
- Estancia media hospitalaria (38,8 frente a 14,2 días; $p < 0,001$).
- Destino al domicilio después del alta (el 35 frente al 74%; $p = 0,009$; OR = 5,22; IC del 95%, 1,2- 24,02).

A continuación se lee textualmente:

El análisis multivariado por regresión logística demostró que la nutrición enteral por sonda nasogástrica (OR = 6,73; IC del 95%, 1,01-45,35) y la duración del tratamiento antibiótico (OR = 1,15; IC del 95%, 1,01-1,28) se asociaban de forma independiente a la infección.

Se presencia una típica y completa algoritmización del proceso, ajena en su desarrollo a un pensamiento racional independiente o anterior al empleo de la técnica estadística (fuera de la selección de variables, claro está). Ya se ha explicado la connotación totalmente ficticia de la selección algorítmica de variables en la determinación de asociaciones estadísticas verdaderamente intrínsecas. La técnica, lejos de identificarlas, solo es capaz de establecer, dentro de un conjunto de características, cuáles mantienen una asociación “significativa” con la variable de respuesta.

En esta investigación, como en muchas otras donde se trata de precisar aquellos factores que influyen sobre un fenómeno, se ha implementado la regresión paso a paso como método discriminatorio que intenta hacer emerger las causas del suceso en estudio. Las posibles deducciones que conlleva distan mucho de la realidad por el simple motivo de que “la receta” es absolutamente ineficaz para obtener lo que se desea descubrir. No obstante, el escudo protector que ofrece una metodología muy bien estructurada y supuestamente dotada de la capacidad para identificar relaciones de causalidad ha hecho atribuirle la condición de factor de riesgo a cualquier atributo estudiado que haya merecido permanecer en el modelo *final*. Sin embargo, en este caso particular es imposible no advertir el descalabro conceptual con el que se ha llevado adelante la investigación.

El procedimiento implementado para descifrar los factores de riesgo de la infección por *Clostridium difficile* va más allá de la ceremonia estadístico-metodológica. Los resultados y las conclusiones se han basado en análisis mucho más superficiales y acientíficos que el propio ritual. Nótese que se han considerado como factores de riesgo no solo aquellas variables que conformaron el modelo final de la regresión (nutrición enteral por sonda nasogástrica y la duración del tratamiento antibiótico) sino también las que solo encierran una asociación bivariada con el desenlace. En otras palabras, se ha interpretado la simple asociación estadística entre dos variables, incluso bajo eventuales efectos ajenos, como una pura relación causal, cuando lo único relevante es que varios factores se han distribuido entre los grupos con una desigualdad lo suficientemente grande como para mostrar cierta relación numérico-estadística o probabilística con la variable dependiente.

Como se ha visto, en el artículo se omite, además del análisis del resto de los requerimientos de un proceso causa-efecto mínimamente aceptable, todo pensamiento científico y razonamiento lógico. Estos han sido suplidos por una secuencia de maniobras estadístico-metodológicas que anula de forma radical el pensamiento crítico e irremediablemente genera conocimientos de dudosa credibilidad.

3. Procedimiento estadístico-metodológico: uso de pruebas de significación para demostrar homogeneidad entre los grupos de comparación de ensayos clínicos.

3.1. Bases teóricas generales:

Es frecuente que una investigación científica transite en determinado momento de su desarrollo por el proceso metodológico de la comparación;¹ no obstante se ha dado en diferenciar aquellos trabajos donde el cotejo se realiza con el objetivo de caracterizar una realidad, de los que buscan descubrir o corroborar las causas que la determinan y que explican el modo en que se manifiesta.¹⁴ El primer caso corresponde a

los estudios *descriptivos* mientras que en el segundo se incluyen a los *analíticos*, también conocidos como estudios *explicativos*.

En el campo de las ciencias biomédicas y de la salud el recurso de comparar comparece en investigaciones tanto de índole observacional como experimental. Los procedimientos y consideraciones que se llevan a cabo para la selección de los integrantes de cada grupo difieren de un diseño a otro, pero en todos los casos ha de procurarse que los conjuntos que se comparan sean internamente heterogéneos pero muy parecidos entre sí; de hecho, tan parecidos como sea posible salvo en lo que concierne a la característica que los define (usualmente, la que es objeto de estudio).¹⁴ La similitud u homogeneidad entre los grupos garantiza, en principio, la “comparabilidad” entre estos a través de la eliminación del efecto confusor de determinadas variables. Ello se logra de manera bastante eficiente en los estudios de corte experimental a través de la asignación aleatoria de las unidades de análisis a los grupos de comparación.^{1,14,65}

Dentro de las diferentes modalidades de experimentos, los ensayos clínicos han sido los de mayor implementación, sistematización y control en la práctica médica investigativa. Su empleo principal se ha orientado generalmente hacia la evaluación de intervenciones sanitarias, médicas o quirúrgicas.⁹³⁻⁹⁵

Todo ensayo terapéutico tras la etapa preclínica, donde se realizan estudios sobre animales de laboratorio, transcurre por fases metodológicas donde se estudian sucesivamente desde la seguridad de un nuevo fármaco (fase I o estudio de inocuidad, verdaderos estudios de farmacología humana) hasta la eficacia del producto y su relación dosis-respuesta; son estudios terapéuticos exploratorios (fase II). Posteriormente en la fase III del ensayo, es necesario evaluar la eficacia del tratamiento mediante la comparación de sus efectos en varios grupos tratados de manera diferente. Uno de estos, típicamente, el llamado grupo experimental, recibe el tratamiento que se desea probar y otro u otros grupos de control no reciben terapia alguna o ésta se basa en un placebo o un tratamiento estándar de eficacia conocida.

En las fases II y III es necesario garantizar que los resultados estén determinados por el tratamiento y no por otros factores que puedan influir de otro modo en el desenlace de cada grupo. De esta manera las variaciones observadas de la respuesta en cada grupo serán atribuidas en exclusiva al tratamiento.

Para contribuir a que los grupos de comparación sean lo suficientemente similares entre sí, los sujetos seleccionados que consientan participar en el ensayo se asignan de manera aleatoria a cada régimen de terapia de forma que todos tengan la misma probabilidad de pertenecer a un grupo en específico. La asignación aleatoria es la única intervención metodológica que teóricamente da lugar a una distribución equilibrada de las características de los pacientes entre los diferentes grupos de tratamiento. Se espera que el azar logre, sobre todo cuando los efectivos son numerosos, gracias a la ley de los grandes números, que las variables que pudieran afectar los resultados se distribuyan homogéneamente en cada uno de los grupos. Sin embargo, es evidente que por meras razones de azar puede ocurrir ocasionalmente que los grupos formados por asignación aleatoria simple no resulten “equivalentes” a los efectos de todas

las variables que sería conveniente controlar, incluso cuando los tamaños de muestra no sean especialmente pequeños.

3.2. Empleo ceremonial del procedimiento:

Otra aplicación ceremonial y espuria que incluye la implementación de las pruebas de significación se desarrolla con mucha frecuencia precisamente en el contexto de los ensayos clínicos aleatorizados.

El procedimiento se basa en la creencia de que conviene corroborar que no se ha producido tal desbalance mediante pruebas de significación. Que se lleve adelante tal ritual es algo exigido incluso por algunas instituciones controladoras de la calidad de los ensayos y por algunas revistas.⁹⁴⁻⁹⁷ La idea consiste en seleccionar determinadas variables que se consideren relevantes y comparar entre los grupos medias o porcentajes mediante respectivas pruebas de hipótesis. La regla regularmente aplicada es asumir que los grupos a comparar están apropiadamente equilibrados cuando no existen diferencias significativas para cada una de las variables.

3.3. Análisis de su improcedencia:

El empleo de este procedimiento metodológico para determinar, a modo de validación del ensayo, que los grupos de comparación no son diferentes entre sí, resulta ser una maniobra sin mayor sentido. Las consideraciones tomadas en cuenta para hacer esta afirmación se exponen a continuación:

- La comparación que convalidaría la igualdad entre los grupos respecto a determinadas variables de interés se realiza a través de pruebas de significación, de modo que el rechazo o no de la hipótesis nula, que enuncia que los parámetros (por ejemplo, dos medias o dos porcentajes) correspondientes son exactamente iguales depende íntimamente del tamaño de muestra empleado en la investigación. Para aceptar que los grupos son homogéneos entre sí, que en principio es lo deseado, lo que se necesita es encontrar diferencias no significativas en relación a todas y cada una de las variables que el investigador considere que puedan actuar como confusoras en la asociación entre el tipo de tratamiento y la medida de su efecto terapéutico. Sin embargo, evitar la significación en las pruebas de hipótesis se logra fácilmente siempre que los grupos de estudio sean lo suficientemente pequeños.^{2,3} La solución verdaderamente atinada sería incorporar la información sobre el posible desbalance (independiente de su magnitud) en el análisis, de modo que se neutralice por esa vía su ocasional efecto confusor. Para conseguirlo no es necesario un paso previo que los identifique, pues se han desarrollado varios métodos estadísticos de eficacia probada que permiten el control durante el proceso de análisis de la información no solo de las variables consideradas como confusoras sino de todas las que se estimen convenientes.

El carácter ceremonial del procedimiento en algunos ensayos es particularmente manifiesto cuando, a pesar de encontrar variables para las que sí se obtiene significación, no se toman medidas para corregir su posible efecto distorsionador.

La sugerencia es la siguiente: donde se haga necesario estudiar grupos pequeños, se hace imprescindible controlar durante el análisis los eventuales efectos confusores mediante procedimientos y técnicas estadísticas a pesar de que supuestamente no existan diferencias significativas para las variables analizadas.

- No obstante, la objeción más notable que merece el procedimiento, es de naturaleza lógica. El objetivo esencial de un *test* de contrastación de hipótesis consiste en dirimir si una diferencia hallada es atribuible al azar o si ésta resulta tan acusada que tal explicación puede descartarse. La realización de la prueba por tanto es en este caso redundante, pues el diseño de los ensayos clínicos requiere que los grupos sean conformados aleatoriamente, de modo que de antemano se sabe la respuesta: cualquier diferencia es necesariamente debida al azar. Dicho de otro modo: si se hallara significación, esta se debe a que en la prueba se cometió el error de primer tipo.

En síntesis, se realiza un procedimiento para evaluar si el azar es el responsable de las posibles diferencias encontradas cuando bajo las condiciones que se han impuesto mediante el diseño solo éste puede causar tales diferencias.

3.4. Ejemplos del empleo ceremonial:

Ejemplo 3.4.a.

Para comparar la respuesta a tratamientos tópicos *versus* sistémicos en un grupo de pacientes VIH/SIDA con candidiasis oral, en el estado de Puebla, México, se desarrolló entre marzo de 2003 y junio de 2004, un ensayo terapéutico para 4 drogas, 2 tópicos y 2 sistémicas, cuyos resultados fueron publicados el pasado año en la Revista Cubana de Medicina Tropical del Instituto Nacional Pedro Kourí.⁹⁸

El experimento incluyó un total de 97 pacientes con VIH/SIDA de los cuales 24 recibieron tratamiento externo con Nistatina, 24 con Clotrimazol tópico, 23 con Ketoconazol y 26 con Itraconazol; estos dos últimos administrados por vía sistémica.

En el análisis estadístico se implementaron pruebas de significación para validar la *comparabilidad* entre los cuatro grupos respecto a las variables edad, sexo, cantidad de episodios de candidiasis antes de recibir el tratamiento y forma clínica de la enfermedad.

En el texto se plantean los siguientes resultados:

La edad promedio en los grupos de tratamiento osciló entre 34 y 35 años, con edades mínimas alrededor de 20-23 años y máximas entre 50-82. Las diferencias observadas no fueron estadísticamente significativas, considerándose los grupos homogéneos en cuanto a la edad.

En todos los grupos de tratamiento el sexo predominante fue el masculino. No hubo diferencias estadísticamente significativas en el sexo por tipos de tratamiento, siendo por tanto homogéneos los grupos para esta variable. La cantidad de episodios de candidiasis al recibir el tratamiento

presentó algunas diferencias entre los grupos de comparación, sin embargo estas no fueron estadísticamente significativas.

La gran mayoría de los pacientes presentó forma pseudomembranosa que se distribuyó uniformemente entre los 4 grupos de tratamiento.

En la Tabla 7, se observa la distribución de las variables en cada grupo y los valores p de las pruebas de hipótesis para los casos de la edad, el sexo y la cantidad de episodios.

Tabla 7: Características de los pacientes por tipo de tratamiento realizado.

Características	Grupos de tratamiento				Prueba ANOVA y chi cuadrado
	Tópicos		Sistémicos		
	Nistatina n=24	Clotrimazol n=24	Ketoconazol n=23	Itaconazo l n=26	
Edad					
Promedio	35,79	35,67	34,52	34,77	
Desviación Estándar	13,240	12,737	9,125	7,665	F=0,0813
Mínimo	23	20	23	22	p>0,05
Máximo	82	68	51	50	
Sexo					
Masculino %	70,8	87,5	73,9	80,8	X2=2,341
Femenino %	29,2	12,5	26,1	19,2	p>0,05
Episodio de candidiasis al recibir el tratamiento					
Primer episodio %	29,2	45,9	43,5	30,8	
De 2 a 4 episodios %	50,0	33,3	39,1	53,8	X2=3,225
5 episodios o más %	20,8	20,8	17,4	15,4	p>0,05

Fuente: Revista Cubana Medicina Tropical 2006; 58(3)

El hecho que no se hayan identificado diferencias estadísticamente significativas entre los grupos de este estudio ha sido interpretado por sus autores como que la asignación aleatoria cumplió satisfactoriamente su objetivo de controlar eventuales factores de confusión. Esto es simplemente falso, ya que lo único que en todo caso se ha conseguido es constatar que no hay suficiente evidencia muestral de que los grupos sean diferentes entre sí a esos efectos, algo totalmente distinto a lo que se afirma.

Para comprender cabalmente algunos aspectos relevantes sobre este asunto, es conveniente incursionar brevemente en materia de pruebas de hipótesis.

Es reconocida la dependencia directa que tienen los resultados de las pruebas de significación del tamaño muestral; en esencia, a medida que aumenta n , la probabilidad de rechazar la hipótesis nula tiende a 1 independientemente de cuán pequeña sea la diferencia cuya significación se valore. Contrariamente, la probabilidad de no rechazar H_0 cuando esta es en verdad falsa (error de tipo II) crece en la medida que n disminuye. Por otra parte, esto cobra especial importancia porque en la práctica virtualmente nunca un parámetro se comporta exactamente igual en dos poblaciones diferentes.

Con la escasa cantidad de pacientes sobre los que se han probado los tratamientos en cuestión es sumamente improbable que tales diferencias fueran detectadas como significativas, pues la potencia de cada una de las pruebas vinculadas a la comparación inicial ha sido mínima.

Ilustremos de la siguiente manera lo que ha ocurrido en esta investigación:

En la tabla anterior se expuso, junto con otras, la distribución basal de la variable sexo en cada grupo de tratamiento. Es evidente que dicha distribución no es exactamente igual en cada uno de los grupos.

Como hemos dicho, las diferencias siempre están presentes, pero lo que se plantean los autores es que resulta necesario precisar si las que han sido encontradas en la muestra son lo suficientemente pequeñas como para ser atribuidas al azar en que se basó la asignación de los pacientes a los tratamientos. Es fácil determinar, a partir de la propia Tabla 7, la distribución absoluta de hombres y mujeres por grupo, valores que se exponen en la Tabla 8.

Tabla 8: Distribución absoluta y relativa de la variable sexo por grupos de tratamiento.

Grupos	Masculino		Femenino		Total	
	#	%	#	%	#	%
Nistatina	17	70.8	7	29.2	24	100
Clotrimazol	21	87.5	3	12.5	24	100
Ketoconazol	17	73.9	6	26.1	23	100
Itraconazol	21	80.8	5	19.2	26	100
Total	76	78.4	21	21.6	97	100

Fuente: Revista Cubana Medicina Tropical 2006; 58(3)

La prueba ji cuadrado con 3 grados de libertad implementada por *González* y colaboradores sobre estos datos, arrojó que las diferencias entre los cuatro grupos resultaron insuficientes como para decidir que eran heterogéneos entre sí respecto al sexo (*ji cuadrado* = 2,3165; *p* = 0,5094). Sin embargo, es sencillo corroborar que aun si estas diferencias fueran mucho más acusadas se llegaría a la misma conclusión de que los grupos son, en efecto, comparables.

Véase según el aparatoso ejemplo que se muestra en la Tabla 9, que no sería posible tampoco rechazar H_0 (*ji cuadrado* = 6,6825; *p* = 0,0827), a pesar de que se han impuesto diferencias muy marcadas entre los grupos.

Se ha visto que el análisis de la comparabilidad entre los grupos ha sido redundante e innecesario y termina constituyendo una mera ceremonia metodológica, con el único dividendo de haber creado la ilusión de que todo está en orden.

Tabla 9: Redistribución absoluta y relativa de la variable sexo por grupos de tratamiento.

Grupos	Masculino		Femenino		Total	
	#	%	#	%	#	%
Nistatina	16	66.67	8	33.3	24	100
Clotrimazol	23	95.83	1	4.2	24	100
Ketoconazol	18	75.00	5	21.7	23	100
Itraconazol	19	76.00	7	26.9	26	100
Total	76	79.17	21	21.6	97	100

Ejemplo 3.4.b.

Algo muy similar ocurre en un sencillo ensayo clínico realizado en una consulta de enfermería en el Centro de Salud de Benimaclet, ciudad española de Valencia y que apareció publicado en 1998 en la revista Atención Primaria,⁹⁹ donde Aleixandre y colaboradores, para evaluar la eficacia de dos tipos de intervenciones de deshabituación tabáquica, estudiaron una muestra de 48 fumadores de los cuales fueron asignados (aleatoriamente) 21 al grupo sobre el que se realizaría una influencia mínima para abandonar el hábito de fumar y 27 sobre el que se llevaría a cabo una intervención más especializada.

Los autores plantean literalmente en el artículo:

Los 2 grupos experimentales que formaron parte del estudio fueron comparados respecto a las variables demográficas y las relacionadas con la historia de consumo de tabaco. No encontramos diferencias estadísticamente significativas en ninguna de ellas.

Los resultados al respecto se mostraron en una tabla similar a la de la página siguiente.

Tabla 10: Comparación de los 2 grupos de intervención.

	Intervención mínima (media)	Intervención avanzada (media)	p
Nº sujetos	21	27	NS
Edad	35,60	36,50	NS
Cigarrillos	24,50	27,20	NS
Cafés	2,19	2,74	NS
Alcohol	4,28	7,63	NS
T. Fagestròm	5,76	5,67	NS
A. fumando	18,04	20,81	NS
I. abandono	3,33	2,33	NS
Fumadores en casa (%)	61,90	66,70	NS

NS: no significativo.

Fuente: Revista Atención Primaria 1998; Volumen 22; Número 7

El esquema expuesto en estos dos ejemplos es completamente canónico. La similitud en el manejo de los dos

casos anteriores es tan notable que no se puede agregar nada a lo ya dicho para el primero de ellos.

Pueden hallarse abundantes ilustraciones similares en una búsqueda regular realizada, por ejemplo, en *Scholar Google*. El lector interesado puede examinar muchos otros encontrados por el autor de esta tesis.¹⁰⁰⁻¹⁰⁷

Ejemplo 3.4.c.

Es lógico que en ocasiones la aleatorización no sea suficiente para alcanzar la ansiada uniformidad, sobre todo cuando se utilizan tamaños de muestras insuficientes como para que la ley de los grandes números se verifique. En estos casos, sin embargo, pueden aparecer entre los grupos de comparación diferencias significativas como las encontradas por González y colaboradores en su ensayo terapéutico a doble ciegas, llevado a cabo en el Hospital General La Mancha-Centro de España y publicado recientemente en la revista *Medicina Clínica*.¹⁰⁸

En este estudio se pretendía determinar el posible efecto beneficioso del tratamiento erradicador del *Helicobacter pylori* en pacientes con dispepsia funcional.

Se estudiaron 93 sujetos que, según dictamen clínico, padecían la enfermedad asociada a la presencia del agente etiológico verificada mediante biopsia de la mucosa gástrica. 46 pacientes recibieron una pauta de tratamiento erradicador consistente en la administración oral durante 7 días de 20 mg de omeprazol, 500 mg de claritromicina y 1 g de amoxicilina cada 12 horas, mientras que a los 47 restantes se le administró un placebo. Los autores de esta investigación también consideraron necesario valorar la comparabilidad entre los dos grupos terapéuticos; pero, como puede apreciarse en el siguiente párrafo, tuvieron un percance:

Las variables estudiadas en ambos grupos de tratamiento fueron homogéneas excepto en el uso de alcohol ($p < 0,05$), con una mayor prevalencia en el grupo de tratamiento, aunque ningún paciente presentó una ingesta superior a 40 g al día.

Para estas contrastaciones se utilizaron pruebas de significación en dependencia del tipo de variable:

Se utilizó la prueba de la *t* de Student para el análisis de datos cuantitativos. Para los datos cualitativos se emplearon la prueba de χ^2 y el test exacto de Fisher en los casos pertinentes.

El resultado del análisis de la información, obtenida antes de los pacientes recibir el tratamiento, fue expuesto en el trabajo a través de la Tabla 11.

Aquí la naturaleza formalista del procedimiento es especialmente notable. En efecto, a pesar de que la diferencia encontrada entre los porcentajes de pacientes que consumían alcohol en el grupo que recibió el tratamiento erradicador y en el que recibió el placebo resultó estadísticamente significativa, esto no

expresa que no se deba al azar. Los tratamientos, se sabe, fueron asignados de forma aleatoria lo que hace, por tanto, a la casualidad la única responsable de la distribución desequilibrada de la variable. La explicación no puede ser otra que la siguiente: ha ocurrido el error de primer tipo en la Prueba exacta de Fisher, realizada en este caso para la comparación de los porcentajes. La ceremonia es inútil, pues de antemano se sabe que esta es la explicación en caso de que aconteciera lo que efectivamente ocurrió.

Tabla 11: Características basales de los pacientes según intención de tratamiento.

Características	Grupo placebo (n=46)	Grupo tratamiento (n=47)
Edad media (DE) años	41,5 (13,5)	38,5 (12,01)
Sexo (V/M)	9/37	15/32
Peso medio (DE), kg.	68,3 (10,5)	65,5 (12,7)
Fumador activo, n (%)	12 (26)	15 (31,9)
Uso de alcohol, n (%)	1 (2,2)	7 (14,9)
		(p<0,05)
Uso de alcohol > 60g/día	0	0
Dispepsia, n (%)		
Ulcerosa-like	11 (23,9)	16 (34)
Dismotilidad	33 (71,7)	30 (63,8)
Mixta	2 (4,3)	1 (2,1)
Altura media (DE), cm	162,2 (7,9)	163 (8)
Duración de los síntomas, n (%)		
< 3 años	16 (32,7)	18 (38,2)
3-8 años	20 (43,4)	20 (42,7)
> 8 años	10 (23,9)	9 (19,1)
Gastritis crónica, n (%)		
Superficial	41 (89)	43 (91,5)
Con metaplasia intestinal	5 (11)	4 (8,5)
Escala de dolor, media (DE)	5,91 (1,38)	5,78 (1,75)
Alteraciones psiquiátricas (%)	10 (21,7)	7 (14,9)
Medición previa al estudio, n (%)		
IBP	35 (76)	33 (70,2)
Procinéticos	40 (86,9)	45 (95,7)
Anti-H2	11 (24)	14 (29,8)
Ansiolíticos	6 (13)	6 (12,7)

Fuente: Revista Medicina Clínica 2004; volumen 122; Número 3

Existe un elemento de mayor notoriedad y que incide con bastante frecuencia en los ensayos clínicos controlados una vez efectuada la ceremonia para valorar la comparabilidad intergrupos. A pesar de identificar distribuciones desiguales de alguna o algunas de las variables, el análisis estadístico que continúa no incluye el control de tal desequilibrio con su posible papel confusor en la asociación que se estudia. Siendo así, ¿para qué se hace la evaluación de marras?

Puede comprobarse que *González* y sus colaboradores no llevaron adelante acción alguna para corregir el efecto de la distribución heterogénea del número de consumidores de alcohol.

Todo el análisis estadístico fluye a través de contrastaciones bivariadas en relación al valor que adoptan las variables de evaluación en cada uno de los grupos cuando lo verdaderamente adecuado hubiera sido postestratificar según la variable dicotómica “consumo de alcohol” y aplicar posteriormente el estadígrafo de *Mantel y Haenzel*, o preferiblemente implementar, tras redimensionar las variables de respuesta, una regresión logística.

El objetivo de advertir si los grupos son lo suficientemente parecidos como para ser cotejados es simplemente superfluo, pues a la larga lo sean o no el problema puede ser resuelto directamente controlando los factores de confusión sin necesidad de conocer previamente su similitud.

Ejemplo 3.4.d.

Un estudio donde se controla la variable que resultó significativa en la comparación basal entre los grupos fue publicado en la Revista Medicina Clínica en el año 2000.¹⁰⁹ Su objetivo era evaluar la efectividad del tratamiento combinado con omeprazol, claritromicina y amoxicilina en la erradicación del *Helicobacter* en pacientes portadores de úlcera péptica activa.

El diseño consistió en un ensayo clínico aleatorizado, abierto (sin enmascaramiento), paralelo, multicéntrico, con asignación aleatoria por bloques de 520 pacientes a dos esquemas de tratamiento con iguales medicamentos y posologías pero diferenciados por su duración (6 y 7 días respectivamente).

En el acápite correspondiente al análisis estadístico se expresa lo siguiente:

Se ha estudiado la comparabilidad de los 2 grupos mediante las variables: edad, sexo, localización del ulcus, ulcus de novo o recidivante y consumo de tabaco. Las variables categóricas se han analizado mediante la prueba de comparación de proporciones (χ^2), las cuantitativas mediante el test de la t de Student-Fisher o, en el caso de que las condiciones de aplicación no se cumplieran, con test de Mann-Whitney. Respecto a las características de los pacientes de ambos grupos, sólo se han encontrado diferencias estadísticamente significativas con relación al sexo: el porcentaje de varones era superior en el grupo de 6 días de tratamiento ($p = 0,04$). Para el resto de características estudiadas (edad, tipo de úlcera, si era recidiva o no, y consumo de tabaco), no se encontraron diferencias estadísticamente significativas.

Los autores complementaron lo anterior con un cuadro estadístico semejante a la Tabla 12. Posteriormente se explica que en el análisis se implementó la postestratificación atendiendo al sexo:

En el análisis estratificado según sexo, no se observaron diferencias estadísticamente significativas en el porcentaje de erradicación ($\chi^2 = 1,41$; $p = 0,24$, prueba de Mantel-Haenzel).

Tabla 12: Características basales de los grupos según el tratamiento

Características	7 días	6 días	p
No de pacientes	115	108	
Media de edad, años (DE)	49,8 (13,4)	47,4 (13,8)	NS
Varones (%)	57,4	70,4	0,04
Úlcus duodenal (%)	93,9	91,7	NS
Recidiva (%)	53,4	56,1	NS
Consumo de tabaco (%)	47,8	55,8	NS

NS: no significativo

Fuente: Revista Medicina Clínica 2000; Volumen 114; Número 12

Como se ve, al menos en este caso, aunque se han empleado las polémicas pruebas de significación, éstas han jugado cierto papel real, ya que se tuvo en cuenta la naturaleza significativa de la diferencia en el momento de hacer el análisis.

4. Procedimiento estadístico-metodológico: Confección de fichas técnicas para comunicar resultados de encuestas.

4.1. Bases teóricas generales:

La labor investigativa resulta en la época contemporánea la principal fuente de obtención de nuevos conocimientos, los cuales, para que alcancen alguna utilidad, deberán ser adecuadamente divulgados.¹

Las formas de comunicación científica se caracterizan hoy en día por su gran versatilidad, determinada en principio por el gran auge de los medios de comunicación, recientemente expandidos por el desarrollo impetuoso de la informática, las redes electrónicas y las telecomunicaciones.¹¹⁰ Estos novedosos medios han logrado desdibujar el límite de las fronteras de la comunicación oral y la escrita, gracias a la aplicación de formas que conjugan los textos con el lenguaje hablado, de la misma manera que asocian a las formas básicas de exposición oral, información audiovisual que enriquecen el mensaje que se transmite y que favorecen su comprensión. La comunicación en ciencias también se ha visto beneficiada gracias a estos adelantos tecnológicos.¹¹¹⁻¹¹⁴

En la actualidad, resulta frecuente que se de cuenta de los resultados de investigaciones a través de cualquier medio de comunicación, ya sea a pequeña escala o de forma masiva (léase televisión, radio o prensa escrita).¹¹¹ La elección de cada cual estará siempre sujeta a elementos de primer orden como las características de la información que se desea divulgar, el apremio por su transmisión, los intereses de los emisores, y los destinatarios a los que va dirigida. Por supuesto que el alcance del estudio, sus objetivos, y el uso que se les dará a los conocimientos aportados, inciden de manera directa en la forma de exponer los hallazgos y el medio que se utilice para ello.

Sin embargo, e independientemente de lo anterior, la comunicación científica tiene exigencias cuyo cumplimiento garantizan su calidad. En relación a este aspecto se han desarrollado estándares de cómo exponer los resultados de los estudios académicos, ya sea de forma verbal o en formato de texto.

Dos de los elementos de mayor importancia a considerar en la exposición de un estudio científico son, razonablemente, los resultados alcanzados en el mismo y la explicación detallada e inteligible de las técnicas y procedimientos empleados para arribar a dichos resultados, es decir, la metodología seguida en la investigación.¹¹⁵

La importancia de esta última radica en que aporta los elementos de rigor para el caso en que se desee replicar la investigación por otros autores con diferentes propósitos, así como que contribuye a evaluar la fiabilidad y el carácter científico de los resultados.

En aquellas situaciones en las que se decide divulgar un estudio a través de medios de comunicación cuyo formato, en aras de los límites del espacio físico (o virtual) de que dispone, requiere frecuentemente de un proceso de síntesis, se decide, razonablemente, sacrificar la exposición detallada de la metodología empleada y sustituir ésta por un breve resumen de los elementos que se consideren de mayor relevancia.

4.2. Empleo ceremonial del procedimiento:

Se ha observado, en el caso específico de la publicación de resultados de investigaciones en diarios o periódicos, particularmente cuando se comunican resultados de encuestas aplicadas a la población con afán descriptivo, que se suele sustituir la exposición de toda la metodología del estudio por una breve reseña a manera de resumen que abarca determinados aspectos en relación al diseño muestral utilizado y que se identifica como "ficha técnica de la encuesta".

Estas fichas consisten, de forma básica, en uno o dos párrafos que ofrecen la información técnica supuestamente necesaria para explicar las bases sobre las que fueron implementadas las encuestas. Por lo general, aunque lógicamente pueden existir variaciones, la tendencia es a repetir un patrón sintáctico de manera ritual, donde se hacen afirmaciones generales que solo precariamente dependen de las especificidades del estudio. Determinados errores y sinsentidos se clonan de unas fichas a otras. El mismo formato se presenta una y otra vez en múltiples artículos de prensa; pero se ha hecho presente igualmente en otros medios que gozan de cierta amplitud de espacio (por ejemplo sitios Web), que tienen la posibilidad de brindar una información mucho más explícita.

4.3. Análisis de su improcedencia:

Las fichas técnicas constituyen un ejemplo sorprendente que refleja cómo el pensamiento mecánico y acrítico puede consolidarse hasta extremos alarmantes. La esencia de su propósito sería en principio la siguiente: brindar información para que los especialistas entiendan lo que se ha hecho y para que los ciudadanos corrientes consideren que la encuesta está asentada sobre bases científicas, convenciéndoles de que los errores asociados a las estimaciones expuestas son suficientemente pequeños como para confiar en ellas. Los estereotipos que se han hecho comunes en la práctica hacen

que la mayoría de tales fichas no cumplan ninguna de las dos funciones anteriormente expuestas. El resumen que se hace de la información, en la gran mayoría de los casos, da al traste con lo que se quiere transmitir por resultar en extremo críptico, ambiguo e incomprensible, sobre todo para la población con poco conocimiento científico, a pesar de que constituye la mayor parte de los receptores de la información. Se obtiene una comunicación defectuosa e inconsistente pero en apariencia de alto rigor científico.

Lo cuestionable de las fichas técnicas no es precisamente su carácter sintetizador, justificado, como ya se ha dicho, por el escaso espacio disponible para informar aspectos metodológicos, que escapa de las manos del comunicador y sobre el cual debe ajustarse su exposición, sino su naturaleza errática. Además, la información contenida, cuando llega a comprenderse, es por lo general falsa en el sentido de que sugiere rasgos que la encuesta no tiene y oculta otros que resultan de importancia vital para poder valorarla. Lo anterior puede argumentarse con algunos ejemplos que ilustran claramente la naturaleza esquemática y falaz que suele asociarse al uso de las fichas técnicas. Queremos aclarar que algunos de ellos no corresponden a estudios estrictamente médicos; la temática de las investigaciones en que se han insertado guardan relación, sin embargo, con este campo por tratarse de aspectos sociales influyentes en el estado de salud individual y de las comunidades.

4.4 Ejemplos del empleo ceremonial:

Ejemplo 4.4.a.

La ficha que sigue es reproducida textualmente de los resultados de una encuesta de opinión sobre el aborto realizada en Argentina en el año 2000 y que se encargó de divulgar el sitio www.rinaweb.com.¹¹⁶

De inmediato se observa cuan poco esclarecedor es el lenguaje utilizado, que no dista mucho del que es generalmente empleado. En este caso, el lector solamente podrá conocer a ciencia cierta que la dirección del estudio estuvo a cargo de la empresa *Graciela Römer y Asociados*, que la investigación concernía a las personas de uno u otro sexo entre 18 y 70 años de edad residentes en la ciudades de Buenos Aires y Gran Buenos Aires, que la muestra estuvo compuesta por 600 casos y que la información se tomó en los domicilios entre el 3 y el 6 de junio del año 2000.

Los demás detalles, de hecho los de mayor perfil técnico, resultan, además de insuficientes, bastante confusos. Tal es el caso, por ejemplo, de la mención de un muestreo basado en una estratificación por zona.

Se desconoce qué se entiende por *zona*, cómo pueden ser delimitadas entre sí y cuántas componen el universo de estudio. Tampoco se menciona la forma en que se aplicó el cuestionario (autoadministrado, entrevista, por vía telefónica, etc.), ni cómo se manejó la “no-respuesta” ni los controles de calidad

realizados (en caso de que se hayan producido). Esto no es un asunto lateral, pues es verosímil que las tasas de rechazo hayan sido enormes, quizás de 60% o 70%, de acuerdo con experiencias bien conocidas en las grandes capitales para estudios de este tipo. Todos estos elementos ausentes o simplemente mal comunicados, contribuyen, en verdad, a una pobre comprensión de la metodología utilizada.

FICHA TÉCNICA

Dirección: Estudio Graciela Römer & Asociados

Universo: Residentes en el Área Metropolitana (Ciudad de Buenos Aires y Gran Buenos Aires) de entre 18 y 70 años.

Cobertura: 600 casos en la Ciudad de Buenos Aires y el Conurbano Bonaerense

Técnica Muestral: Probabilística, domiciliaria. Estratificada por zona.

Error estimado: +/- 3,98%. Grado de confiabilidad: 95%

Relevamiento: Entre el 3 y el 6 de junio de 2000.

Fuente de datos históricos: Base de datos propia.

Puede aducirse que todos estos detalles técnicos son exageradamente complicados para algo que solo pretende ser una "ficha", que por definición debería ser ágil y breve. Pero precisamente ahí se revela la naturaleza no funcional de la susodicha ficha. Con tal de construirla e incluirla en el informe, se sacrifican los elementos de verdadero o mayor interés. Si se tratara de una publicación en un periódico, podría admitirse; pero estando la información en un sitio Web, no hay exigencias de espacio que pueda justificar que todo se reduzca a un resumen ininformativo.

Comentado lo que no se dice, concentrémonos ahora en lo que sí se dice. Supuestamente, la técnica muestral incluyó la estratificación del universo por zonas y la posterior selección de viviendas o domicilios dentro de cada una de éstas como unidades de primera etapa. No se especifica en la ficha cómo se seleccionaron los integrantes (unidades de análisis) dentro de cada vivienda elegida ¿Encuestaron a todos los que cumplieran con los requisitos de edad para el estudio? ¿Se realizó una subselección muestral dentro de cada vivienda? En tal caso. ¿Cómo fue ésta? Habría que suponer que estamos en el primer caso, que es lo coherente con que no se dé más información. Pero a nuestro juicio esto es de todo punto imposible. Buenos Aires y el Conurbano bonaerense albergan aproximadamente nada menos que 13 millones de personas. Esto significa que se requeriría disponer o construir una lista compuesta por un número cercano a 2 600 000 viviendas, para luego hacer una elección aleatoria de éstas, algo simplemente impensable desde el punto de vista práctico. En el mejor de los casos, con seguridad se emplearon unidades de primera etapa anteriores a las viviendas.

Es decir, si de algo podemos estar seguros es de que no se trató de una muestra simple aleatoria de viviendas, mucho menos de sujetos (no solo por lo que se nos informa, sino porque tal estrategia muestral hubiera sido demencial en términos prácticos). Tengamos en consideración que el área geográfica donde se ha desarrollado el estudio abarca aproximadamente 4 000 km². Un muestreo simple

aleatorio a partir del directorio de todas las viviendas del área (en caso que este sea viable) supone que los efectivos seleccionados se distribuyen uniformemente por toda la zona, es decir, dispersos por todo el territorio lo que entorpecerá seriamente el trabajo de campo al mismo tiempo que implicará prohibitivos costos de ejecución.

Sin embargo, la fórmula del error empleada es la que corresponde a ese diseño muestral que no se realizó. En efecto, es fácil corroborar que el error estándar que se declara es aproximadamente del 4%, valor que corresponde a haber aplicado la fórmula [4.1] del muestreo simple aleatorio:

$$E = 1,96 \sqrt{\frac{P(100 - P)}{n}}$$

cuando se toma $P=50$, $n=600$ y confiabilidad del 95% ($Z=1,96$).

Este dato no solo es falso por la razón apuntada (es bien sabido que si se usan las fórmulas correspondientes a diseños complejos, los errores son mucho mayores que para el muestreo simple aleatorio) sino también porque, en todo caso, es el que corresponde solo para estimaciones realizadas con toda la muestra (de 600).

Si observamos algunos de los resultados expresados en la página web, podemos esperar que cualquier lector normal, quizás no muy experto en cuestiones de estadística, creará (equivocadamente) que a cada estimación realizada se le puede asociar su respectivo intervalo de confianza obtenido de restar y sumar al valor puntual el error declarado (4%). Sin embargo, procede advertir que para cada estimación el margen de error será diferente teniendo en cuenta sobre todo el tamaño de los subgrupos que son analizados, información que ya no atañe a la población general que ha sido estudiada.

En el informe del estudio se comunica:

La aceptación del aborto sin condiciones se profundiza en población de nivel económico social medio y medio alto (52%), entre los varones (56%), entre los más educados (60%) y entre los que no profesan una religión o se consideran ateos (15%). El rechazo es mayor en la población de nivel bajo y medio bajo (53%), entre los que tienen menor nivel educativo (50%) y entre los que profesan la religión católica (88%).

Consideremos, por ejemplo, el caso de quienes aceptan el aborto entre los “más educados”. Supongamos que este grupo contuviese al 20% de la población de la capital argentina. Siendo así, el tamaño muestral real para hacer la estimación sería aproximadamente la quinta parte de 600, es decir, 120 individuos.

Aplicando la fórmula, tendríamos:

$$E = 1,96 \sqrt{\frac{(50)(50)}{120}} = 8.95$$

Es decir, el error sería en principio más del doble del declarado en la ficha. En realidad, el error cometido es mucho mayor, pues usando las fórmulas correctas, una estimación razonable pudiera estar alrededor de $E=14\%$ (véase el procedimiento para tal estimación en Silva, 2000).¹

Dos asuntos más cabe señalar en este contexto: Primero, llama poderosamente la atención que el tamaño de muestra haya sido exactamente igual a un múltiplo de 100. Es cierto que el azar, centrado en la tasa de no respuesta acontecida, pudo haber determinado tan precisa cifra de personas a las que fue posible aplicar el cuestionario, sin embargo es razonable desconfiar de ella sobre todo cuando es típico en las fichas técnicas declarar tamaños redondos de n . Cualquiera en su sano juicio puede sospechar que tras un muestreo por conglomerados (que es el que sugiere la lógica que se haya implementado en este caso) fue necesario en la práctica alguna manipulación encaminada a procesar exactamente los datos obtenidos de 600 personas.

En segundo lugar, al informar que “el grado de confiabilidad fue del 95%” se está haciendo un clarísimo y “conveniente” abuso de lenguaje. Con la expresión se pretende insinuar un alto nivel de calidad a los resultados obtenidos por la encuesta. Eso no tiene sentido, pues el nivel de confianza no es inherente a la encuesta sino que es prefijada de antemano por el investigador y significa única y exclusivamente que los intervalos de confianza de las estimaciones han sido contruidos con ese alto margen de confiabilidad. Esto solo trae consigo una consecuencia: los intervalos resultantes serán de mayor amplitud lo que permite, con gran seguridad, que el valor verdadero del parámetro poblacional esté incluido en el intervalo calculado (de 100 veces que se seleccione una muestra de 600 personas del área metropolitana de Buenos Aires, 95 aportarán intervalos que contienen el valor real del parámetro en la población).

Pudiera pensarse que hemos hecho un inventario de errores cometidos por estos autores que pudieron ser descubiertos al examinar con detenimiento la ficha técnica. Pero no es este el sentido del análisis que nos interesa. Estos deslices o pifias no son más que típicos disfraces que oscurecen la realidad tras el afán de satisfacer una simple expectativa formal. Con el anterior compendio de errores, omisiones y declaraciones sin sentido hemos puesto en franca evidencia que la ficha técnica es por lo común equívoca y conduce, casi siempre, a conformar juicios esquemáticos y falaces.

Ejemplo 4.4.b.

Otra ficha técnica, en esencia muy similar, se presenta en la versión electrónica del periódico *El observador*, a propósito de una encuesta que recogía información acerca de las características de la familia española, la vivienda, la situación del empleo y la utilización del tiempo libre y que fue bautizada con el nombre *Encuesta Familia, Casa, Trabajo y Tiempo Libre*.¹¹⁷

FICHA TÉCNICA

803 encuestas telefónicas a mayores de 18 años. Muestreo aleatorio simple, con una distribución geográfica proporcional a la población española. Para un nivel de confianza del 95,5%, el error muestral es del + - 3,53%.

Trabajo de campo realizado del 18 al 22 de febrero de 2002 por la Empresa Iberphone.

Además de padecer de los mismos males de laconismo y de incluir información incorrecta sobre los errores, es una incógnita lo que se ha querido expresar en la segunda oración (*Muestreo aleatorio simple, con una distribución geográfica proporcional a la población española*) que indudablemente, lejos de aclarar cualquier aspecto metodológico relacionado con el diseño de la muestra, puede confundir hasta al lector más agudo y defraudar al medianamente exigente.

Ejemplo 4.4.c.

Este ejemplo concierne a otro estudio de gran envergadura realizado en España durante 1997 bajo el nombre de *Encuesta Nacional de Salud*. Como se verá, la ficha correspondiente es algo más extensa e informativa, pero presenta los mismos clichés de las anteriores, aunque se agregan algunos otros giros, frecuentes en algunas fichas y tan crípticos e incomprensibles como algunos de los que ya hemos examinado.

Se procuraba conocer, a través de un cuestionario, el estado de salud de la población española. La estrategia seguida condujo a realizar varias etapas u “oleadas” en la aplicación del cuestionario que fue realizado bajo el amparo del Ministerio de Sanidad y Consumo.

El universo se dividió en adultos (16 años y más) y niños (de 0 a 15 años) cada cual representado por una muestra de 6 400 y 2000 unidades respectivamente. Los resultados de la tercera etapa de implementación son expuestos en el sitio www.cis.es ¹¹⁸ donde se incluyen elementos generales del estudio.

Aparte del ya consabido texto completamente inaceptable sobre los errores de muestreo que parece clonado en todas las fichas, la que ahora nos ocupa se destaca por la inclusión de oraciones cuyo significado es misterioso, tales como que la selección de los individuos se realizó “por rutas aleatorias y cuotas de sexo y edad” y que los estratos se han conformado por el “cruce de las regiones independientes con el tamaño de hábitat”.

No resulta posible pasar por alto, finalmente, la impresionante tasa de no respuesta alcanzada por este estudio, donde solo se dejó de entrevistar 1 persona de los 1 600 previstos. Es bien sabido el problema que entraña la no respuesta en investigaciones que utilizan muestras de este tipo donde tasas cercanas a 10% se consideran un resultado formidable. En este caso la cifra se iguala a 0,06% lo que representaría un verdadero y absoluto record mundial en materia de no-respuesta.

A continuación mostramos los datos correspondientes a la llamada *tercera oleada* de los cuales solo hemos incluido los concernientes a la población adulta:

FICHA TÉCNICA

Tamaño diseñado: 1.600 entrevistas

Realizada: 1.599 entrevistas

Afijación: Proporcional a la población de los diferentes estratos

Ponderación: Si se tabula cada submuestra, la de adultos y la de niños, de forma independiente no procede la ponderación porque son autoponderadas.

Si se desea explotar ambas submuestras conjuntamente es necesario aplicar los siguientes coeficientes de ponderación: submuestra de niños 0,888; submuestra de adultos 1,035.

Puntos de muestreo: 97 municipios y 43 provincias.

Procedimiento de muestreo: Polietápico, estratificado por conglomerados, con selección de las unidades primarias de muestreo (municipios), y de las unidades secundarias (secciones) de forma aleatoria proporcional y de las unidades últimas (individuos) por rutas aleatorias y cuotas de sexo y edad.

Los estratos se han formado por el cruce de las 17 regiones autonómicas con el tamaño de hábitat, dividido en 7 categorías: menos o igual a 2.000 habitantes; de 2.001 a 10.000; de 10.001 a 50.000; de 50.001 a 100.000; de 100.001 a 400.000; de 400.001 a 1.000.000; más de 1.000.000 de habitantes.

Error muestral: Para un nivel de confianza del 95,5% y $P = Q$, el error es de $\pm 2,50$.

Fecha de realización: Del 15 de septiembre al 12 de octubre, 4 semanas. Se han seguido las normas del "Diseño General" en cuanto a la aplicación de la submuestra por bloques.

Ejemplo 4.4.d.

Un estudio financiado y publicado por el diario El País acerca de los conocimientos sobre educación sexual de la juventud en Cali, Colombia, *Los jóvenes saben todo del condón pero no lo usan*,¹¹⁹ apareció en las páginas de dicho periódico el 15 de junio del 2005. El artículo mostró los resultados de la encuesta confeccionada y llevada a cabo en el mismo año por la empresa especializada *Analizar y Asociados*. La ficha técnica correspondiente reza así:

FICHA TÉCNICA

Evaluación cuantitativa de la opinión de los caleños sobre el sexo, contratada por El País S.A. y realizada por Analizar & Asociados, con horario restringido el 1 y 2 de junio de 2005. 300 encuestas autoadministradas en hogares a hombres y mujeres mayores de 18 años, de todos los estratos, escogidos al azar. Margen de confiabilidad: 95%.

Los autores (o quizás los editores) han decidido realizar un exiguo esbozo del diseño de la investigación y su metodología que, evidentemente, nada aporta a quien, conocedor de estos asuntos, pueda interesarse en ellos por una u otra causa. Nótese que se desconocen cuáles son todos los estratos a que se refieren y, sobre todo, el hecho de que la selección haya sido efectuada al azar no precisa, de ninguna manera, el tipo de muestreo que se ha utilizado. Se nota nuevamente la presencia de la absurda información según la cual la confiabilidad del estudio fue del 95%.

4.5. Una última observación:

La bibliografía consultada en la presente investigación no arrojó muchos resultados de estudios de corte médico o salubrista que incluyeran fichas técnicas en sus informes, caso de haber implementado algún cuestionario. Esto se explica razonablemente, porque la mayoría de esta clase de trabajos se exponen en medios de mayor especialización y consagración científica. No obstante, además de los 4 ejemplos arriba mostrados, el lector puede encontrar otros muy similares, de connotación médico-social, en la bibliografía acotada.¹²⁰⁻¹³²

También dentro del ramo del comercio y la economía se desarrollan estudios que incluyen encuestas de opinión pública que posteriormente se presentan en los medios con sus respectivas *fichas técnicas*¹³³⁻¹⁴² con deficiencias idénticas o muy similares a las expuestas arriba.

El desarrollo de esta investigación ha puesto de manifiesto que el fenómeno de las fichas técnicas deficientes, equívocas, que contienen datos falsos y errores, pero sobre todo fórmulas sintácticas que nada claro significan, es harto común en otras zonas, quizás de menos rigurosidad académica, pero donde se emplea también la estadística como parte del método científico y que incluye la comunicación de los resultados obtenidos en investigaciones. Sin lugar a dudas, este sustituto de la sección metodológica en los artículos de ciencia, se ha popularizado sobre todo en el sector político donde ha recibido una amplia aplicación en la comunicación de los resultados de encuestas para los escrutinios presidenciales en las campañas electorales. Los ejemplos en estos casos pueden resultar asombrosos sobre todo si se analiza la gran similitud, a pesar de ser estudios realizados por diferentes empresas o compañías y divulgados en medios de diversa índole.

Consideramos muy importante resaltar este aspecto y ofrecer al lector la oportunidad de apreciar tal manifestación en otra bibliografía.¹⁴³⁻¹⁴⁶ Estos ejemplos, a pesar de desviarse un poco del tema de esta tesis (solo un poco, pues debe recordarse que el concepto salud abarca esferas tanto biológicas, psíquicas como sociales) dan cuenta, en definitiva, de una tendencia ceremonial de la estadística como disciplina, por lo que hemos considerado de utilidad mencionarlas en nuestro informe.

A continuación, sin merecer comentarios que aporten elementos novedosos a los ya tocados en relación a las fichas técnicas exponemos una serie de varios ejemplares correspondientes a diferentes áreas del quehacer humano:

Encuesta "Intención de compra a 12 meses".¹³³

FICHA TÉCNICA

800 encuestas telefónicas a mayores de 18 años. Muestreo aleatorio simple, con una distribución geográfica proporcional a la población española. Para un nivel de confianza del 95,5%, el error muestral es del + - 3,54%. Trabajo de campo realizado del 18 al 22 de febrero de 2002 por la Empresa Iberphone.

Estudio "Consumo de Muebles".¹³⁴

FICHA TÉCNICA

1 837 encuestas válidas a hogares representativos de la población española. Entrevistas autoadministradas mediante Homescan de ACNielsen. Para un nivel de confianza del 95,5%, el error muestral es del + - 2,3%. Trabajo de campo del 28 de enero al 21 de febrero de 2002.

Encuesta "Consumo de Muebles y otros datos sobre Consumo en general".¹³⁵

FICHA TÉCNICA

2 000 encuestas telefónicas a mayores de 18 años. Muestreo aleatorio simple, con una distribución geográfica proporcional a la población española. Para un nivel de confianza del 95,5%, el error muestral es del + - 2,23%. Se ha realizado un total de 411 encuestas a personas que han comprado muebles en los últimos 12 meses, lo que supone el 20,6% de la muestra total. Trabajo de campo realizado del 10 al 18 de julio de 2001 por la Empresa Iberphone.

Perú: Simulación de Voto. Segunda Vuelta Presidencial 2006.¹⁴⁵

FICHA TÉCNICA

Encuestadora: Idice del Perú

Registro: 101/ree/jne

Encargada: Universidad Inca Garcilazo de la Vega

Técnica: urna-simulación de voto

Representatividad: a nivel nacional

Muestra: 4850 hogares

Fecha: del 23 al 26 de mayo del 2006

Nivel de confianza: 95.0%

Margen de error: ± 4.5% nacional

Correo electrónico: idiceperu@yahoo.com

Página web: www.idice.com.pe

FICHA TÉCNICA

Tamaño y distribución de la muestra: 800 entrevistas fijadas mediante muestreo estratificado por región y tamaño de hábitat proporcional a la distribución de la población y con ampliación de cuotas de sexo y edad.

Error de muestreo: asumiendo los criterios de muestreo aleatorio simple, para un nivel de confianza de 95,5% (dos sigmas) y para la hipótesis más desfavorable ($p=q=50$), el error para el total de la muestra sería de $\pm 3,5\%$.

CONCLUSIONES GENERALES Y RECOMENDACIONES

Ha quedado evidenciado claramente que las cuatro ceremonias estadístico-metodológicas elegidas para nuestro estudio están vigentes en la investigación médica contemporánea.

Si bien el campo que se ha explorado compete preferentemente a las ciencias de la salud con algunas incursiones en el marco social, vale mencionar que cualquiera de los procedimientos puede ser utilizado en cualquier área de la investigación científica, pues es fácilmente reconocible que la esencia ceremonial no se debe al método propiamente sino a los usuarios que lo utilizan.

Los efectos que generan estas ceremonias o rituales, según se ha demostrado en este trabajo, suelen ser poco provechosos, principalmente en aquellos casos donde la aplicación del procedimiento trae consigo el arribo a conocimientos que constituyen un reflejo desacertado de la realidad que se investiga. Podemos asegurar que cualquier ceremonia estadístico-metodológica implica, en una u otra medida, resultados contraproducentes, incluso aquellas, como el cálculo del tamaño de muestra, que no afectan directamente los resultados. Lo impugnable radica sobre todo en su carencia de utilidad aunque constituyen, además, una forma de deshonestidad y una expresión de puro formalismo metodológico que acostumbra al pensamiento inercial.

Coincidimos en que el lenguaje de las ciencias debe ser nítido, pero además sencillo y fundamentalmente decoroso. Urge que no sea acompañado de artificios que simulen tecnicismo y rigor. En ocasiones un "humilde" estudio que implementa técnicas y procedimientos, quizás a los ojos de algunos, poco sofisticados metodológicamente puede aportar conocimientos de gran utilidad.

Cabe puntualizar que en este trabajo no se encontró ninguna de las aplicaciones rituales estudiadas en la literatura de alto impacto científico internacional. Tal es el caso de la prestigiosa revista *New England Journal of Medicine* de la cual fueron revisados, infructuosamente a nuestros efectos, varios artículos que por su temática, podrían padecerla.¹⁴⁸⁻¹⁵⁴

Como ya se ha mencionado, la estadística es solo una herramienta que puede facilitar el trabajo investigativo, siempre y cuando vaya de la mano del razonamiento lógico y juicioso del investigador. En la mayoría de las ocasiones, al acudir mecánica e inercialmente a recursos metodológicos, se produce un divorcio entre la razón y la técnica lo cual dificulta el hallazgo de una solución real al problema que se aborda.

Indiscutiblemente, no pocas conductas ceremoniales están actualmente muy arraigadas en la comunidad científica; en ocasiones se hace necesario contemporizar con ellas, sin embargo, no podemos convalidar ni el error ni el engaño que en ocasiones está asociado a ellas. Es menester comprender que identificar determinados errores, comprenderlos, es una manera de ayudar a que no se cometan. Es por ello que,

teniendo en cuenta todo lo que se ha expuesto en esta tesis, consideramos oportuno hacer las siguientes recomendaciones:

- Divulgar la crítica a las ceremonias estadístico-metodológicas en cursos de bioestadística y metodología de la investigación, así como en las especialidades de bioestadística e higiene y epidemiología.
- Desarrollar investigaciones donde se identifiquen y examinen otros procedimientos empleados de manera ceremonial.
- Realizar estudios bibliométricos y de series cronológicas que permitan, respectivamente, aquilatar la magnitud del fenómeno y describir su evolución histórica.

BIBLIOGRAFÍA

1. Silva LC. Cultura estadística e investigación científica en el campo de la salud: una mirada crítica. Madrid: Díaz de Santos; 1997.
2. Benavides A, Silva LC. Contra la sumisión estadística: un apunte sobre las pruebas de significación. *Metas de Enfermería*. 2000; 3(27).
3. Silva LC, Sarría M. La pruebas de significación estadística en tres revistas biomédicas: una revisión crítica. *Revista Panamericana de Salud Pública*. 2004; 15(5).
4. Altman D, Goodman S. How statistical expertise is used in medical research published. *J Am Med Assoc*. 2002.
5. Medawar P. Consejos a un joven científico. México DF: Fondo de Cultura Económica; 1984.
6. Gigerenzer G. We need statistical thinking, not statistical rituals. *Behav Brain Sci*. 1998.
7. Silva LC. Diseño razonado de muestras y captación de datos para la investigación sanitaria. Madrid: Díaz de Santos, 2000.
8. Ordúñez P, Silva LC, Rodríguez MP, Robles S. Prevalence estimates for hypertension in Latin America and the Caribbean: are they useful for surveillance? *Revista Panamericana de Salud Pública*. 2001; 10(4).
9. Schulz KF, Grimes DA. Sample size calculations in randomised trials: mandatory and mystical. *Lancet*. 2005; 365: 1348–1353.
10. Nickerson RS. Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods*. 2000; 5(2): 241–301.
11. Matthews RA. Facts versus Factions: the use and abuse of subjectivity in scientific research. *European Science and Environment Forum Working Paper*; reprinted in *Rethinking Risk and the Precautionary Principle*. Ed: Morris, J. Oxford: Butterworth. 2000.
12. Comité Internacional de Directores de Revistas Médicas. Requisitos uniformes para la preparación de manuscritos. *Revista Panamericana de Salud Pública*. 2004; 15(1).
13. Goodman SN. Toward evidence based medical statistics (I): The p value fallacy. *Annals of Internal Medicine*. 1999; 130: 995-1004.
14. Silva LC. Una ceremonia estadística para identificar factores de riesgo. *Salud Colectiva*. 2005; 1(3): 309-322.
15. Irala J, Faulín FJ. *Bioestadística amigable*. Madrid: Díaz de Santos; 2001.
16. Daniel W. *Bioestadística. Base para el análisis de las ciencias de la salud*. Limusa. Distrito Federal, 2004.
17. Camel F. *Estadísticas médicas y de salud pública*. La Habana, Pueblo y Educación, 1978.
18. Azorín F, Sanchez-crespo JL. *Métodos y aplicaciones del muestreo*. Alianza, Madrid, 1986.
19. Argimón JM, Jiménez J. *Métodos de investigación. Aplicados a la atención primaria de salud*. Doyma, Barcelona, 1991.
20. Jenicek M, Cleroux R. *Epidemiología: principios-técnicas-aplicaciones*. Salvat, Barcelona, 1987.

21. Doménech JM. Métodos estadísticos en ciencias de salud. Unidad Didáctica 5, Gráficas Signo, Barcelona, 1990.
22. Lemeshow S, Hosmer Jr DW, Clark J, Lwanga SK. Adequacy of sample size in health studies. New York, Wiley, 1990.
23. Lwanga SK, Lemeshow S. Sample size determination in health studies: a user's manual. Geneva, World Health Organization, 1989.
24. Cochran WG. Sampling techniques. New York, Wiley, 1963.
25. Yamane T. Elementary sampling theory. La Habana, Editorial R, 1970.
26. Mejía JM, Fajardo A, Gómez A, Cuevas ML, Hernández H, Garduño J et al. El tamaño de muestra: un enfoque práctico en la investigación clínica pediátrica. Boletín Médico del Hospital Infantil de México 1995; 52.
27. Hansen MH, Hurwitz WN, Madow WG. Sample survey methods and theory. New York, Wiley, 1953.
28. Santana F, G Ovies G, Verdeja OL, R Fleitas R. Características de la primera relación sexual en adolescentes escolares de Ciudad de La Habana. Revista Cubana de Salud Pública 2006; 32(3).
29. Hernández M, García RG, Pérez D, Ramos D. Información sobre prevención de accidentes que poseen adultos y niños a su cuidado. Revista Cubana de Higiene y Epidemiología 2001; 39(2):95-100.
30. Montero S. Hilo quirúrgico Agasut QE como una alternativa para el cierre de aponeurosis y piel en incisiones de laparotomía [tesis]. Instituto Superior de Medicina Militar "Dr. Luis Díaz Soto". Ministerio de las Fuerzas Armadas; 2004.
31. Mur T, Franch J, Morato J, Mena A, Vilarrubias M, Llobera A. Macroangiopatía en la diabetes tipo II. El estudio Raval Sud. Atención Primaria 1995; 16 (2).
32. Dios JA. Prevalencia de Cardiopatía Isquémica y factores de riesgo coronario en oficiales del Territorio del Ejército Oriental (1997-1999) [tesis]. Facultad de Ciencias Médicas No.2 de Santiago de Cuba. Instituto Superior de Ciencias Médicas de Santiago de Cuba. Ministerio de Salud Pública.
33. López R. Fiebre Reumática. Estudio clínico-epidemiológico. Pinar del Río. Cuba 1986-1996 [tesis]. Hospital Provincial Docente Clínico Quirúrgico León Cuervo Rubio. Ministerio de Salud Pública; 2000.
34. Gómez N. Factores pronósticos de deterioro cognitivo en pacientes con infarto cerebral [tesis]. Hospital Clínico Quirúrgico "Hermanos Amejeiras". Ministerio de Salud Pública; 2003.
35. Ochoa RF. Sistemas ELISA en ensayos clínicos de vacunas y estudios seroepidemiológicos [tesis]. Instituto Carlos J. Finlay. Centro de Investigación-Producción de Vacunas y Sueros. Consejo de Estado; 2004.
36. Fernández JC. Efectos del Policosanol en ancianos con hipercolesterolemia tipo II [tesis]. Centro de Productos Naturales. Centro Nacional de Investigaciones Científicas. Ministerio de Educación Superior; 2002.
37. Fernández LC. Eficacia, seguridad y tolerabilidad del Policosanol en pacientes con riesgo aterosclerótico [tesis]. Centro Nacional de Investigaciones Científicas. Ministerio de Educación Superior; 1999.

38. Bermúdez IB. Actividad antimicrobiana, balance redox celular, seguridad y ventajas económicas de una crema elaborada con propóleos de Manzanillo, en pacientes de pie diabético [tesis]. Instituto de Farmacia y Alimentos. Universidad de La Habana. Ministerio de Educación Superior; 2003.
39. Bermúdez B. Eficacia clínica, seguridad y ventajas económicas de una crema de propóleos, en el tratamiento del pie diabético neuroinfeccioso [tesis]. Instituto de Farmacia y Alimentos (IFAL). Universidad de La Habana (UH). Ministerio de Educación Superior; 2004.
40. Alemany EJ. Alcance del modelo del especialista en Medicina General Integral [tesis]. Ciudad de La Habana. Facultad de Ciencias Médicas "Miguel Enríquez". Instituto Superior de Ciencias Médicas de La Habana. Ministerio de Salud Pública; 2004.
41. Díez JP, Debesa F, Sosa IA, García RG, Bonet M, Fernández N. Comportamiento de la asistencia prestada al paciente en las farmacias comunitarias de Ciudad de La Habana, 1997. *Rev Cubana Hig Epidemiol* 2001;39(3).
42. Díaz O, Valenciaga JL, Domínguez E. Comportamiento epidemiológico de la diabetes mellitus en el municipio de Güines, 2002. *Rev Cub Hig y Epi* 2004; 42 (1).
43. Collado AM, Barberis AE, Aguilar J, López JF. Condiciones de vida y morbilidad en niños y adolescentes en el municipio La Habana Vieja. *Rev Cubana Hig Epidemiol* 2003; 42 (3).
44. Castañeda I, Solís S. Desigualdades atribuibles al género entre cirujanos(as). Ciudad de La Habana. *Rev Cubana Salud Pública* 2002; 28(2).
45. Díaz G, Alonso R, Bayarre H. Factores de riesgo de insatisfacción con la vida en mujeres de edad mediana. *Rev Cubana Hig y Epidemiol* 2002; 40 (3).
46. Ochoa RF, Martínez JC, Ferriol XR, Sotolongo FT. Niveles de antitoxina tetánica y diftérica en recién nacidos y niños preescolares cubanos. *Rev Cubana Med Trop* 2006; 58(1).
47. Núñez FA, González OM, Bravo JR, Escobedo AA, González I. Parasitosis intestinales en niños ingresados en el Hospital Universitario Pediátrico del Cerro, La Habana, Cuba. *Rev Cubana Med Trop* 2003;55(1).
48. Aranda C, Pando M, Salazar JG, Torres TM, Aldrete MG, Pérez MB. Síndrome de Burnout en médicos familiares del Instituto Mexicano del Seguro Social, Guadalajara, México. *Rev Cubana Salud Pública* 2005; 31(2).
49. Iglesias M, Farrús M, Gimbert R, Montaner I, Mundet X, Santos PT. Análisis de los costes de la atención a la hipercolesterolemia en un equipo de atención primaria. *Atención Primaria* 1995; 16 (5).
50. Mur T, Franch J, Morató J, Llobera A, Vilarrubias M, Ros C. Nefropatía y microalbuminuria en la diabetes tipo II. *Atención Primaria* 1995; 16 (9).
51. Cortés JA, Costa P, Pérez-Ventosa V, Ramos R, Fernández I, Ahijado D, Bravo P, et. al. Prevalencia de infección por hepatitis B y factores de riesgo con motivo de una campaña de vacunación en adolescentes. *Atención Primaria* 1995; 16 (10).
52. Fernández de Mendiola J et al. Evaluación de la población diabética tipo II atendida en un equipo de atención primaria. *Atención Primaria* 1996; 17 (7).
53. Martínez V, Salcedo F, Franquelo R, Jarabo Y, Muelas F, García P, Cebrián C. Factores de riesgo cardiovascular en escolares de la ciudad de Cuenca. *Atención Primaria* 1996; 17 (9).

54. Serna MC, Serra L, Miret M, Gascó E, Perera A, Mitjavila J. Prevalencia de tabaquismo en 6 comarcas de Lleida. *Atención Primaria* 1997; 19 (5).
55. Rodríguez JC, Calonge S, Bichara G. Prevalencia de los factores de riesgo de cardiopatía isquémica en la isla de Lanzarote. *Medicina Clínica* 1993; 101 (2).
56. Muñiza J, Juanea R, Castro-Beirasa A, Fernández I, Lorenzo S, Sánchez J. Estudio de casos y controles de factores de riesgo de accidente cerebrovascular agudo. *Medicina Clínica* 1993; 101 (11).
57. M Monras, N Freixa, LI Ortega, A Lligoña, S Mondón, A Gual. Eficacia de la terapia de grupo para alcohólicos. Resultados de un ensayo clínico controlado. *Medicina Clínica* 2000; 115 (4).
58. Conesa C, Ríos A, Ramírez P, Rodríguez MM, Canteras M, Parrilla P. Actitud de la población hacia una legislación de consentimiento presunto a la donación de órganos de cadáver. *Medicina Clínica* 2004; 122 (2).
59. Baré M, Jordana R, Toribio R, Montull S, Solá J, Torné E, Moreno C, Bonfil X. Decisiones terapéuticas en los pacientes hospitalizados. *Medicina Clínica* 2003; 122 (4).
60. Salvá A, Bolívar I, Pera G, Arias C. Incidence and consequences of falls among elderly people living in the community. *Medicina Clínica* 2004; 122 (5).
61. S Balanza, F Mestre. Factores de riesgo cardiovascular en la población pesquera de Cartagena y Castellón. *Revista Española de Salud Pública* 1995; 69 (3-4).
62. M J Tormo, C Nauamo, M D Chirlaque, D Pérez. Factores de riesgo cardiovascular en la región de Murcia, España. *Revista Española de Salud Pública* 1997; 71 (6).
63. Rothman JK. *Modern epidemiology*. Little, Brown and col. Boston, 1986.
64. Resik P. *La Causalidad en epidemiología*. Ciudad de la Habana: Científico-Técnica; 2003.
65. De Irala J, Martínez MA, Guillén F (2001) ¿Qué es una variable de confusión? *Medicina Clínica*. 2001;(117).
66. Silva LC, Barroso IM. *Regresión Logística*. Madrid: La Muralla; 2004.
67. Silva LC. *Excursión a la regresión logística en ciencias de la salud*. Madrid: Díaz de Santos; 1994.
68. Pardo G, Cedeño M. *Investigación en salud. Factores sociales*. Santafé de Bogotá: McGraw-Hill Interamericana; 1997.
69. Silva LC, Benavides A. Causalidad e inobservancia de la premisa de precedencia temporal en la investigación biomédica. *Methodologica (Bélgica)* 1999; 7: 1-1.
70. Kleinbaum DG, Kupper LL, Muller KE, Nizam A. *Applied regression analysis and multivariable methods*. 3rd ed. USA: Duxbury Press; 1997.
71. Taubes G. Epidemiology faces its limits. *Science*. 1995 ;(269):164-169.
72. Barroso IM, Silva LC. Selección algorítmica de modelos en las aplicaciones biomédicas de la regresión múltiple. *Medicina Clínica*. 2001 116 (19): 741-745.
73. Gardner MJ, Altman DG. Confidence intervals rather than P values: estimation rather than hypothesis testing. *British Medical Journal*. 1986; 292: 746–50.
74. Chia KS. "Significant-itis" an obsession with the p values. *Scand J Work Environ Health*. 1997; 23:152–4.
75. Greenwald AG, Gonzalez R, Harris RJ, Guthrie D. Effect sizes and p-values: what should be reported and what should be replicated? *Psychophysiology*. 1996; 33:175–83.

76. Silva LC, Muñoz A. Debate sobre métodos frecuentistas vs. bayesianos. *Gac Sanit.* 2000; 14:482–94.
77. Sarria M, Silva LC. Las pruebas de significación estadística en tres revistas biomédicas: una revisión crítica. *Revista Panamericana de Salud Publica* 2004; 15(5):300-306.
78. Thompson B. Various editorial policies regarding statistical significance tests and effect sizes [en línea]. 2003 [fecha de acceso 20 de Marzo de 2003]. URL disponible en: <http://www.coe.tamu.edu/~bthompson>.
79. Beaglehole R, Bonita R, Kjellström T (1994) *Epidemiología básica* Organización Panamericana de la Salud, Publicación Científica Nº 551, Washington.
80. Chatterjee S, Hadi AS, Price B. *Regression analysis by example*. Nueva York: Willey, 2000.
81. Draper N, Smith H. *Applied regression analysis* (2.ª ed.). Nueva York: Willey and Sons, 1981.
82. Bundó M, Aubà J, Vallés R, Torner O, Pérez AM, Massons J. Arteriopatía periférica en la diabetes mellitus tipo 2. *Atención Primaria* 1998; 22 (1).
83. Robles JA, Hernández R, Fonte R. Diagnóstico y tratamiento de las hiperlipoproteinemias. *Revista Cubana de Medicina General Integral*. 1999; 4.
84. Gómez O, Fernández J E, Núñez M, Menau T, Ferrer M, Mígueles R, et.al. Factores de riesgo aterogénico en una población de adultos mayores, *Revista Cubana de Enfermería* 2005 ; 21.
85. Escobar N V. *Lípidos y Ateroesclerosis*. Santiago de Cuba: Editorial Oriente: 2004.
86. Toros H, Castellanos R, Fernández JE. La asociación de dislipidemia y trombosis en la inestabilización de la placa aterosclerótica. *Rev Cubana Invest Bioméd* 2005; 22.
87. Acevedo SM, Aguillón RA. Manejo de la dislipidemia en paciente diabético tipo 2 [publicación periódica en Internet] 2006 vol,pag [citado 31 Ene. 2007]. Disponible en URL:<http://www.medigraphic.com/espanol/e-htms/e-iner/e-in2006/e-in06-2/er-in062c.htm> - 10k.
88. Alvares-Li FC. Epidemiología de las enfermedades cerebrovasculares en Cuba. *Revista de Neurología* 1999; 29: 868-71.
89. Sánchez M, Rodríguez AL, Martínez LL. Desordenes lipídicos: una puesta al día [publicación periódica en Internet] 2003 n pag [Citado 1feb2007]. Disponible en URL:http://www.bvs.sld.cu/revistas/end/vol14_1_03/end07103.htm - 87k.
90. A Ríos, JM Rodríguez, NM Torregrosa, A Piñero, M Canteras, P Parrilla. Hipoparatiroidismo e hipocalcemia en el postoperatorio de la cirugía del bocio multinodular. Estudio multivariante de los factores de riesgo. *Medicina Clínica (Barcelona)* 2004;122 (10).
91. Garrido G, Madero R, Herruzo R, García J. Infección nosocomial en una unidad de cuidados intensivos: análisis multivariante de los factores de riesgo. *Medicina Clínica (Barcelona)* 1997; 108 (11).
92. Selva A, Yuste M, Armadans LI, Almirante B, San José A, Vilardel M. Factores de riesgo de infección por *Clostridium difficile* en pacientes ancianos. Estudio de casos y controles. *Medicina Clínica (Barcelona)* 2000; 115 (13).
93. Gómez-Gómez M, Danglot-Banck C, Velásquez-Jones L. Bases para la revisión crítica de artículos médicos. *Revista Mexicana de Pediatría* 2001; 68 (4).
94. Alfaro V. Ensayos clínicos, buena práctica de publicación y regulaciones legislativas. *Medicina Clínica (Barcelona)* 2004;123 (3).

95. García FJ. Impacto de la normativa legal en la calidad de los ensayos clínicos realizados en España. *Medicina Clínica* (Barcelona) 1993; 100 (20).
96. Soto J, Galende I, Sacristánb JA. Calidad de los ensayos clínicos publicados en España: valoración a través del análisis de 3 revistas durante el período 1985-1991. *Medicina Clínica* 1994; 102 (12).
97. I Galendea, JA Sacristánb, J Sotoc. Cómo mejorar la calidad de los ensayos clínicos. *Medicina Clínica* 1994; 102 (12).
98. LM Prieto, MT Illnait, EG Ramos, B Lazcano, N Márquez, N Cantelar et. al. Candidiasis oral en pacientes seropositivos al VIH y casos SIDA. Aspectos clínicos, micológicos y terapéuticos. *Revista Cubana Medicina Tropical* 2006; 58(3).
99. E. Aleixandre, MA Casanova, J. Mitjans, J Sánchez, A Sanmartín. Ensayo clínico de dos intervenciones de deshabituación tabáquica en atención primaria. *Atención Primaria* 1998; 22 (7).
100. Camp J, Artigas V, Millá J, Allende L, Domínguez R, Moreno E. Eficacia del flurbiprofeno inyectable en el tratamiento sintomático del cólico biliar. *Medicina Clínica* (Barcelona) 1992; 98 (6).
101. Aliaga A, Armijo M, Camacho F, Castro A, Cruces M, Díaz JL, et. al. Solución tópica de idoxuridina al 40 % en dimetilsulfóxido frente a aciclovir oral en el tratamiento del herpes zoster. Ensayo clínico multicéntrico a doble ciego. *Medicina Clínica* (Barcelona) 1992; 98 (7).
102. Fernández LC. Eficacia, seguridad y tolerabilidad del Policosanol en pacientes con riesgo aterosclerótico [tesis]. Centro Nacional de Investigaciones Científicas. Ministerio de Educación Superior. 1999.
103. Camarellas F, Asensioa A, Jiménez-Ruizb C, Becerril B, Rodero D, Vidaller O. Efectividad de la intervención grupal para la deshabituación tabáquica. Ensayo clínico aleatorizado. *Medicina Clínica* (Barcelona) 2002;119(2).
104. Fernández JC. Efectos del Policosanol en ancianos con hipercolesterolemia tipo II [tesis]. Centro de Productos Naturales. Centro Nacional de Investigaciones Científicas. Ministerio de Educación Superior. 2000.
105. Bermúdez IB. Actividad antimicrobiana, balance redox celular, seguridad y ventajas económicas de una crema elaborada con propóleos de Manzanillo, en pacientes de pie diabético [tesis]. Instituto de Farmacia y Alimentos. Universidad de La Habana. Ministerio de Educación Superior. 2003.
106. Montero S. Hilo quirúrgico Agasut QE como una alternativa para el cierre de aponeurosis y piel en incisiones de laparotomía [tesis]. Instituto Superior de Medicina Militar "Dr. Luis Díaz Soto". Ministerio de las Fuerzas Armadas Revolucionarias. 2004.
107. Bermúdez B. Eficacia clínica, seguridad y ventajas económicas de una crema de propóleos, en el tratamiento del pie diabético neuroinfeccioso. Instituto de Farmacia y Alimentos. Universidad de La Habana. Ministerio de Educación Superior. 2004.
108. P González, ML Legaz, F Pérez, JM López, JC Valenzuela, A Ponte et. al. Eficacia de la erradicación de *Helicobacter pylori* en la dispepsia no ulcerosa. *Medicina Clínica* 2004; 122 (3).
109. Bañosa F, Madrdejosa R, Cabezasb C, Burrullb M, Lafuenteb C, Morerac R et. al. Efectividad de la combinación de omeprazol, claritromicina y amoxicilina en la erradicación de *Helicobacter pylori* en pacientes con ulcus péptico activo: Resultados preliminares del Estudio GEHPY. *Medicina Clínica* (Barcelona) 2000; 114 (12).

110. Marrero I, Bermello, Kessell I, Mirabal F. La información computadorizada: una forma de divulgación al usuario. ACIMED 1995; 3 (1).
111. Thompson JB. La comunicación masiva y la cultura moderna. Contribución a una teoría crítica de la ideología. Versión. Estudios de comunicación y política 1991; 1(1).
112. Borgman CL. Digital libraries and the continuum of scholarly communication. Journal of Documentation 2000; 56 (4):412-430.
113. LLanusa SB. Evaluación del programa de Introducción de Tecnologías de Información y Comunicación en la Atención Primaria de Salud [tesis]. Escuela Nacional de Salud Pública. Ministerio de Salud Pública; 2004.
114. Casher O, Rzepa HS. Chemical laboratories using World-Wide Web servers and EyeChem based viewers. Journal of Molecular Graphics 1995; 13:268-271 [citado, 1 de agosto, 2000]. Disponible en <http://www.ch.ic.ac.uk/talks/acs/anaheim.html>.
115. Alonso ME, Rodríguez B, Zamora MA. Errores más frecuentes cometidos en los trabajos de terminación de la especialidad. Revista de Ciencias Médicas de La Habana 2003; 9(2).
116. Aborto en Argentina [en línea]. Buenos Aires [actualizado 12 de junio 2007; citado 25 de julio 2007]; [aprox. 3 pantallas]. Disponible en: http://www.rimaweb.com.ar/aborto/encuesta_romer.html.
117. Relación de Estudios Realizados para el Observador de la Distribución 2002 [en línea]. [actualizado 12 de junio 2007; citado 25 de julio 2007]; [aprox. 4 pantallas]. Disponible en: <http://www.cetelem.es/Institucional/EIObservador/Obervador2002/con6-0.htm#>
118. Estudio C.I.S. Nº 2.256. Encuesta Nacional de Salud (tercera oleada). Ficha técnica [en línea]. [actualizado 12 de junio 2007; citado 25 de julio 2007]; [aprox. 7 pantallas]. Disponible en: http://www.cis.es/cis/export/sites/default/-Archivos/Marginales/2240_2259/2256/FT2256.pdf.
119. Los jóvenes saben todo del condón [en línea] [actualizado 12 de junio 2007; citado 25 de julio de 2007]; [aprox. 5 pantallas]. Disponible en: <http://www.klip7.cl/blogsalud/sexo/2005/06/los-jovenes-saben-todo-del-condon-pero.html>.
120. Estudio C.I.S. Nº 2.614-2.629. Encuesta de Salud en Extremadura. Ficha técnica [en línea]. [actualizado 12 de junio 2007; citado 25 de julio 2007]; [aprox. 7 pantallas]. Disponible en: http://www.cis.es/cis/export/sites/default/-Archivos/Marginales/2600_2619/2614/ft2614-2629.pdf.
121. Il encuesta nacional sobre consumo de drogas en estudiantes de enseñanza media. Resumen preliminar abril 2006. [en línea] [actualizado 12 de junio 2007; citado 25 de julio 2007]; [aprox. 23 pantallas]. Disponible en: <http://www.conacedrogas.cl/inicio/pdf/EstudioEscolaresUruguay2005.pdf>.
122. Informe de prensa. Encuesta Latinobarómetro 1997 [en línea] [actualizado 12 de junio 2007; citado 25 de julio de 2007]; [aprox. 15 pantallas]. Disponible en: http://www.latinobarometro.org/uploads/media/1997_02.pdf.
123. Anexo 2: Ficha técnica de la encuesta y cuestionario [en línea] [actualizado 12 de junio 2007; citado 25 de julio de 2007]; [aprox. 11 pantallas]. Disponible en: <http://www.pdf.lacaixa.comunicacions.com/ee/esp/ee04anx2esp.pdf>.
124. Defensoría presenta estudio sobre niños y niñas. [en línea] [actualizado 12 de junio 2007; citado 27 de julio de 2007]; [aprox. 5 pantallas]. Disponible en: <http://www.reliefweb.int/rw/rwb.nsf/db900SID/VBOL-6WFEWM?OpenDocument>.

125. Encuesta nacional de hogares sobre medición de nivel de vida, 2001. [en línea] [actualizado 12 de junio 2007; citado 27 de julio de 2007]; [aprox. 3 pantallas].
Disponible en: <http://www.worldbank.org/html/prdph/lsm/s/country/ni2001/docs/fichatec.pdf>.
126. Encuesta de condiciones de vida de los mayores. Estudio 4.597, IMSERSO, 2004. [en línea] [actualizado 12 de junio 2007; citado 27 de julio de 2007]; [aprox. 3 pantallas].
Disponible en: <http://www.imsersomayores.csic.es/estadisticas/encuestas/index.htm>.
127. Apoyo informal a las personas mayores. Estudio 4.178, IMSERSO, 2004 [en línea] [actualizado 14 de junio de 2007; citado 27 de julio de 2007]; [aprox. 3 pantallas].
Disponible en: <http://www.imsersomayores.csic.es/estadisticas/encuestas/index.htm>.
128. Encuesta sobre Discapacidades, Deficiencias y Estado de Salud (EDDES), IMSERSO, INE, F. ONCE, 1999. [en línea] [actualizado 14 de junio de 2007; citado 27 de julio de 2007]; [aprox. 3 pantallas].
Disponible en: <http://www.imsersomayores.csic.es/estadisticas/encuestas/index.htm>.
129. La soledad en las personas mayores. Estudio 2.279. IMSERSO-CIS, 1998 [en línea] [actualizado 12 de junio 2007; citado 27 de julio de 2007]; [aprox. 3 pantallas].
Disponible en: <http://www.imsersomayores.csic.es/estadisticas/encuestas/index.htm>.
130. Apoyo informal a las personas mayores (1ª fase). Estudio 2.072, IMSERSO-CIS, 1993. [en línea] [actualizado 14 de junio de 2007; citado 2 de agosto de 2007]; [aprox. 3 pantallas].
Disponible en: <http://www.imsersomayores.csic.es/estadisticas/encuestas/index.htm>.
131. Apoyo informal a las personas mayores (2ª fase). Estudio 2.117, IMSERSO-CIS, 1994. [en línea] [actualizado 14 de junio de 2007; citado 25 de julio de 2007]; [aprox. 3 pantallas].
Disponible en: <http://www.imsersomayores.csic.es/estadisticas/encuestas/index.htm>.
132. La importancia de la salud del colectivo médico. Fundación Galatea. en línea] [actualizado 12 de junio 2007; citado 2 de agosto de 2007]; [aprox. 5 pantallas].
Disponible en: <http://www.fgalatea.org/cast/enquesta.htm>.
133. Encuesta Intención de compra a 12 meses [en línea] [actualizado 14 de junio de 2007; citado 2 de agosto de 2007]; [aprox. 3 pantallas].
Disponible en: <http://www.elcomerciooperu.com.pe/especiales/fichatecnica/apoyo.htm>.
134. Estudio Consumo de Muebles [en línea] [actualizado 14 de junio de 2007; citado 25 de julio de 2007]; [aprox. 3 pantallas].
Disponible en: <http://www.elcomerciooperu.com.pe/especiales/fichatecnica/apoyo.htm>.
135. Encuesta Consumo de Muebles y otros datos sobre Consumo en general [en línea] [actualizado 12 de junio 2007; citado 2 de agosto de 2007]; [aprox. 3 pantallas].
Disponible en: <http://www.elcomerciooperu.com.pe/especiales/fichatecnica/apoyo.htm>.
136. Encuesta Consumo Europeo [en línea] [actualizado 14 de junio de 2007; citado 27 de julio de 2007]; [aprox. 3 pantallas].
Disponible en: <http://www.elcomerciooperu.com.pe/especiales/fichatecnica/apoyo.htm>.
137. Encuesta cualitativa Consumo Europeo. [en línea] [actualizado 14 de junio de 2007; citado 25 de julio de 2007]; [aprox. 3 pantallas].
Disponible en: <http://www.elcomerciooperu.com.pe/especiales/fichatecnica/apoyo.htm>.

- 138 Encuesta Intenciones de Compra a 6 meses [en línea] [actualizado 12 de junio 2007; citado 27 de julio de 2007]; [aprox. 3 pantallas].
Disponible en: <http://www.elcomercio Peru.com.pe/especiales/fichatecnica/apoyo.htm>.
139. Ficha Técnica de la encuesta nacional de Apoyo Opinión y Mercado S.A. 19/03/2006 [en línea] [actualizado 14 de junio de 2007; citado 25 de julio de 2007]; [aprox. 2 pantallas]. Disponible en: <http://www.elcomercio Peru.com.pe/especiales/fichatecnica/apoyo.htm>.
140. Encuesta de Ocupación Hotelera [en línea] [actualizado 14 de junio de 2007; citado 2 de agosto de 2007]; [aprox. 2 pantallas]. Disponible en: http://www.ine.es/daco/fichas/ficha_eoh.htm#inicio.
141. El 95% de los uruguayos cree que hoy es difícil conseguir empleo [en línea] [actualizado 14 de junio de 2007; citado 27 de julio de 2007]; [aprox. 1 pantalla].
Disponible en: <http://www.cifra.com.uy/co260401.htm>.
142. Más del 30% de los uruguayos está pagando al menos un préstamo del sistema financiero [en línea] [actualizado 14 de junio de 2007; citado 27 de julio de 2007]; [aprox. 1 pantalla].
Disponible en: <http://www.cifra.com.uy/co170501.htm>.
- 143 Arena, arriba en simpatías [en línea] [actualizado 14 de junio de 2007; citado 27 de julio de 2007]; [aprox. 2 pantallas]. Disponible en: <http://www.laprensagrafica.com/lpgdatos/733152.asp>.
144. González LE. Empate técnico con ventaja para Batlle. El país 1999 septiembre 14.
145. Perú: Simulación de Voto 2ª Vuelta Presidencial 2006 [en línea] [actualizado 12 de junio 2007; citado 25 de julio de 2007]; [aprox. 2 pantallas].
Disponible en: <http://weblogs.elearning.ubc.ca/peru/archives/idicesimulacion23-26mayo.pdf>.
146. La encuesta nos da tres diputados y la posibilidad de ser una fuerza decisiva en el Parlament [en línea] [actualizado 12 de junio 2007; citado 27 de julio de 2007]; [aprox. 3 pantallas].
Disponible en: http://www.ciudadans-ciudadanos.net/news.php?id_pagina=23&id_post=1389.
147. Triunfo claro de Aznar en el primer debate. El País. 1993 mayo 26.
148. Tesfaye S, Chaturvedi N, Eaton S, Ward JD, Manes C, Ionescu-Tirgoviste C, et. al. Vascular Risk Factors and Diabetic Neuropathy. *New England Journal of Medicine* 2005;352 (23).
149. Berenson GS, Srinivasan SR, Bao W, Newman WP, Tracy RE, Wattigney A. Association between multiple cardiovascular risk factors and atherosclerosis in children and young adults. *New England Journal of Medicine* 1998; 338 (23).
150. Kyriacou D, Anglin D, Taliaferro E, Stone S, Tubb T, Linden JA, et. al. Risk factors for injury to women from domestic violence. *New England Journal of Medicine* 1999; 341 (20).
151. Talbot T, Hartert T, Mitchel E, Halasa N, Arbogast P, Poehling K, et. al. Asthma as a Risk Factor for Invasive Pneumococcal Disease. *New England Journal of Medicine* 2005; 352 (25).
152. Schilling F, Spix C, Berthold F, Erttmann R, N Fehse, B Hero. Neuroblastoma screening at one year of age; 2002; 346 (14).
153. Martin GS, Mannino DM, Eaton S, Moss M. The Epidemiology of Sepsis in the United States from 1979 through 2000. *New England Journal of Medicine* 2003; 348 (16).
154. Development and Validation of a Risk Score for Predicting Death in Chagas' Heart Disease. A Rassi, A Rassi, WC Little, SS Xavier, SG. Rassi, AG Rassi, et al. *New England Journal of Medicine* 2006; 355 (8).