

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

RAÚL A. BORRACCI<sup>MTSAC</sup>, EDUARDO B. ARRIBALZAGA

Received: 06/30/2013

Accepted: 11/13/2013

**Address for reprints:**

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

## ABSTRACT

### 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.

### Objectives

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 student 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 risk presentations (RRR: 57.7% vs. ARR: 51.5%,  $p = 0.319$ ); however, performance was worse when information was expressed as NNT (RRR: 57.7% versus 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 perfect students' training in this type of quantitative analysis in order to improve the medical decision-making process.

REV ARGENT CARDIOL 2014;82:210-214. <http://dx.doi.org/10.7775/rac.v82.i3.2847>

**Key words >** Biostatistics - Effect Size - Decision-making - Clinical Trials Results - Medical Education

**Abbreviations >**

<b>AAR</b>	Absolute risk reduction	<b>OR</b>	Odds ratio
<b>95%CI</b>	confidence interval	<b>RRR</b>	Relative risk reduction
<b>NNT</b>	Number needed to treat		

## INTRODUCTION

The method of reporting clinical trial results seems to influence the process of medical decision-making, an effect which is manifested both in doctors and in patients. (1-3). For decades, cognitive psychology studies have shown that the method in which quantitative information is expressed has a profound effect on its possible interpretation (4 -5). This means that the interpretation of numerical information may

differ depending on the format of data presentation. The importance of how to display clinical trial results may be summarized with a simple example. A relative risk reduction of 25 % may be generated both from a difference in the event rate of two groups of 40% and 30 %, as well as from a difference between 4% and 3 %; which in the first case is an absolute difference of 10% and in the second of only 1%. If data were expressed only as relative risk reduction (RRR),

there would be a tendency to accept these differences as clinically important, bypassing or ignoring the true reduction in the event rate. Several studies have shown that the presentation of therapeutic benefits as RRR rather than as absolute risk reduction (ARR) or number needed to treat (NNT) would give physicians a more favorable impression of these benefits. In a recent meta-analysis Covey (6) observed these same effects, albeit heterogeneously, probably due to differences in the procedures used in each individual study. Given the importance of result reporting methodologies in medical decision-making, and taking into account the heterogeneity of previous studies and the relevance of teaching these aspects, the following study was undertaken to learn through a cognitive assessment, the degree of influence that the mode of presentation of clinical benefits arising from diagnostic and therapeutic methods might exert in the decision making of a group of medical

students who were attending a Biostatistics course.

**METHODS**

The study included 65 students who attended the Biostatistics course during the second year of the medical career in 2012. They were subjected to an assessment of the biostatistics knowledge acquired during the course. The first test contained a first section which assessed knowledge of hypothesis tests and a second test the effect size. This second test consisted of nine questions with multiple choice answers. The questions presented the summarized results of different controlled clinical trials, with data displayed as RRR, ARR and / or NNT with the addition of graphs in some cases. The series of nine questions was randomized and organized as follows:

Group 1 (first question proposed in four different ways): in this question results were expressed as RRR and NNT; the same question was reworked to express them only as RRR; then the same question was expressed as ARR and finally only as NNT.

Group 2 (second question proposed in three different

**Table 1.** Question's summary (7-9)

ARR	RRR	NNT														
<p><b>G1.</b> In a controlled clinical trial 1900 men who received a drug to lower cholesterol were compared with 1906 men who received placebo. After 7 years follow up, the rate of cardiovascular death was 2% in the placebo group and 1.6 % in the drug group, which meant an absolute risk reduction of 0.4 % (statistically significant).</p>	<p><b>G1.</b> A controlled clinical trial of nearly 4000 men with hypercholesterolemia treated with drug or placebo showed a relative risk reduction of 20 % in cardiovascular death rate at 7 years. This difference in favor of the drug was statistically significant.</p>	<p><b>G1.</b> In a controlled clinical trial, 3806 patients were divided into two groups to compare a drug versus placebo. Assessment of cardiovascular death rate showed a statistically significant difference in favor of the drug, with a number needed to treat of 250 patients to prevent one death at 7 years.</p>														
<p><b>G2.</b> A medical intervention had an absolute reduction of 1.4 % (2.5 % vs. 3.9%) in the 5-year incidence rate of fatal and non- fatal myocardial infarction (this was a controlled clinical trial with significant statistical difference).</p>	<p><b>G2.</b> A medical intervention generated 36 % relative risk reduction in the incidence of fatal and non- fatal myocardial infarction at 5 years (it was a controlled clinical trial with significant statistical difference).</p>	<p><b>G2.</b> A medical intervention requires treating 72 subjects for 5 years to prevent fatal or non-fatal myocardial infarction (it was a controlled clinical trial with significant statistical difference).</p>														
<p><b>G3.</b> The application of a certain noninvasive diagnostic method for a population screening yielded the following survival results:</p>	<p><b>G3.</b> The application of a certain noninvasive diagnostic method for a population screening yielded the following death results in the long term follow-up</p>	<p><b>NNT+RRR</b></p> <p><b>G1.</b> Nearly 4000 subjects participated in a controlled clinical trial evaluating a drug vs. placebo in hypercholesterolemia. Drug treatment favored results at 7 years, with a statistically significant relative cardiovascular death rate reduction of 20% and a number needed to treat of 250 to prevent an event.</p>														
<table border="1"> <caption>Survival %</caption> <tr><th>Screening Status</th><th>Survival %</th></tr> <tr><td>without screening</td><td>97.6</td></tr> <tr><td>with screening</td><td>98.3</td></tr> </table>	Screening Status	Survival %	without screening	97.6	with screening	98.3	<table border="1"> <caption>Mortality %</caption> <tr><th>Screening Status</th><th>Mortality %</th></tr> <tr><td>without screening</td><td>2.4</td></tr> <tr><td>with screening</td><td>1.7</td></tr> <tr><td>RRR</td><td>30%</td></tr> </table>	Screening Status	Mortality %	without screening	2.4	with screening	1.7	RRR	30%	
Screening Status	Survival %															
without screening	97.6															
with screening	98.3															
Screening Status	Mortality %															
without screening	2.4															
with screening	1.7															
RRR	30%															
<p>The use of screening improves survival by an absolute value of 0.7% (98.3% vs. 97.6%)</p>	<p>The use of screening reduces mortality risk by 30%</p>															

ways): in this question results were only expressed as ARR; then the same question was reworked to express them only as RRR and finally only as NNT.

Group 3 (third question proposed in two different ways): in this question results were only expressed as RRR and then the same question only as ARR, both accompanied by graph representations. In this case, the scale or the graphs' relative proportion was modified in order to favor the choice of RRR. The graph modification consisted in exaggerating or reducing bar heights to favor or disfavor relative differences. Question contents are summarized in Table 1. For comparison purposes, the test questions were prepared from three previous studies that addressed the same topic of this study. (7 -9) The composition of these questions attempted to remedy previous study shortcomings, especially by adding examples containing NNT, the combination of RRR + NNT and supporting graphs. At the end of each sample the student was asked whether he would consider these results as sufficient evidence for the proposed intervention. The multiple choice answers to each question allowed on some occasions to answer "yes, no or do not have enough information"; and at other times "very helpful, moderately or scarcely helpful, not helpful or do not have enough information."

To analyze the results, the favorable rate of choice for RRRs vs. ARRs or NNTs was compared. The proposed hypothesis was that respondents would choose more easily RRR than ARR or NNT as treatment evidence. Questions in groups 1 and 2 were evaluated together; therefore, the total number of answers was 130 (65 students per 2 questions each) in the case of ARR, 130 (65 students per 2 questions each) in the case of RRR and 195 (65 students per 3 questions) in the case of NNT and RRR + NNT (2 questions in group 1 and one question in group 2).

Finally, the individual score obtained in this test based on the effect size was compared with that achieved in the section that assessed knowledge of hypothesis tests in order to analyze whether the understanding of both was related. The groups were compared with  $\chi^2$  and the odds ratio logarithm (ln OR) and its 95% confidence interval (95% CI) was estimated in order to compare these results with those previously reported in the same format. The distribution of correct answers was adjusted with Poisson distribution and the correlation between the tests with Spearman's rho.

## RESULTS

The scores of the test analyzing the influence of effect size measurements followed a Poisson distribution with  $\lambda = 3.6$  (adjusted  $p = 0.706$ ) (Figure 1). Similarly, the test scores on hypothesis tests also coincided with a Poisson distribution with  $\lambda = 4.8$  (adjusted  $p = 0.095$ ). In the joint analysis of the questions in groups 1 and 2, 51.5% (67/130) of the answers was correct when results were expressed as ARR; likewise, 57.7% (75/130) was also correct when results were expressed as RRR; finally, 37.7% (61/195) was correct when data were expressed as NNT ( $p < 0.0001$ ). The comparison between correct answers with RRR and ARR showed no statistical difference (ln OR = 0.25, 95% CI 0.274-0.770;  $p = 0.319$ ); although there was a difference when ARR and NNT were compared (31.3%, 61/195) (ln OR = 1.10, 95% CI 0.610-1.585;  $p = 0.00025$ ), indicating

that students handled the evidence statistics better with ARR than with its equivalent NNT. Finally, the comparison between RRR and NNT was ln OR = 0.85, 95% CI 0.365 -1.332;  $p = 0.000002$ . In group 3, where questions were expressed with rescaled graphs, 98.5% (64/65) of the students accepted as recommended the proposed screening method where data was first expressed as RRR; however, when the same problem was proposed as ARR, 43.1% (28/65) accepted the method as useful (ln OR = 4.44, 95% CI 2.434-7.461;  $p < 0.000001$ ) (Figure 2).

Finally, Figure 3 shows Spearman's correlation between the marks obtained in the exam on hypothesis tests and in the exam on effect size measurements.

## DISCUSSION

Although some medical journals now require the publication of the true event rates as ARR and / or NNT, there is no uniform policy to present the results of clinical studies, either by researchers or the pharmaceutical industry when advertising. (8, 10) The "illusion of benefit" generated when results are expressed as RRR has strong implications on the acceptance and adoption of innovations. In our work, although students had been trained to recognize the different presentations of effect size measurements, flaws in the interpretation of results were found. On the one hand, the fact that students have correctly recognized a similar response rate with ARR (51.5%) and RRR (57.7%) could be interpreted as their ability to differentiate between these measurements, albeit with errors in the interpretation of problem presentation or in the clinical significance of differences. The rate of errors made when results were expressed as NNT indicate that students had difficulty relating this measurement with ARR (NNT = 1 / ARR). Most errors occurred in graph interpretations. Figures intentionally modified generated the most significant confusion in the study. The exaggeration of the relative proportions in the bar graphs was easily

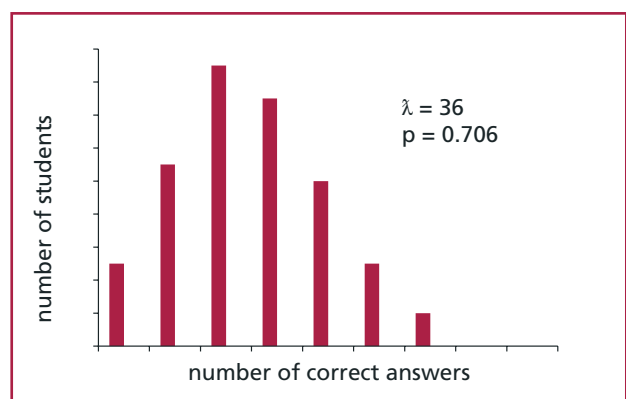


Fig. 1. Distribution of correct answers in the influence test of relative and absolute risks in decision-making (the  $\lambda$  value corresponds to a Poisson distribution and the  $p$  level to the goodness-of-fit test)

mistaken for a real difference in the data.

Students' and doctors' training in quantitative analysis of statistical techniques is essential to correct these problems and improve the decision making process. So far, other study results, as well as ours, have shown that these learning methods do not seem to be sufficient or adequate. (11-13) In this sense, Covey's meta-analysis (6) concluded that both doctors and students and the general public may be equally affected by the presentation of results; i.e. medical training and clinical experience are not enough to immunize people against the tendency to interpret data in a different way when they are expressed as RRR, ARR or NNT. This inability to acquire adequate knowledge of mathematics or statistics to interpret decision making in health has been intensively studied under the name of numeracy. (14, 15) The importance of this concept is that teaching and understanding statistical tools might not be available to many decision makers in health, and in particular to patients who would like to take part in an autonomous decision regarding their health. (16)

Only four previous studies have analyzed the influence in the presentation of the therapeutic benefits or risks in Medicine or Pharmacy students, showing a wide dispersion of results. (4, 17-19) Overall, the estimated In OR was 2.54 when comparing RRR vs. ARR and 2.16 when comparing RRR vs. NNT, indicating a greater chance of accepting a result as valid when expressed as relative risk. Although these findings as a whole differ from those observed in our work, when considering the individual results of Lacy et al. (18) and Chao et al. (19) the differences are negligible.

#### Limitations

Among the limitations of this study we may underline the inability to extend the findings to the entire population of medical students, as the sample only comprised 65 individuals belonging to a single university. Likewise, the School of Medicine where the study was performed is one of the few which teaches Biostatistics as an undergraduate subject.

#### CONCLUSIONS

In conclusion, the evaluation of a sample of students who attended a biostatistics course during their medical career showed a similar performance when comparing the therapeutic benefits observed in a number of clinical trials, both when the presentation was expressed as RRR and as ARR; however, the performance was worse when data were expressed as NNT. In addition, the inclusion of graphs with a modified scale was interpreted by students as a real data difference. This study demonstrated the risks associated with the misinterpretation of statistical results and emphasizes the need for student training in this type of quantitative analysis to improve the process of medical decision making

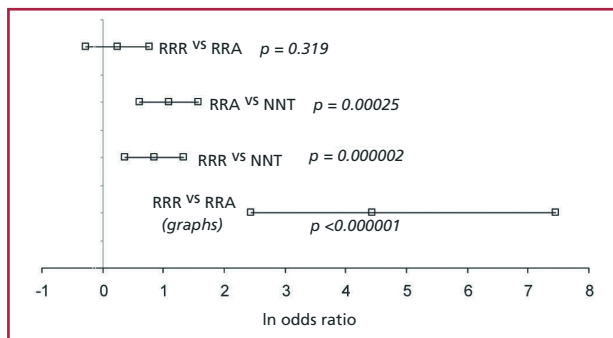


Fig. 2. Analysis of the results according to ln odds ratio and its 95% confidence interval for each type of comparison. NNT = Number Needed to Treat; ARR = Absolute Risk Reduction; RRR = Relative Risk Reduction.

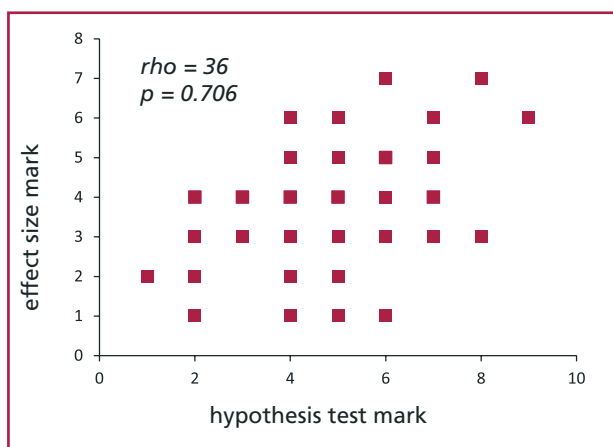


Fig. 3. Spearman correlation between the marks obtained in the hypothesis test exams and the marks in the relative and absolute risks tests.

#### BASIC CONCEPTS

Clinical trial results may be expressed as relative risks and/or absolute risks.

According to the method chosen, the treatment risks and benefits may be misinterpreted.

Medical students often confuse relative risks and absolute risks in their decision making.

Student training in this type of quantitative analysis is essential to correct these issues and improve the process of medical decision making.

#### RESUMEN

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

#### 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. El



objetivo fue conocer a través de una evaluación cognitiva, el grado de influencia que podría ejercer la forma de presentación de los riesgos y 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 quienes se le presentaron los resultados resumidos de distintos ensayos clínicos, de forma de mostrar los datos como reducción relativa del riesgo (RRR), reducción absoluta del riesgo (RA) 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 RA: 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 RA: 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.

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

### Conflicts of interest

Dr. Borracci is a professor at Universidad Austral.  
Dr. Arribalzaga is professor of Biostatistics at Universidad Austral.

### REFERENCES

1. 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>

2. 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>
3. 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>
4. Tversky A, Kahneman D. The framing of decisions and the rationality of choice. *Science* 1981;211:453-8. <http://doi.org/fj3z3r>
5. Kahneman D, Tversky A. Choices, values, and frames. *Am Psychologist* 1984;39:341-50. <http://doi.org/b4mb2g>
6. 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>
7. Farrow 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>
8. 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/qzw>
9. 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>
10. 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>
11. Woloshin S, Schwartz LM. Press releases: translating research into news. *JAMA* 2002;287:2856-8. <http://doi.org/d4xkbp>
12. Moynihan 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>
13. Thomson R, Edwards A, Grey J. Risk communication in the clinical consultation. *Clin Med* 2005;5:465-9. <http://doi.org/qzx>
14. 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/cg4qcq>
15. 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>