

Introducción

Erik Cobo^a, Albert Cobos^b y Albert Sorribas^c

^aDepartamento de Estadística e Investigación Operativa. Universitat Politècnica de Catalunya. Barcelona.

^bUniversitat de Barcelona y RDES SL. Barcelona. España.

^cEstadística. Departament de Ciències Mèdiques Bàsiques. Universitat de Lleida. Lleida. España.

Los métodos estadísticos son una herramienta fundamental en la investigación médica, tanto en las etapas de diseño del estudio como en los procesos de selección de muestras y análisis de resultados¹. Últimamente hemos asistido a un cambio importante en la utilización de los métodos estadísticos por la comunidad médica española. Por una parte, la mayoría de los artículos médicos incluye un análisis estadístico² como herramienta metodológica para establecer la validez de las conclusiones del estudio. Por otra, se ha pasado de utilizar básicamente métodos univariantes a utilizar procedimientos multivariantes específicos para el problema a resolver³. Con todo ello, se ha producido un aumento importante en la sofisticación de los análisis, sin duda facilitada por la gran difusión de la microinformática y por la accesibilidad de algunos programas de análisis de datos, aunque esto no implica necesariamente su correcta aplicación⁴⁻⁸.

La producción científica de nuestro entorno ha tenido una evolución positiva⁹ ayudada, al menos en parte, por el establecimiento de unidades específicas dedicadas a la asesoría estadística, tanto en la etapa de diseño como en la de análisis de datos. Las revistas médicas han tenido un papel muy importante en este proceso y han exigido la inclusión de lo que ha dado en llamarse «evidencia». De la sinergia entre los especialistas médicos, que conocen el problema, y los especialistas metodológicos, que conocen las técnicas de análisis, surgen proyectos y resultados de mayor relevancia científica. Es nuestra obligación recordar que la calidad de las publicaciones científicas mejora y está avalada por la inclusión de estadísticos¹⁰ como autores del original, tal y como aconseja la propuesta de autoría del denominado «Grupo de Vancouver»¹¹. Esta participación ha sido clave para asegurar la calidad de los análisis y la correcta aplicación de las técnicas.

A pesar de estos cambios, el profesional médico en general se enfrenta todavía a serios problemas cuando se plantea un proyecto de investigación. Por una parte, existe un cierto desconocimiento acerca del planteamiento y del diseño del estudio que determina importantes limitaciones. A veces ocurre que los investigadores deciden un diseño más en función de la disponibilidad de los pacientes y de la actividad del servicio que en función de criterios objetivos que permitan optimizar el esfuerzo y proporcionen los datos adecuados a la finalidad del estudio. Por otra parte, aunque se ha producido un avance significativo en este aspecto, el conocimiento de las técnicas estadísticas por parte de los equipos de investigación biomédica es a veces algo superficial y se observa una cierta tendencia a utilizar métodos conocidos y adaptarlos a situaciones en las que no son los más adecuados.

Las técnicas de análisis de datos, contempladas en un sentido amplio, incluyen aproximaciones muy diversas en función del tipo de objetivos, del diseño y de las variables estudiadas. Algunas de estas técnicas se basan en metodología clásica, mientras que otras pueden parecer muy alejadas de lo que se entiende por estadística^{12,13}. En particular, la disponibilidad de ordenadores potentes y asequibles ha determinado un creciente interés por técnicas computacionales que se basan en el cálculo intensivo y que proporcionan una alternativa a los procedimientos clásicos. Un ejemplo de esto son las redes neuronales, que pueden aportar una alternativa al uso de la regresión logística en algunos casos. Sin embargo, existe un cierto desconocimiento general de las posibilidades de estas alternativas emergentes que pueden ser de mayor utilidad en algunas situaciones.

En este monográfico hemos querido ofrecer una selección de artículos sobre temas muy diversos que hemos creído importantes, ya sea por la frecuencia de su aplicación o porque son potencialmente interesantes pero poco conocidos y utilizados.

Así, este número se inicia con una visión de conjunto acerca del uso, posibilidades y problemas de las representaciones gráficas. Los gráficos permiten transmitir con claridad ciertas características de los datos disponibles. Sin embargo, es evidente que una mala utilización o un uso excesivamente subjetivo de las representaciones gráficas pueden inducir percepciones erróneas. González y Jover repasan el uso de gráficos clásicos e introducen algunas posibilidades menos habituales pero útiles para extraer información relevante. Este artículo aporta consejos en forma de guía y muestra cómo se puede mejorar la calidad de la información que se transmite por medio de un gráfico.

Dados el elevado impacto que tienen los métodos de recolección de datos en su calidad y el de ésta sobre los resultados de cualquier análisis, es llamativa la escasa atención que en general se presta a la captura de datos en la bibliografía. A modo de reflexión acerca de la importancia del proceso de recogida de datos, hemos incluido 2 trabajos que se centran en esta problemática. Por una parte, el trabajo de Velázquez, Navarro y Cobos aborda las posibilidades que ofrece la utilización de sistemas electrónicos para registrar datos de investigaciones clínicas. Las ventajas de estos sistemas son evidentes, ya que determinan una optimización en el proceso de introducción de datos en estudios prospectivos. En otras ocasiones, interesa analizar datos ya disponibles, por ejemplo, a partir de bases de datos administrativas. El trabajo de Arribas, Cirera y Tristán trata la cuestión de la concatenación o conexión de registros (*record linkage*) de diferente origen, reto de la salud pública y de la epidemiología para, respetando las leyes de protección de datos personales, explotar con ventaja la información disponible en diversos registros.

Toda investigación está basada en el análisis de datos y, en un sentido amplio, los datos siempre son medidas de los objetos de la investigación. En tanto que medidas, siempre

Correspondencia: Dr. E. Cobo.
Departamento de Estadística e Investigación Operativa.
Universitat Politècnica de Catalunya.
C/ Pau Gargallo, 5. 08025 Barcelona. España.
Correo electrónico: erik.cobo@upc.es

están afectadas de error, en mayor o menor grado. Estos errores complican notablemente el análisis de datos y, de hecho, han motivado el desarrollo de diversos métodos estadísticos. Desde un punto de vista práctico, estos errores motivan que 2 procedimientos de medida (o 2 observadores) distintos puedan discrepar a pesar de estar midiendo lo mismo. Los conceptos de validez y fiabilidad, asociados respectivamente con los de errores sistemáticos y aleatorios, son fundamentales por cuanto determinan qué información aporta una variable. En su artículo, Batista, Coenders y Alonso exponen las técnicas más actuales, basadas en la teoría estadística de la medida, que permiten cuantificar la validez y la fiabilidad de las observaciones clínicas. A su vez, Carrasco y Jover revisan las posibles aproximaciones a la evaluación de la concordancia, a través del concepto de error de medida.

Las técnicas de muestreo y su impacto en el análisis de datos constituyen un tema clásico de la estadística. Sin embargo, el llamado efecto del diseño no siempre se tiene en cuenta en investigación clínica y, aunque se ha llamado la atención acerca de su importancia en diseños aleatorizados¹⁴, sigue ignorándose en la mayoría de los estudios observacionales. El trabajo de Guillén representa una introducción a los tipos de muestreo utilizados en encuestas de salud.

Las dos contribuciones siguientes abordan sendas tareas médicas fundamentales: la clasificación diagnóstica y el pronóstico. Sanz, Sancho y Bonis resumen el estado actual de los sistemas de ayuda en la toma de decisiones clínicas exponiendo los métodos para su evaluación y revisando el rendimiento de los sistemas más utilizados. Por su parte, Muñoz trata el incipiente uso de técnicas de previsión temporal, muy extendidas por otra parte en economía o en epidemiología, para el seguimiento individualizado de pacientes. Su uso debería permitir no sólo establecer un pronóstico individualizado, sino también detectar, en un momento más temprano, los posibles desvíos del curso previsto de la evolución de la enfermedad.

Algunos de los trabajos incluidos en este monográfico presentan técnicas especiales con el objetivo de mostrar la necesidad de recurrir a métodos apropiados cuando el problema que hay que resolver se escapa de un esquema básico. Así, el trabajo de Ascaso y Abellana se centra en la estimación del riesgo de mortalidad en áreas geográficas donde la distribución espacial del riesgo sigue un patrón de agrupamiento. Este problema debe abordarse mediante métodos estadísticos que consideren la correlación entre las distintas unidades. En caso contrario, las estimaciones que se obtengan no reflejarán el estado real de la población.

El trabajo de Ruiz de Villa trata del análisis de datos correspondientes a medidas repetidas. Este tipo de datos, muy común en medicina, suele analizarse mediante técnicas inadecuadas. La autora examina las distintas opciones existentes, muestra sus limitaciones y enfatiza las ventajas de un planteamiento adecuado. En este sentido, los modelos mixtos proporcionan una aproximación adecuada y muy flexible a este tipo de problema.

Trujillano, March y Sorribas presentan la aplicación de la metodología de redes neuronales a la caracterización de la probabilidad de un suceso en función de distintas variables de predicción. Este problema suele enfocarse desde el pun-

to de vista de la regresión logística, cuyas limitaciones la hacen poco adecuada en ciertas situaciones. Las redes neuronales, basadas en cálculo intensivo, proporcionan una alternativa muy interesante para el análisis de este tipo de problemas. En este artículo se ofrece una introducción general a esta metodología y a su aplicación en el ámbito de la medicina, así como recomendaciones para la utilización de los tipos de redes más sencillos.

Finalmente, el trabajo de Moreno y Solé introduce la problemática asociada al uso de *microarrays* de ADN. Esta técnica de biología molecular permite obtener el perfil de expresión de gran cantidad de genes. Con esto se abren perspectivas muy amplias para la identificación de enfermedades y para su uso clínico en general. Sin embargo, en su estado actual, esta técnica conlleva algunos problemas metodológicos en cuanto al análisis de los resultados. En este artículo, los autores proporcionan una visión de conjunto de la técnica de *microarrays* y presentan una discusión crítica de sus ventajas y limitaciones.

Esperamos que los lectores de este monográfico descubran nuevas ideas que puedan proporcionarles alternativas de utilidad para su trabajo. La creciente complejidad de la investigación biomédica en todos los ámbitos, incluido el metodológico, determina que profesionales de distintas especialidades debe participar en un esfuerzo común. Este monográfico quiere contribuir a facilitar esta cooperación.

REFERENCIAS BIBLIOGRÁFICAS

1. Delgado-Rodríguez M, Ruiz-Canela M, De Irala-Estévez J, Llorca J, Martínez-González A. Participation of epidemiologists and/or biostatisticians and methodological quality of published controlled clinical trials. *J Epidemiol Community Health* 2001;55:569-72.
2. Mora R, Ascaso C, Sentís J. Tendencias actuales en la utilización de la estadística en Medicina. Estudio de los originales publicados en Medicina Clínica (1991-1992). *Med Clin (Barc)* 1995;104:444-7.
3. Mora R, Ascaso C, Sentís J. Uso actual de la estadística en investigación biomédica: una comparación entre revistas de medicina general. *Med Clin (Barc)* 1996;106:451-6.
4. McGuigan SM. The use of statistics in the British Journal of Psychiatry. *Br J Psychiatry* 1995;167:683-8.
5. Gardner MJ, Bond J. An exploratory study of statistical assessment of papers published in the British Medical Journal. *JAMA* 1990;263:1355-7.
6. Altman DG. Statistics in medical journals: some recent trends. *Stat Med* 2000;19:3275-89.
7. Welch LI GE, Gabbe SG. Statistics usage in the American Journal of Obstetrics and Gynecology: Has anything changed? *Am J Obstet Gynecol* 2002;186:584-6.
8. Arnau C, Cobo E, Ribera JM, Cardellach F, Selva A, Urrutia A. Efecto de la revisión estadística en la calidad de los manuscritos publicados en Medicina Clínica: estudio aleatorizado. *Med Clin (Barc)* 2003;121:690-4.
9. Camí J, Zulueta MA, Fernández MT, Bardons M, Gómez I. Producción científica española en biomedicina y ciencias de la salud durante el período 1990-1993 (Science Citation Index y Social Science Citation Index) y comparación del período 1986-1989. *Med Clin (Barc)* 1997;109:481-96.
10. Müller M, Mathews H, Altman DG. Reporting of statistical methods to adjust for confounding: a cross-sectional survey. *Ann Intern Med* 2002;136:122-6.
11. International Committee of Medical Journal Editors. Uniform Requirements for manuscripts submitted to Biomedical Journals. *Ann Intern Med* 1997;126:36-47.
12. Murray GD. How we should approach the future. *Stat Med* 1990;9:1063-8.
13. Van Houwelingen HC. The future of biostatistics: expecting the unexpected. *Stat Med* 1997;16:2773-84.
14. Kerry SM, Bland JM. Analysis of a trial randomised in clusters. *BMJ* 1998;316:54.