ORIGINAL REPORTS

The Incidence of Overconfidence and Underconfidence Effects in Medical Student Examinations

Raúl A. Borracci, MD, and Eduardo B. Arribalzaga, MD

Department of Surgery, School of Medicine, University of Buenos Aires, Buenos Aires, Argentina

BACKGROUND: Overconfidence is the tendency to overestimate the knowledge, capacity, or performance one really possesses. This cognitive bias could be potentially dangerous in medical decision-making, considering the impact it could have on patient health care. The aim of this study was to evaluate the incidence of overconfidence and underconfidence in medical student knowledge on general surgery by using traditional and new statistical approaches.

METHODS: During the application of a multiple-choice examination, 251 next-to-graduate medical students were invited to express the accuracy of their responses by choosing their own perceived confidence level for a set of questions. Analysis was done by comparing the difference between percentage of right answers (student's actual knowledge or accuracy) and self-estimated confidence level (student's perceived knowledge or confidence). Overconfidence was defined as a positive difference between confidence and accuracy, and underconfidence as a negative difference.

RESULTS: Nearly 12% of students showed significant overconfidence regarding their actual knowledge or accuracy levels. Better students showed a lower overconfidence effect than students with poorer performance. On the other hand, underconfidence was less likely than overconfidence (8.3% of students), and that effect was most frequently found in students who performed better in examinations.

CONCLUSIONS: The small proportion of our students exhibiting overconfidence or underconfidence behaviors moderates the need for educational interventions. Nevertheless, promoting prudence in individualized students manifesting overconfidence, and trust in those reporting significant underconfidence could increase the reliability of medical judgment during their future professional life. Overconfidence in individuals with lower scores in examinations may depend on a ceiling-like effect, since worst ranked students have a wider upper margin to manifest their confidence perceptions. The most confident students showed higher scores in examinations than the less confident ones. From this point of view, confidence could be considered an essential ingredient of success in examination performance. (J Surg Ed ▮:▮▮▮-▮▮▮. © 2018 Association of Program Directors in Surgery. Published by Elsevier Inc. All rights reserved.)

KEY WORDS: medical students, general surgery, education, cognitive sciences, trust

COMPETENCIES: Medical Knowledge

INTRODUCTION

Charles Darwin wrote that ignorance often generates greater confidence than does knowledge.[1] More recently, the Dunning-Kruger effect was described, stating that the incompetent are often ill-suited to recognize their incompetence.[2,3] Originally, Kruger and Dunning[2] suggested that people who tend to overestimate their intellectual or social abilities suffer a dual burden: not only they reach erroneous conclusions and make unfortunate choices, but their incompetence robs them of the metacognitive ability to realize it. The overconfidence effect is a well-known bias in which a person's subjective confidence in his own judgments is reliably greater than the objective accuracy of those judgments.[4] Somehow, overconfidence is one example of miscalibration of subjective probabilities. Pallier et al.[5] have defined overconfidence in 3 different ways: (1) as an overestimation of one's actual performance, (2) as an overplacement of one's performance relative to others, and (3) as an excessive certainty regarding the accuracy of one's beliefs or knowledge, also termed overprecision.

Statement: The manuscript has been read and approved by all the authors, the requirements for authorship have been met, and each author believes that the manuscript represents honest work. The authors confirm all personal identifiers have been removed or disguised so the person(s) described are not identifiable and cannot be identified through the details of the story.

Correspondence: Inquiries to Raúl A. Borracci, Department of Surgery, School of Medicine, University of Buenos Aires, Av. Córdoba 2351, (C1120AAR), Buenos Aires, Argentina. Fax: +(54) 02322-48-2204; e-mail: raborracci@gmail.com

  1931-7204/$30.00 https://doi.org/10.1016/j.jsurg.2018.01.015

Researchers in the field of cognitive psychology have shown that most people show an excessive level of confidence to answer a series of questions on general knowledge, after asking them to express the accuracy of their responses.[6,7] By analogy, overconfidence could be potentially dangerous in medical decision-making, considering the direct impact it could have on patient health care.[8] Berner et al.[9] argued that physicians in general underappreciate the likelihood that their diagnoses may be wrong.

While doctors need to demonstrate a certain level of confidence in their skills to interact with patients, they also need to know the wrong foundations of that confidence. Academic tests measured by a confidence-weighted scoring technique showed that overconfidence and underconfidence indices help to predict medical students' subsequent academic achievements.[10] Other investigations found that students tended toward underconfidence in their diagnostic judgments when classifying heart arrhythmias.[11] Furthermore, nearly 19% of medical residents showed to be overconfident when diagnosing complex clinical cases,[12] while faculty physicians were overconfident in 13% of diagnoses.[13] Cross-cultural variations in probability judgment accuracy and confidence were also recognized, even in medical diagnoses.[14] Some concerns exist about some counterintuitive effects of confidence (hard-easy and underconfidence-with-practice effects), and about the correspondence between subjective and objective probabilities, since the same data can appear to reveal both overconfidence and underconfidence, depending on the method of data analysis.[15,16] According to these observations, other authors proposed that somehow, confidence level could be considered a statistical artifact.[17,18]

Based on this theoretical framework we hypothesized that next-to-graduate medical students could show a higher or lower level of confidence than actually justified by their knowledge or performance. Therefore, the aim of this work was to evaluate the incidence of overconfidence and underconfidence in medical student knowledge by using traditional and new statistical approaches.

MATERIAL AND METHODS

From May 2015 to September 2016, a prospective study was conducted at the Buenos Aires University School of Medicine. While taking a multiple-choice examination of general surgery, 251 next-to-graduate medical students (Fifth-year students in Argentina) expressed the accuracy of their responses by choosing their own perceived confidence level for each question on a 5-point scale (20, 40, 60, 80, and 100). Each multiple-choice question had 5 options, and only 1 right answer. After selecting the presumed right answer, students pointed out the confidence level they estimated for the response. The procedure was repeated for a total of 60 questions, in order to obtain the percentage of right answers and the average percentage of confidence levels selected by each student for the whole examination. Statistical analysis was done by comparing the difference between percentage of right answers (student's actual knowledge or accuracy) and self-estimated confidence level (student's perceived knowledge or confidence). Overconfidence (that is to say, overprecision) was operatively defined as a positive difference between confidence and accuracy, and underconfidence (that is to say, underprecision) as a negative difference between those values. Statistically significant values of overconfidence and underconfidence were considered when a particular confidence-to-accuracy difference exceeded the 95% confidence limit.

Statistical Analysis

Continuous variables were expressed as mean and standard deviation or 95% CI, or median and 25-75 percentile ($P_{25\%\text{-}75\%}$). Kolmogorov-Smirnov goodness-of-fit test was used to analyze normal distributions. Independent metric variables were compared with Student *t*-test, and paired variables (concordance/accuracy) with paired Student *t*-test. Comparison of dichotomous variables was performed using the χ^2 test and the odds ratio (OR) with the associated 95% CI. Two-tailed Fisher exact test was employed when cell expected values were ≤5. A modified (nonbinary) mean probability score, or Brier score, was used to assess the statistical consistency between students' confidence and accuracy, and calibration was determined with Spiegelhalter's *Z*-statistical method.[19] The Brier score is useful to measure the predictive accuracy of a judgment. It compares the observed probability (y) (in the current study, the percentage of right answers) with the expected probability (p) (the mean confidence level) for each student (i), according to the following equation:

$\text{Brier} = 1/n \; \Sigma \; [(y_i - p_i)/100]^2$, where n is the number of students. Nonparametric Kendall τ_b was used as a measure of association to test the significance of the confidence/accuracy relationship. The output of the expected-to-observed (E-O) (confidence-to-accuracy) percentage of right answers for each student was calculated and plotted. The plot y-axes computed the expected minus the observed percentage of right answers. The expected percentage corresponds to the confidence level as estimated by each student, and the observed percentage indicates the accuracy or proportion of right answers obtained in the examination. To determine statistically significant overconfidence or underconfidence, 95% confidence limits of E-O difference were calculated for each student's performance with the confidence interval of the difference between 2 proportions. In addition, the statistical analyses were performed with SPSS Statistics for Windows, Version 17.0. Chicago: SPSS