

Influencia de la forma de presentación de los resultados de estudios clínicos en la toma de decisiones de los estudiantes de Medicina

The Effects of Presenting Clinical Trials Results in Different Formats to Medical Students

RAÚL A. BORRACCI^{MTSAC}, EDUARDO B. ARRIBALZAGA

Recibido: 30/06/2013

Aceptado: 13/11/2013

Dirección para separatas:

Dr. Raúl A. Borracci
Bioestadística, Facultad de
Ciencias Biomédicas
Universidad Austral
Av. J. D. Perón 1500
(B1629AHJ) Derqui, Pilar
Pcia. de Buenos Aires, Argentina
e-mail: raborracci@gmail.com

RESUMEN

Introducción

Varios estudios demostraron que la presentación de los beneficios terapéuticos en forma de riesgos relativos en lugar de riesgos absolutos o número necesario a tratar produciría una impresión más favorable de dichos beneficios.

Objetivos

Conocer a través de una evaluación cognitiva el grado de influencia que podría ejercer la forma de presentación de los riesgos y los beneficios clínicos de métodos diagnósticos o tratamientos en la toma de decisiones de estudiantes de Medicina.

Material y métodos

Se estudiaron 65 alumnos que cursaban la materia Bioestadística, a los que se les presentaron los resultados resumidos de distintos ensayos clínicos, de forma de mostrar los datos como reducción del riesgo relativa (RRR), reducción del riesgo absoluta (RRA) y/o número necesario a tratar (NNT), con o sin gráficos asociados.

Resultados

Los alumnos mostraron un desempeño similar al comparar los beneficios terapéuticos, tanto cuando la presentación se realizó con riesgos relativos como con riesgos absolutos (RRR: 57,7% vs. RRA: 51,5%; $p = 0,319$); sin embargo, el desempeño fue peor cuando los datos se expresaron como NNT (RRR: 57,7% vs. NNT: 31,3%; $p = 0,000002$). Por su parte, la inclusión de gráficos con escala modificada fue interpretada por los alumnos como una diferencia real en los datos (RRR: 98,5% vs. RRA: 43,1%; $p < 0,000001$).

Conclusiones

Este estudio demostró los riesgos relacionados con la mala interpretación de los resultados estadísticos, así como la necesidad de insistir en el entrenamiento de los alumnos en este tipo de análisis cuantitativo, a fin de mejorar el proceso de toma de decisiones médicas.

REV ARGENT CARDIOL 2014;82:225-230. <http://dx.doi.org/10.7775/rac.es.v82.i3.2847>

Palabras clave > Bioestadística - Magnitud de efecto - Toma de decisiones - Resultados de estudios clínicos - Educación médica

Abreviaturas >

IC _{95%}	Intervalo de confianza del 95%	RRA	Reducción del riesgo absoluta
NNT	Número necesario a tratar	RRR	Reducción del riesgo relativa
OR	Odds ratio		

INTRODUCCIÓN

La forma de presentar los resultados de los ensayos clínicos parece tener influencia en el proceso de toma de decisiones médicas, efecto que se manifiesta tanto en los médicos como en los pacientes. (1-3) Los estudios

sobre psicología cognitiva demostraron hace ya décadas que la forma en que se presente la información cuantitativa tiene un profundo efecto sobre su interpretación posible. (4, 5) Es decir que la interpretación de la información numérica puede divergir según el formato de la presentación de los datos. La importancia que

tiene la forma de mostrar los resultados de un ensayo clínico puede resumirse con un ejemplo sencillo. Una reducción del riesgo relativa del 25% puede generarse tanto a partir de una diferencia en la tasa de eventos de dos grupos de 40% y 30%, como también a partir de una diferencia entre 4% y 3%, lo que en el primer caso representa una diferencia absoluta del 10% y en el segundo de solo el 1%. Si los datos se presentaran solo como reducción relativa del riesgo, existiría una tendencia a aceptar estas diferencias como de importancia clínica, obviando o ignorando la verdadera reducción de la tasa de eventos. Varios estudios demostraron que la presentación de los beneficios terapéuticos en forma de riesgos relativos produciría en los médicos una impresión más favorable de dichos beneficios que si la presentación de los mismos resultados se hiciera en forma de riesgos absolutos o número necesario a tratar. En un metaanálisis reciente, Covey (6) observó estos mismos efectos, aunque en forma heterogénea, probablemente debido a diferencias en las metodologías usadas en cada estudio individual. Dada la importancia que tendría la forma de comunicar los resultados en la toma de decisiones de los médicos, y teniendo en cuenta la heterogeneidad de los estudios previos y la importancia de la enseñanza de estos aspectos, se desarrolló este estudio con el objetivo de conocer a través de una evaluación cognitiva el grado de influencia que podría ejercer la forma de presentación de los beneficios clínicos de métodos diagnósticos o terapéuticos en la toma de decisiones de un grupo de estudiantes de Medicina que cursaban la materia Bioestadística.

MATERIAL Y MÉTODOS

Se estudiaron 65 alumnos que cursaban la materia Bioestadística en el segundo año de la carrera de Medicina durante 2012. Fueron sometidos a una evaluación de sus conocimientos de bioestadística aprendidos durante la cursada de la materia. El examen contuvo una primera sección en la que se evaluó el conocimiento sobre las pruebas de hipótesis y una segunda sobre pruebas de magnitud de efecto. Esta segunda prueba estuvo compuesta por nueve preguntas con respuesta de opción múltiple. Las preguntas presentaron los resultados clínicos resumidos de distintos ensayos controlados, de tal forma que los datos podían mostrarse como reducción del riesgo relativa (RRR), reducción del riesgo absoluta (RRA) y/o número necesario a tratar (NNT), con el agregado de gráficos en algunos casos. La serie de nueve preguntas se presentó en un orden aleatorio y estuvo compuesta de la siguiente manera:

Grupo 1 (planteo de la primera pregunta de cuatro formas distintas): esta pregunta planteó los resultados con RRR y NNT; la misma pregunta se reelaboró para expresarla solo con RRR; luego la misma pregunta planteada con RRA y finalmente la misma solo con NNT.

Grupo 2 (planteo de la segunda pregunta de tres formas distintas): esta pregunta planteó los resultados solo con RRA; luego la misma pregunta se reelaboró para presentarla solo con RRR y finalmente la misma solo con NNT.

Grupo 3 (planteo de la tercera pregunta de dos formas distintas): esta pregunta presentó los resultados solo con RRR y luego la misma pregunta solo con RRA, acompañadas de sendas representaciones gráficas. En este caso, la escala o proporción relativa de los gráficos fue modificada a fin de

favorecer la elección de la RRR. La modificación de los gráficos consistió en exagerar o reducir la altura de las barras para favorecer o desfavorecer las diferencias relativas. El contenido resumido de las preguntas se muestra en la Tabla 1. A los fines comparativos, las preguntas de la prueba se confeccionaron a partir de tres estudios previos que abordaron la misma temática que este estudio. (7-9) La composición de estas preguntas intentó subsanar los inconvenientes de los estudios previos, sobre todo mediante el agregado de ejemplos que contenían NNT, la combinación RRR + NNT y gráficos de apoyo. Al final de cada ejemplo se le preguntaba al alumno si consideraría estos resultados como una evidencia suficiente a favor de la intervención propuesta. La opción múltiple de respuestas a cada pregunta permitió responder en unas ocasiones con "sí, no o no tengo información suficiente" y en otras ocasiones con "de mucha utilidad, de moderada o poca utilidad, de ninguna utilidad o no tengo información suficiente".

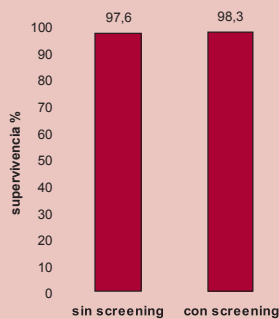
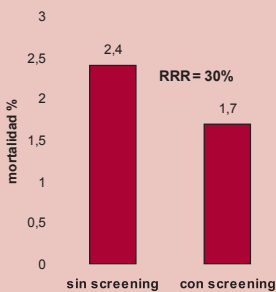
Para el análisis de los resultados se comparó la tasa de elección favorable para las RRR *versus* las RRA o NNT. La hipótesis propuesta fue que los encuestados elegirían más fácilmente a las RRR que a las RRA o NNT como evidencia a favor de un tratamiento. Por su parte, las preguntas de los grupos 1 y 2 se evaluaron en conjunto; por lo tanto, el número total de respuestas fue 130 (65 alumnos por 2 preguntas cada uno) para el caso de RRA, 130 (65 alumnos por 2 preguntas cada uno) para el caso de RRR y 195 (65 alumnos por 3 preguntas) para el caso de NNT y RRR + NNT (2 preguntas en el grupo 1 y una en el grupo 2).

Por último, el puntaje individual obtenido en esta prueba basada en la magnitud del efecto se comparó con el conseguido en la sección que evaluó los conocimientos sobre pruebas de hipótesis, a fin de analizar si la comprensión de uno y otro tema estuvo relacionada. Los grupos se compararon con χ^2 y se calculó el logaritmo del *odds ratio* (ln OR) con su intervalo de confianza del 95% (IC_{95%}), a fin de poder comparar estos resultados con los publicados previamente con el mismo formato. La distribución de las respuestas correctas se ajustó con Poisson y la correlación entre los exámenes con *rho* de Spearman.

RESULTADOS

Los puntajes de la prueba de influencia de las medidas de magnitud del efecto siguieron una distribución de Poisson con $\lambda = 3,6$ (p de ajuste = 0,706) (Figura 1). De la misma manera, los puntajes de la prueba sobre pruebas de hipótesis también coincidieron con una distribución de Poisson con $\lambda = 4,8$ (p de ajuste = 0,095). En el análisis conjunto de las preguntas de los grupos 1 y 2, el 51,5% (67/130) de las respuestas fue correcto cuando se presentaron los datos como RRA; por su parte, el 57,7% (75/130) de las respuestas fue también correcto al mostrar los resultados como RRR; por último, el 37,7% (61/195) fue correcto cuando los datos se expresaron como NNT (p < 0,0001). La comparación entre las respuestas correctas con RRR y RRA no mostró diferencia estadística (ln OR = 0,25, IC_{95%} -0,274-0,770; p = 0,319), aunque sí hubo al comparar RRA y NNT (31,3%, 61/195) (ln OR = 1,10, IC_{95%} 0,610-1,585; p = 0,00025), lo que indicaría que los alumnos manejan mejor la evidencia estadística con la RRA que con su equivalente NNT. Finalmente, la comparación entre RRR y NNT fue ln OR = 0,85, IC_{95%} 0,365-1,332; p = 0,000002. En el grupo 3, cuyas preguntas se presentaron con gráficos con escala mo-

Tabla 1. Resumen de las preguntas (7-9)

RRA	RRR	NNT
G1. 1.900 hombres que recibieron una droga para bajar el colesterol se compararon con otros 1.906 que recibieron placebo en un ensayo clínico controlado. Tras 7 años de seguimiento, la tasa de muerte cardiovascular fue del 2% para el grupo placebo y del 1,6% en el que se administró la droga, lo que significó una reducción del riesgo absoluto del 0,4% (estadísticamente significativa). G2. Una intervención médica generó una reducción absoluta del 1,4% (2,5% vs. 3,9%) en la incidencia a los 5 años de infarto mortal y no mortal (se trató de un ensayo clínico controlado y la diferencia estadística fue significativa). G3. La aplicación de cierto método diagnóstico no invasivo para el <i>screening</i> de una población arrojó los siguientes resultados de supervivencia:  El uso de <i>screening</i> mejora la supervivencia en un valor absoluto de 0,7% (98,3% vs. 97,6%)	G1. Un ensayo clínico controlado de casi 4.000 varones con hipercolesterolemia tratados con una droga o placebo mostró una reducción relativa del riesgo del 20% en la tasa de muerte cardiovascular a los 7 años. Esta diferencia a favor de la droga fue estadísticamente significativa. G2. Una intervención médica generó una reducción relativa del 36% en la incidencia de infarto de miocardio mortal y no mortal a los 5 años (se trató de un ensayo clínico controlado y la diferencia estadística fue significativa). G3. La aplicación de cierto método diagnóstico no invasivo para el <i>screening</i> de una población generó los siguientes resultados de mortalidad en el seguimiento alejado:  El uso de <i>screening</i> reduce la chance de morir en un 30%	G1. En un ensayo clínico controlado, 3.806 pacientes se dividieron en dos grupos para comparar una droga <i>versus</i> placebo. La evaluación de la tasa de muerte cardiovascular mostró una diferencia estadísticamente significativa a favor de la droga, con un número necesario a tratar de 250 pacientes para evitar una muerte a los 7 años. G2. Cierta intervención médica requiere tratar 72 personas durante 5 años para evitar un infarto de miocardio mortal o no mortal (se trató de un ensayo clínico controlado y la diferencia estadística fue significativa).
NNT + RRR G1. Alrededor de 4.000 personas participaron de un estudio clínico controlado para evaluar una droga contra placebo, en la hipercolesterolemia. Los resultados a los 7 años favorecieron al tratamiento con la droga, con una reducción relativa de la tasa de muerte cardiovascular del 20% (estadísticamente significativa), y un número necesario a tratar para evitar un evento de 250.		

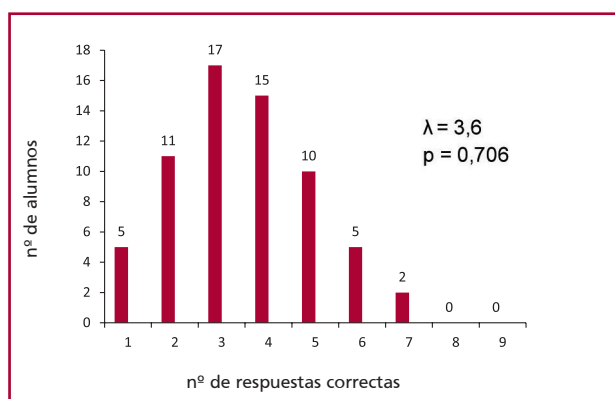


Fig. 1. Distribución de las respuestas correctas en la prueba de influencia de los riesgos relativos y absolutos en la toma de decisiones (el valor de λ corresponde a una distribución de Poisson y el nivel de p , a la prueba de bondad de ajuste).

dificada, el 98,5% (64/65) de los alumnos aceptó como recomendable el método de *screening* propuesto cuando los datos se presentaron en primer lugar como RRR; sin embargo, cuando se planteó el mismo problema con RRA, el 43,1% (28/65) aceptó el método como útil (ln OR = 4,44, IC_{95%} 2,434-7,461; $p < 0,000001$) (Figura 2). Por último, en la Figura 3 se muestra la correlación de Spearman entre las notas obtenidas en el examen sobre pruebas de hipótesis y las notas de la prueba sobre medidas de magnitud del efecto.

DISCUSIÓN

Aunque algunas revistas médicas en la actualidad exigen la publicación de las verdaderas tasas de eventos en forma de RRA y/o NNT, no existe una política uniforme para presentar los resultados de los estudios clínicos, ya sea por parte de los investigadores o de la

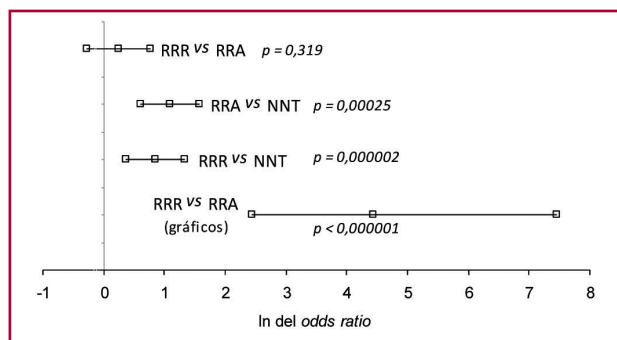


Fig. 2. Análisis de los resultados de acuerdo al $\ln$ del odds ratio y su intervalo de confianza del 95% para cada tipo de comparación. NNT: Número necesario a tratar. RRA: Reducción del riesgo absoluta. RRR: Reducción del riesgo relativa.

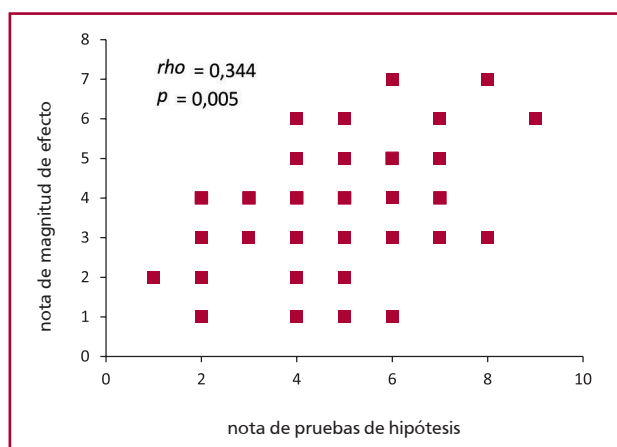


Fig. 3. Correlación de Spearman entre las notas obtenidas en el examen sobre pruebas de hipótesis y las notas de la prueba sobre riesgos relativos y absolutos.

industria farmacéutica cuando realiza publicidad. (8, 10) La “ilusión del beneficio” que se genera al mostrar los resultados como RRR tiene fuerte implicación en la aceptación y adopción de las innovaciones. En nuestro trabajo, a pesar de que los alumnos habían sido entrenados para reconocer las diferentes presentaciones de las medidas de magnitud del efecto, se encontraron fallas en la interpretación de los resultados. Por un lado, el hecho de que los estudiantes hayan reconocido correctamente un porcentaje similar de respuestas con RRA (51,5%) y con RRR (57,7%) podría interpretarse que fueron capaces de diferenciar entre estas medidas, aunque con errores de interpretación de los planteos de los problemas o de la importancia clínica de las diferencias. La tasa de errores cometidos cuando los resultados se presentaron como NNT indicaría que los alumnos tuvieron dificultad para relacionar esta medida con la RRA ($NNT = 1 / RRA$). La mayor cantidad de errores se produjo en las interpretaciones gráficas.

Las figuras modificadas intencionalmente generaron la confusión más significativa en el estudio. La exageración de las proporciones relativas en los gráficos de barras fue fácilmente confundida con una diferencia real de los datos.

El entrenamiento de los estudiantes y de los médicos en el análisis cuantitativo de las técnicas estadísticas es esencial para corregir estos problemas y mejorar el proceso de toma de decisiones. Los resultados de otros estudios y del nuestro muestran que el aprendizaje de estos métodos como se realiza hasta el momento no parece ser suficiente ni adecuado. (11-13) En este sentido, el metaanálisis de Covey (6) concluyó que tanto los médicos como los estudiantes y el público pueden ser afectados por igual por la forma de presentación de los resultados; es decir que el entrenamiento médico y la experiencia clínica no son suficientes para inmunizar a las personas contra la tendencia a interpretar en forma diferente cuando los datos se presentan como RRR, RRA o NNT. Esta dificultad para adquirir un conocimiento adecuado de la matemática o la estadística para interpretar la toma de decisiones en salud se ha estudiado intensamente bajo la denominación de *numeracy*. (14, 15) Lo importante de este concepto radica en que la enseñanza y la comprensión de las herramientas estadísticas podrían no estar al alcance de muchos decisores en salud y de los pacientes en particular que quisieran participar en una decisión autónoma sobre su salud. (16)

Solamente cuatro estudios previos han analizado la influencia en la forma de presentación de los beneficios o riesgos terapéuticos en estudiantes de Medicina o Farmacia, observándose una amplia dispersión de los resultados. (4, 17-19) En conjunto, el $\ln$ OR ponderado fue 2,54 al comparar RRR *versus* RRA y 2,16 al enfrentar RRR *versus* NNT, lo que indicaría una mayor chance de aceptar un resultado como válido cuando se presenta como riesgo relativo. Aunque estos hallazgos en conjunto difieren de los observados en nuestro trabajo, al considerar los resultados individuales de Lacy y colaboradores (18) y Chao y colaboradores (19) las diferencias se hacen mínimas.

Limitaciones

Entre las limitaciones de este estudio se destaca la imposibilidad de extender las conclusiones a toda la población de estudiantes de Medicina, ya que la muestra fue de solo 65 individuos pertenecientes a una única Universidad. De la misma forma, la Escuela de Medicina donde se realizó el estudio es una de las pocas que imparte la materia Bioestadística en el pregrado.

CONCLUSIONES

La evaluación de una muestra de alumnos que cursaban la materia Bioestadística de la carrera de Medicina mostró un desempeño similar al comparar los beneficios terapéuticos observados en una serie de ensayos clínicos, tanto cuando la presentación se realizó con riesgos relativos como con riesgos absolutos; sin embargo, el desempeño fue peor cuando los datos

se expresaron como número necesario a tratar. Por su parte, la inclusión de gráficos con escala modificada fue interpretada por los alumnos como una diferencia real en los datos. Este estudio demostró los riesgos relacionados con la mala interpretación de los resultados estadísticos, así como la necesidad de insistir en el entrenamiento de los alumnos en este tipo de análisis cuantitativo, a fin de mejorar el proceso de toma de decisiones médicas.

CONCEPTOS BÁSICOS

Los ensayos clínicos suelen presentar sus resultados en forma de riesgos relativos y/o riesgos absolutos.

Según se elija una u otra forma de presentación, se podrían malinterpretar los beneficios y los riesgos de un tratamiento.

Los estudiantes de Medicina suelen confundir el significado de riesgos relativos y absolutos al tomar decisiones.

El entrenamiento de los estudiantes en el análisis cuantitativo de las técnicas estadísticas es esencial para corregir estos problemas y mejorar el proceso de toma de decisiones.

ABSTRACT

The Effects of Presenting Clinical Trials Results in Different Formats to Medical Students

Introduction

Previous studies have demonstrated that presenting treatment benefits in terms of relative risk reduction rather than in terms of absolute risk reduction or number of needed to treat patients, should favor the sense of outcome effectiveness.

Purposes

The purpose of this study was to perform a cognitive evaluation to assess how the form of presenting the risks and benefits of screening methods and treatments affects medical students' decision-making.

Methods

Sixty-five medical students attending a Biostatistics course answered a questionnaire reporting the results of clinical trials expressed as relative risk reduction (RRR), absolute risk reduction (ARR) and/or number needed to treat (NNT), with or without associated graphs.

Results

Students' performance was similar when comparing treatment benefits, both in relative and absolute risks presentations (RRR: 57.7% vs. ARR: 51.5%, $p=0.319$); however, performance was worse when information was expressed as NNT (RRR: 57.7% vs. NNT: 31.3%, $p=0.000002$). Inclusion of modified-scale graphs was misinterpreted as a real data difference (RRR: 98.5% vs. ARR: 43.1%, $p<0.000001$).

Conclusions

This study demonstrated the risks related to misinterpretation of statistical results, and the need to improve the students' training in this type of quantitative analysis, in order to improve the medical decision-making process.

Key words > Biostatistics - Measure of Effect - Decision-making - Clinical Trials Results - Medical Education

Declaración de conflicto de intereses

El Dr. Borracci se desempeña como docente de la Universidad Austral.

El Dr. Arribalzaga se desempeña como docente de la materia Bioestadística en la Universidad Austral.

BIBLIOGRAFÍA

- McGettigan P, Sly K, O'Connell D, Hill S, Henry D. The effects of information framing on the practices of physicians. *J Gen Intern Med* 1999;14:633-42. <http://doi.org/d5h79q>
- Edwards A, Elwyn G, Covey J, Matthews E, Pill R. Presenting risk information: a review of the effects of "framing" and others manipulations on patients outcomes. *J Health Commun* 2001;6:61-82. <http://doi.org/cz49dh>
- Moxey A, O'Connell D, McGettigan P, Henry D. Describing treatment effects to patients: how they are expressed makes a difference. *J Gen Intern Med* 2003;18:948-59. <http://doi.org/d87vhs>
- Tversky A, Kahneman D. The framing of decisions and the rationality of choice. *Science* 1981;211:453-8. <http://doi.org/fj3z3r>
- Kahneman D, Tversky A. Choices, values, and frames. *Am Psychologist* 1984;39:341-50. <http://doi.org/b4mb2g>
- Covey J. A meta-analysis of the effects of presenting treatment benefits in different formats. *Med Decis Making* 2007;27:638-654. <http://doi.org/b9f5hn>
- Forrow L, Taylor WC, Arnold RM. Absolutely relative: how research results are summarized can affect treatment decisions. *Am J Med* 1992;92:121-4. <http://doi.org/bm2j5p>
- Naylor CD, Chen E, Strauss B. Measured enthusiasm: does the method of reporting trial results alter perceptions of therapeutic effectiveness? *Ann Intern Med* 1992;117:916-21. <http://doi.org/qzww>
- Wolf AMD, Schorling JB. Does informed consent alter elderly patients' preferences for colorectal cancer screening? Results of a randomized trial. *J Gen Intern Med* 2000;15:24-30. <http://doi.org/brj99w>
- Dryver E, Hux JE. Reporting of numerical and statistical differences in abstracts: improving but not optimal. *J Gen Intern Med* 2002;17:203-6. <http://doi.org/b7zr52>
- Woloshin S, Schwartz LM. Press releases: translating research into news. *JAMA* 2002;287:2856-8. <http://doi.org/d4xkbp>
- Moynehan RL, Bero L, Ross-Degnan R, Henry D, Lee K, Watkins J, et al. Coverage of the news media of the benefits and risks of medications. *N Engl J Med* 2000;342:1645-50. <http://doi.org/bt58db>
- Thomson R, Edwards A, Grey J. Risk communication in the clinical consultation. *Clin Med* 2005;5:465-9. <http://doi.org/qzxx>
- Nelson W, Reyna VF, Fagerlin A, Lipkus I, Peters E. Clinical implications of numeracy: Theory and practice. *Ann Behav Med* 2008;35:261-74. <http://doi.org/cg4qcc>
- Lexchin J. Statistics can affect treatment decisions. *Can Fam Physician* 2003;49:1587-8.

16. Reyna VF, Nelson WL, Han PK, Dieckmann NF. How numeracy influences risk comprehension and medical decision making. *Psychol Bull* 2009;135:943-73. <http://doi.org/fsz693>
17. Damur C, Steurer J. Do doctors judge therapy results differently from students? *Schweizerische Medizinische Wochenschrift* 2000;130:171-6.
18. Lacy CR, Barone JA, Suh DC, Malini PL, Bueno M, Moylan DM. Impact of presentation of research results on likelihood of prescribing medications to patients with left ventricular dysfunction. *Am J Cardiol* 2001;87:203-7. <http://doi.org/fg9vqz>
19. Chao C, Studts JL, Abell T, Hatley T, Roetzer L, Dinne S, et al. Adjuvant chemotherapy for breast cancer: how presentation of recurrence risk influences decision-making. *J Clin Oncol* 2003;21:4299-305. <http://doi.org/cjn364>