

# Anesthesiologists' Overconfidence in Their Perceived Knowledge of Neuromuscular Monitoring and Its Relevance to All Aspects of Medical Practice: An International Survey

Mohamed Naguib, MD, MSc, FCARCSI,\* Sorin J. Brull, MD, FCARCSI (Hon),†  
Jennifer M. Hunter, MBE, MBChB, PhD, FRCA, FCARCSI (Hon),‡ Aaron F. Kopman, MD,§  
Béla Fülesdi, MD, PhD, DSci,|| Ken B. Johnson, MD,¶ and Hal R. Arkes, BA, PhD#

See Editorial, p 1057

**BACKGROUND:** In patients who receive a nondepolarizing neuromuscular blocking drug (NMBD) during anesthesia, undetected postoperative residual neuromuscular block is a common occurrence that carries a risk of potentially serious adverse events, particularly postoperative pulmonary complications. There is abundant evidence that residual block can be prevented when real-time (quantitative) neuromuscular monitoring with measurement of the train-of-four ratio is used to guide NMBD administration and reversal. Nevertheless, a significant percentage of anesthesiologists fail to use quantitative devices or even conventional peripheral nerve stimulators routinely. Our hypothesis was that a contributing factor to the nonutilization of neuromuscular monitoring was anesthesiologists' overconfidence in their knowledge and ability to manage the use of NMBDs without such guidance.

**METHODS:** We conducted an Internet-based multilingual survey among anesthesiologists worldwide. We asked respondents to answer 9 true/false questions related to the use of neuromuscular blocking drugs. Participants were also asked to rate their confidence in the accuracy of each of their answers on a scale of 50% (pure guess) to 100% (certain of answer).

**RESULTS:** Two thousand five hundred sixty persons accessed the website; of these, 1629 anesthesiologists from 80 countries completed the 9-question survey. The respondents correctly answered only 57% of the questions. In contrast, the mean confidence exhibited by the respondents was 84%, which was significantly greater than their accuracy. Of the 1629 respondents, 1496 (92%) were overconfident.

**CONCLUSIONS:** The anesthesiologists surveyed expressed overconfidence in their knowledge and ability to manage the use of NMBDs. This overconfidence may be partially responsible for the failure to adopt routine perioperative neuromuscular monitoring. When clinicians are highly confident in their knowledge about a procedure, they are less likely to modify their clinical practice or seek further guidance on its use. (Anesth Analg 2019;128:1118–26)

## KEY POINTS

- **Question:** What is the confidence level among anesthesiologists in their personal knowledge of how to manage the administration of neuromuscular blocking drugs and reversal agents?
- **Findings:** Anesthesiologists surveyed are overconfident in their knowledge of how to monitor neuromuscular blockade.
- **Meaning:** Anesthesiologists' overconfidence may contribute to their belief that they can intuitively manage neuromuscular blockade without neuromuscular monitoring.

*All you need in this life is ignorance and confidence, and then success is sure.*

—Samuel Langhorne Clemens  
(known as Mark Twain) (1835–1910)

A 2010 survey found that almost 20% of European and 10% of American anesthesiologists never use neuromuscular monitors of any kind.<sup>1</sup> Most respondents reported that neither conventional peripheral nerve

From the \*Department of General Anesthesia, Cleveland Clinic Lerner College of Medicine of Case Western Reserve University, Cleveland Clinic, Cleveland, Ohio; †Department of Anesthesiology and Perioperative Medicine, Mayo Clinic College of Medicine, Jacksonville, Florida; ‡Department of Musculoskeletal Biology, University of Liverpool, Liverpool, United Kingdom; §Weill Cornell Medical College, New York City, New York; ||Department of Anesthesiology and Intensive Care, University of Debrecen Faculty of Medicine, Debrecen, Hungary; ¶Department of Anesthesiology, University of Utah, Salt Lake City, Utah; and #Department of Psychology, The Ohio State University, Columbus, Ohio.

Accepted for publication July 3, 2018.

Copyright © 2018 International Anesthesia Research Society

DOI: 10.1213/ANE.0000000000003714

Funding: S.J.B. has received research funding from Merck & Co, Inc (Kenilworth, NJ) (funds to Mayo Clinic).

Conflicts of Interest: See Disclosures at the end of the article.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website ([www.anesthesia-analgesia.org](http://www.anesthesia-analgesia.org)).

M. Naguib and S. J. Brull contributed equally to this work and share first authorship.

A. F. Kopman is now retired.

Reprints will not be available from the authors.

Address correspondence to Mohamed Naguib, MD, MSc, FCARCSI, Department of General Anesthesia, Anesthesiology Institute, Cleveland Clinic, 9500 Euclid Ave, NE6-306, Cleveland, OH 44195. Address e-mail to [naguibm@ccf.org](mailto:naguibm@ccf.org).

stimulators (PNSs) nor quantitative train-of-four monitors should be part of minimum monitoring standards.<sup>1</sup> Thus, neuromuscular blockers and their antagonists are often administered without proper guidance.<sup>1-3</sup> The incidence of residual paralysis in the immediate postoperative period (defined as a train-of-four ratio, <0.90) remains around 40%.<sup>2,4-6</sup> It is estimated that as many as 112,000 patients annually in the United States are at risk of critical respiratory events associated with undetected residual neuromuscular blockade.<sup>7</sup> Residual neuromuscular weakness may result in tracheal reintubation in the postanesthesia care unit (PACU), postoperative pulmonary complications, and delays in hospital discharge.<sup>8-10</sup> The incidence of tracheal reintubation in the PACU directly attributed to inadequate recovery of neuromuscular function is not known. However, different studies from large medical centers report reintubation rates of 0.05%–0.19% in this context.<sup>11-14</sup>

Despite a plethora of publications that stress the need for routine neuromuscular monitoring,<sup>15-18</sup> many anesthesiologists feel sufficiently confident about their medical knowledge and clinical expertise to believe that they can safely manage neuromuscular blockade without quantitative monitoring or even the use of a PNS.<sup>1</sup> Evidence contradicts these beliefs.<sup>2,5,19,20</sup>

The aim of our study was to explore anesthesiologists' confidence in their knowledge of the core concepts in neuromuscular monitoring. Overconfidence in this clinical domain may contribute to the persistence of inadequate detection of residual neuromuscular blockade by well-trained anesthesiologists. In the presence of overconfidence, the perceived need to confirm clinical assessments with quantitative monitoring tools is probably considered superfluous.<sup>21,22</sup>

We hypothesized that many anesthesiologists are overconfident in their knowledge of the intraoperative management of neuromuscular monitoring. To explore this hypothesis, we conducted a survey of anesthesiologists that consisted of 9 questions relevant to the clinical management of neuromuscular blockade and assessed their confidence in their response to each question. Inappropriately high confidence in their (incorrect) answers would be consistent with our hypothesis.

## Confidence

The assessment of confidence has a long history in psychology and decision sciences. Such assessment is usually couched in terms of "calibration." To the extent that one's confidence is appropriate for one's level of accuracy, the person is said to be well calibrated. Assume that you have been presented with one hundred 2-option forced-choice questions such as "What is the capital of Arizona? A. Tucson. B. Phoenix." If you have assigned a confidence level of 70% to 10 of your 100 answers to such questions, you should correctly answer exactly 7 of those 10 questions to be perfectly calibrated; your accuracy and confidence level for that subset of questions would be identical, 70%. If you have assigned a confidence level of 100% to 27 of the 100 questions, every one of those 27 questions must be answered correctly for you to be perfectly calibrated for that subset of questions. Figure 1 contains 3 curves. The 45° line represents perfect calibration of confidence; each confidence level

is appropriate for each level of accuracy. Over- and underconfidence are also indicated. Note that the abscissa cannot go <50% for the confidence one has in one's answers to 2-option questions because if one were <50% confident that the correct answer was chosen, she/he would have selected the other answer. Also note that under- and overconfidence are both examples of poor calibration; both diverge from the perfect calibration, 45° line in Figure 1.

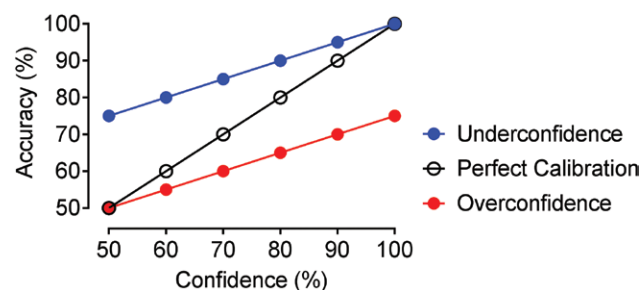
It is important to understand what the calibration of confidence does not represent: it is not a measure of accuracy. People who answer a small number of questions correctly can be very well calibrated if they assign appropriately low confidence levels to their answers. Our hypothesis is that anesthesiologists answering questions about neuromuscular monitoring will provide data similar to the curve labeled "overconfident" in Figure 1. We offer no hypothesis as to what their accuracy data will be.

## Calibration in the Medical Decision-Making Literature

Calibration is an important aspect of judgment performance, but the medical decision-making literature suggests that calibration is often deficient. For example, Dawson et al<sup>23</sup> asked 198 physicians to estimate values of pulmonary capillary wedge pressure, cardiac index, and systemic vascular resistance before right heart catheterization in 846 patients. The physicians also provided confidence in each of their estimates. The estimates' deviation from the actual measurements was unrelated to physician confidence in the estimates. The lack of relationship between accuracy and confidence signifies extremely poor calibration. Similarly, Stiegler et al<sup>24</sup> noted the presence of overconfidence among over half of resident physicians' management of a simulated anesthesia emergency.

## Why Appropriate Confidence Is Important

Arkes et al<sup>21</sup> asked baseball experts and nonexperts to examine the batting statistics for each of 3 players who might have won the National League Most Valuable Player Award for each of 19 years. The participants were given the following helpful decision rule: "If you choose the player whose team finished highest in the standings that year, you will choose



**Figure 1.** The black line with open circles represents perfect calibration of confidence; each confidence level is appropriate for each level of accuracy. The red and blue lines with filled circles represent over- and underconfidence, respectively. For 2-option questions (eg, true/false), the abscissa starts at 50%. Respondents would not indicate that their confidence was <50%, because to do so would indicate that they preferred the other answer. Under- and overconfidence are both examples of poor calibration; both diverge from the perfect calibration, 45° line.

correctly 75% of the time.” The experts were seriously overconfident in their selection of the most valuable player for each of the 19 years. The nonexperts actually made significantly more correct choices than did the experts, due to the fact that the nonexperts used the very helpful decision rule more. We propose that an inappropriately high degree of confidence might also be responsible for anesthesiologists’ reluctance to utilize quantitative assessment of perioperative neuromuscular monitoring; this hypothesis is consistent with a conjecture by Stiegler and Tung<sup>25</sup> that overconfidence would diminish a physician’s likelihood to seek assistance.

### Slope As an Additional Dependent Variable

Not only does the assessment of accuracy and confidence allow us to measure calibration, it also allows us to examine an index called “slope,”<sup>26</sup> which is defined as the difference between the confidence assigned to correct answers and the confidence assigned to incorrect answers. Slope is not a measure of under- or overconfidence. It is a measure of discrimination; can the respondent discriminate between correct and incorrect answers by assigning different confidence levels to these 2 categories of answers? People can be overconfident, underconfident, or perfectly calibrated and yet manifest the exact same slope. Whereas calibration pertains to the overall relationship between confidence and accuracy, slope is an important index of “meta-knowledge”—knowledge about what one knows and what one does not know. True experts can have good slope even if they are not well calibrated. Slope and calibration are not controversial concepts: no one would accept weak calibration and poor slope. We sought to address these issues by administering a 9-item test on neuromuscular monitoring to anesthesiologists and assessing the calibration and slope of the confidence ratings of their answers.

### METHODS

After obtaining institutional approvals from Cleveland Clinic (Cleveland, OH) and Mayo Clinic (Jacksonville, FL), we conducted an Internet-based survey among anesthesiologists worldwide. Respondents gave their informed consent online before completing the survey. The International Anesthesia Research Society (United States), the Royal College of Anaesthetists (United Kingdom), The European Society of Anaesthesiology, the São Paulo State Society of Anesthesiologists (Brazil), the Hungarian Society of Anesthesiology and Intensive Therapy, the Swiss Society of Anesthesiology, and the Society of Anaesthetists of Hong Kong e-mailed all of their active members on the authors’ behalf, inviting them to complete a 9-question survey anonymously (Supplemental Digital Content, Document, <http://links.lww.com/AA/C535>).

The survey questions were developed and critiqued by all investigators and were validated in a pilot study. The surveys were piloted twice. Our aim was to select 8–9 questions (with indisputable answers) and to omit any question that was not answered correctly (or any question that was answered correctly) by a large majority of respondents. Three questions were deleted after the first pilot, and 1 was deleted after the second pilot. We aimed to exclude nondiscriminatory questions in the pilot study. A question was deemed to be nondiscriminatory

if it did not discriminate well between anesthesiologists who answered most of the questions correctly and those who did not. For instance, if there were a question for which almost no pilot subject knew the answer, the question would have been removed from the questionnaire. Also, if 99% of respondents had answered correctly, the question would have been removed from the questionnaire. Survey face validity was deemed adequate. Each survey question was developed based on previously published findings related to neuromuscular blockade monitoring. (See References Supporting Survey Questions in the Supplemental Digital Content, Document, <http://links.lww.com/AA/C535>.)

We provided additional separate unique links to the survey in different languages (German, French, Spanish, Portuguese, and Hungarian). To ensure accurate translation of the survey from English to other languages, a 2-step process was used. First, anesthesiologists who were native speakers of these languages translated the survey from English into their native language. Second, different anesthesiologists who were native speakers of these languages back-translated the surveys into English. The original and the back-translated versions were compared independently by 2 investigators (M.N. and S.J.B.) to assess translation accuracy.

The survey was stored on a dedicated website (<https://www.surveymonkey.com/r/neuromuscularsurvey>) accessed via computer. The survey was designed to be completed in <10 minutes, and all questions were formatted into a Hypertext Markup Language interface. Responses were stored electronically in separate customized databases. The survey was available online for 90 days (August 1, 2017 to October 31, 2017). An electronic link to the survey was provided within the body of the e-mail invitation and was also published on the Royal College of Anaesthetists (United Kingdom) website. Reminder e-mails were sent 30 days after the initial invitations.

All questions were in a true/false format. After each question, we asked the respondent to indicate his/her level of confidence in the answer using a scale from 50% to 100% in 5% increments using a drop-down menu. Participants were orientated to the confidence scale using the following definition: 50% indicated complete uncertainty; and 100% indicated complete certainty.<sup>27,28</sup>

This combination of 2-option questions and a 50%–100% confidence rating scale is extremely common in the judgment/decision-making literature.<sup>29</sup> As the scales of the 2 factors—accuracy and confidence—align perfectly, no respondent would ever give a confidence rating <50% because if a person were <50% confident in 1 of the 2 possible answers, he or she would choose the other answer, whose confidence level would therefore be >50%. If a person were totally ignorant about the 2 possible options for any question, then respondent would answer correctly with a probability of 50%. Thus, a perfectly calibrated but ignorant person would assign a confidence level of 50% to a question that had an accuracy probability of 50%. Confidence and accuracy would align perfectly. Analogously, a perfectly calibrated person who was fully informed about a domain would assign confidence levels of 100% to answers that were 100% likely to be accurate. Thus, the use of 2-option questions with a 50%–100% confidence scale describes a straightforward interaction between accuracy and confidence, which is the definition of calibration.



Respondents were also asked the number of years in practice (0–5, 6–10, 11–15, 16–20, >20), their degree, and the country in which they practiced.

The raw data (excluding the country of each respondent) are deposited in the Open Science Framework (<https://osf.io/rn2mv>).

## STATISTICAL CONSIDERATIONS

Calibration is a measure of the correspondence between accuracy and confidence. To calculate calibration, we subtracted the average proportion of the questions answered correctly from the average confidence assigned to the answers to the questions. For example, if a person answered 5 of the 9 questions correctly, the proportion of correct answers is 0.556. If the average confidence one assigns to the answers to these 9 questions is 95%, then overconfidence is  $0.950 - 0.556 = 0.394$  or 39.4%. We estimated that a total sample size of 272 respondents will be needed for a  $t$  test to have a 95% power to detect an effect size (Cohen's  $d$ ) of 0.2 to test whether the mean overall calibration is  $>0$  (1-tailed test).<sup>30</sup> Although both over- and underconfidence signify poor calibration, we use overconfidence in this example because underconfidence is extraordinarily rare and therefore was unanticipated in our data.

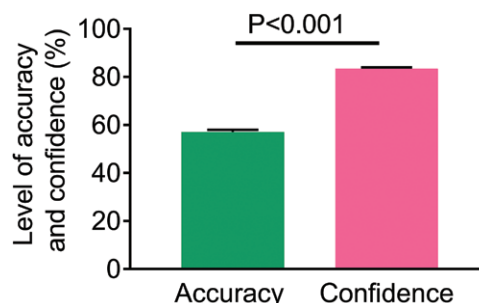
We performed two 1-way analyses of variance to ascertain whether the number of years of professional experience was related either to the magnitude of overconfidence or to the percentage of questions answered correctly.

Slope was indexed by subtracting the confidence assigned to the incorrect answers from the confidence assigned to the correct answers. We calculated overall slope, and because of the heterogeneity of the questions, we also calculated the slope separately for each question using a Bonferroni correction thereby making  $\alpha = .0056$  (.05/9). A  $P$  value  $<.05$  was considered statistically significant for all other analyses.

## RESULTS

A total of 2560 persons accessed the website and provided demographic information; of these, 1629 anesthesiologists from 80 different countries completed the 9-question survey. Supplemental Digital Content, Table 1, <http://links.lww.com/AA/C535>, contains the number of respondents from each country who completed the survey. It is impossible to determine the number of society members who received or viewed the survey. This number would be needed to calculate the response rate. Lacking that number, we only can assert that the number of respondents who completed the survey was large (1629) and geographically diverse. Supplemental Digital Content, Table 2, <http://links.lww.com/AA/C535>, contains information about the number of respondents at each level of experience.

The 1629 respondents correctly answered an average of 57.1% of the 9 questions. The mean confidence assigned by the respondents was 83.5% (95% confidence interval [CI], 83%–84%), which was greater than their accuracy of 57.1% (95% CI, 56.2%–58.0%) ( $t_{(1628)} = 55.48$ ,  $P < .001$ ). The magnitude of overconfidence was thus 26.4% (95% CI, 25.48%–27.34%) (Figure 2). Of the 1629 respondents, 1496 (91.8%) were overconfident, 119 (7.3%) were underconfident, and 14 (0.9%) were perfectly calibrated. Supplemental Digital



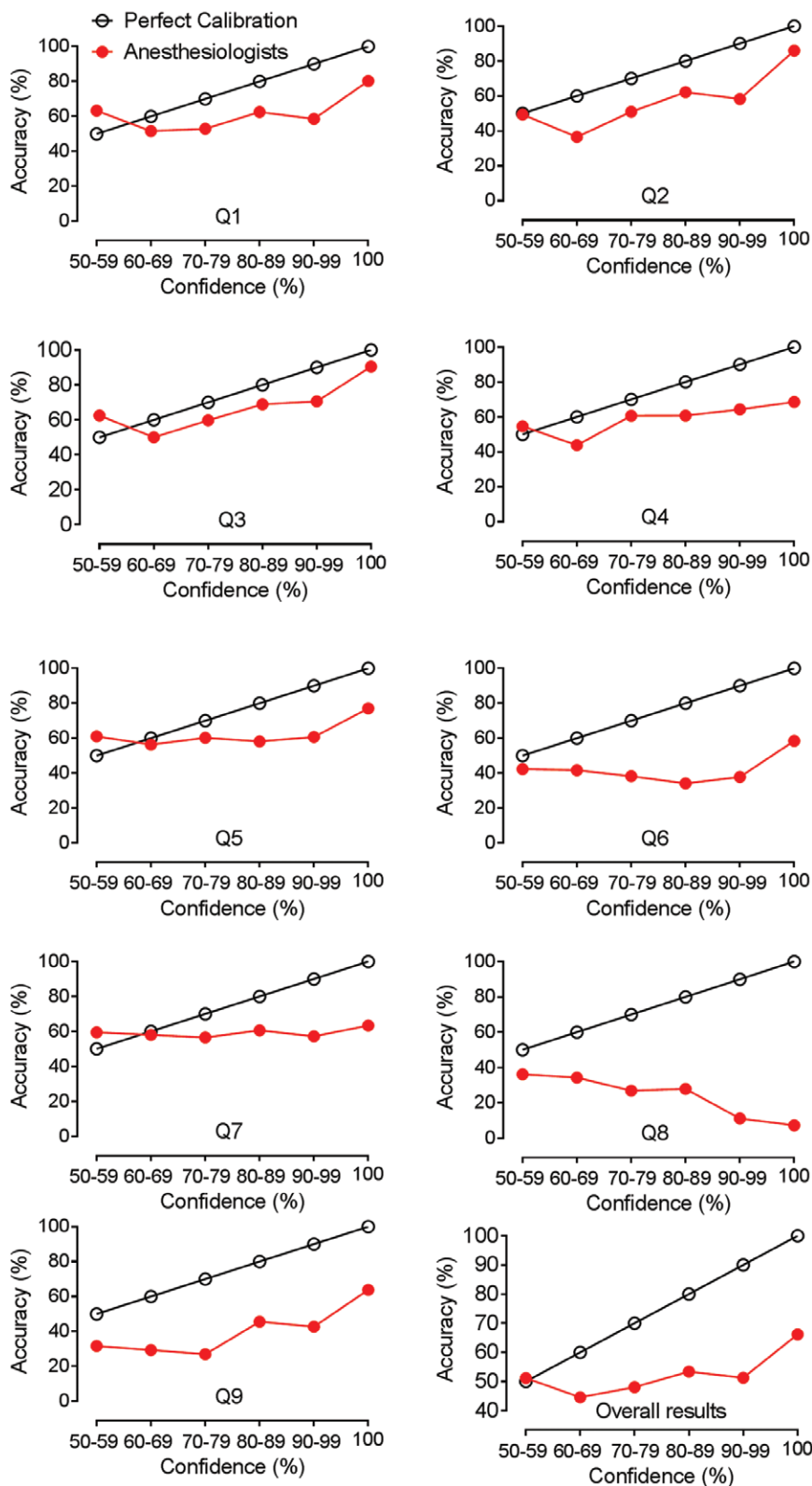
**Figure 2.** The level of accuracy and the level of confidence among the surveyed anesthesiologists. The mean confidence assigned by the respondents was 83.5% (95% CI, 83%–84%), which was greater than their accuracy of 57.1% (95% CI, 56.2%–58.0%) ( $t_{(1628)} = 55.48$ ,  $P < .001$ ). The magnitude of overconfidence was thus 26.4% (95% CI, 25.48%–27.34%). Data represent mean and 95% CI. CI indicates confidence interval.

Content, Table 3, <http://links.lww.com/AA/C535>, presents the proportion of correct responses for each question at each level of confidence and overall accuracy at each level of confidence. Figure 3 presents the calibration curves for each question and an overall calibration curve for all questions combined. For the purpose of constructing the graphs, we noted the accuracy of each individual answer which was assigned a confidence level within each of the 5 deciles (50%–59%, 60%–69%, 70%–79%, 80%–89%, 90%–99%) and 100% and plotted the mean accuracy of those answers in the middle of the appropriate decile and for 100% on the x-axis.<sup>a</sup>

Survey respondents manifested overconfidence on every question. When survey respondents were certain that they answered the question correctly (100% confidence), their accuracy was only 66.2%. When the respondents expressed a confidence of 50%, they were correct 46.4% of the time.

We also calculated the slope for the respondents' confidence ratings. Slope is indexed by the difference between the confidence assigned to the correct and incorrect answers. The average confidence for the correct answers was 85.6, and the average confidence for the incorrect answers was 79.7, yielding a slope of 5.8 (note the rounding error; 95% CI, 5.18–6.40) ( $t_{(1611)} = 18.53$ ,  $P < .001$ ). The lower degrees of freedom are due to the fact that for respondents who answered all of the questions correctly ( $n = 14$ ) or all of the questions incorrectly ( $n = 3$ ), the slope was not calculable. In terms of each question, for all but questions 7 and 8, the slope was positive and significant (each of 7  $t_{(1627)} > 4.3$ , each  $P < .001$ ). The slope for question 8 was significant and negative (–13.8),  $t_{(1627)} = 12.79$ ,  $P < .001$ , which means that when anesthesiologists answered incorrectly they were more confident of the correctness of their answer than when they responded correctly. The slope for question 7 was not significantly different from zero. Experience was not significantly related to either calibration or slope (both  $F_s < 1.0$ ).

<sup>a</sup>Because the graphs have 6 decile subdivisions for confidence, a large number of respondents did not use all of the deciles in his or her answers. The accuracy level for such deciles is therefore undefined for many respondents. Many other respondents used a decile (eg, 70%) only once. The accuracy level for such deciles must be either 0% or 100%. These complications render error bars inappropriate. This is the reason that calibration graphs contain no error bars.



**Figure 3.** Calibration curves for each of the 9 questions and for the overall results summed over 9 questions. The black lines with open circles represent perfect calibration of confidence; each confidence level is appropriate for each level of accuracy. The red lines with filled circles represent overconfidence.

## DISCUSSION

The results confirmed our primary study hypothesis: the findings of this survey of substantial statistical power suggested that the anesthesiologists we sampled are overconfident regarding their knowledge of intra-operative neuromuscular blockade management and

monitoring. The most important results of this investigation were that the respondents were overconfident in their knowledge (26.4%), while their slope was only 5.8%. The mean overconfidence of 26.4% is larger than that generally found in the current judgment/decision-making literature.<sup>28,29</sup>

We attempted to characterize the magnitude of overconfidence and slope by considering previous research on these 2 factors. There are no prior studies using anesthesia professionals as respondents to 2-option, half-scale (50%–100%) true-false questions. Given this lack of prior comparable data, what level of calibration might we have expected from our respondents?

Juslin et al<sup>29</sup> summarized the data from numerous studies involving 29 samples of 2-option half-scale questions. They reported a mean proportion correct of 0.65. This is slightly above the accuracy level observed in our respondents. The mean magnitude of overconfidence in these studies was 10% (95% CI,  $\pm 2\%$ ). By comparison, our observed level of overconfidence in surveyed anesthesiologists (26.4%) is much greater than we might have expected.<sup>b</sup>

There are numerous studies in which physicians and medical students of various specialties plus other health professionals were asked to assign confidence ratings to their performance on factual questions, various clinical skills, and diagnoses. (These studies do not utilize 2-option half-scale response scales.) Such studies regularly document poor calibration between confidence and the level of performance with overconfidence being the usual finding.<sup>26,31–37</sup> The results of these studies support the conclusion that our result of significant overconfidence is representative and that our findings are not anomalous. In addition, we used vastly more participants (from many different countries) than these earlier studies, thus increasing our confidence in the results.

A second consideration that might have guided our expectations concerning the calibration of our respondents is known as the “hard-easy effect.”<sup>29</sup> Figure 1 presents an idealized depiction of this effect. For hard questions with a low probability of being answered correctly, they are often found to be associated with overconfidence. As the questions become easier, the curve rises, even to the level that on extremely easy questions, underconfidence is occasionally observed.<sup>28</sup> The anesthesiologists’ relatively modest accuracy on the true/false test pertaining to perioperative neuromuscular monitoring was 1 factor contributing to their acute overconfidence. The “hard-easy” effect is the reason why question 8 manifested the most overconfidence; it was the question with the lowest proportion of correct answers.

A third consideration pertains to the feedback environment in which confidence judgments are rendered. Horserace handicappers,<sup>38</sup> expert bridge players,<sup>39</sup> and weather forecasters<sup>40</sup> exhibit outstanding calibration. These 3 groups obtain prompt, incontrovertible feedback after they make an estimate. A weather forecaster who estimates 100% chance of rain might need to seek alternative employment if such estimates occur frequently for days without any rain. Do anesthesiologists receive prompt, incontrovertible feedback about adverse outcomes every time they make a subjective estimate pertaining to neuromuscular function? Probably not, and so their calibration will be deficient

compared to the excellent calibration of those professionals who do receive such feedback.

A fourth factor to consider is the role of expertise. Expertise should certainly help the test taker to answer more questions correctly, but would it confer some advantage in calibration? Lichtenstein and Fischhoff<sup>41</sup> attempted to answer this question by asking graduate students in psychology to answer questions both on that topic and general-knowledge questions pertaining to a wide variety of other domains. The 2 sets of questions were matched in proportion answered correctly. The investigators found no difference in the calibration of the 2 sets of questions even though 1 was in the participants’ field of expertise. Thus, expertise in a domain conferred no calibration advantage, so by analogy we might not have expected the anesthesiologists’ superior domain knowledge to bolster their calibration.

Might medical expertise foster better—that is, larger—slope on questions of a medical context? Prior research has examined the ability of medical personnel to assign higher confidence levels to correct answers than incorrect ones.<sup>26,42</sup> Unfortunately, these studies involve predicting the occurrence of an external event. Such studies therefore involve a task that is not analogous to the prediction of the correctness of one’s own answers, which was the task in our study. Nevertheless, such studies might guide our expectations to some extent. Physicians predicting the 6-month survival of seriously ill patients had a mean slope of 26%.<sup>26</sup> The patients themselves, possessing more subjective knowledge but much less medical knowledge than their physicians, had a mean slope of only 13%. In a separate study of physicians’ predictions of the 6-month survival of lung cancer patients, the most accurate physician had a slope of 13%; the least accurate physician had a slope of 2%.<sup>42</sup> The slope we obtained was 5.8%, which does not compare favorably with prior data. Slope reflects the judge’s ability to discriminate correct answers from incorrect ones; not surprisingly, such discrimination ability fosters better calibration.<sup>42</sup>

### Why Would Anesthesiologists Not Utilize Perioperative Neuromuscular Monitoring?

We suggest that 1 reason why anesthesiologists might eschew the use of perioperative neuromuscular monitoring and especially the use of quantitative/objective monitoring of neuromuscular function is that those who believe that they have high levels of expertise think they do not need such assistance.<sup>21,22</sup> Closely related to this reason is the “better-than-average effect.”<sup>43</sup> For example, Svenson<sup>44</sup> reported that 46% of Americans placed themselves in the top 20% in driving skills, and 82% placed themselves in the top 30% in automotive safety. Thus, many anesthesiologists might agree that the average anesthesiologist should use quantitative monitoring, but as above-average clinicians, they may think, “I certainly don’t need to use it.” Those who are grossly overconfident would seem to be particularly vulnerable to the “better-than-average effect.”

Psychologists<sup>45</sup> and operations researchers<sup>46</sup> have long been frustrated by the fact that practitioners are highly reluctant to use quantitative actuarial formulae and decision aids that are more accurate than subjective intuitive

<sup>b</sup>Different levels of overconfidence occur depending on the type of task and dependent variable.<sup>29</sup> Those summarized by Juslin et al<sup>29</sup> on their page 390 are the studies most comparable to ours.



estimates. This has been documented previously in medical contexts.<sup>47,48</sup>

The reasons for the low adoption rate of quantitative neuromuscular monitoring are no doubt multifactorial. Overconfidence is likely to be only 1 cause of this reality<sup>22</sup>; other possible causes include overperception by practitioners of reasons why the quantitative technique might not be appropriate in a particular instance.<sup>49</sup> In addition, reticence to use monitors to guide management of neuromuscular block may be compounded by the relative paucity of easy-to-use, reliable objective monitors. One of the most frequently used such monitors (TOF-Watch; Organon, Cork, Ireland) is no longer manufactured or available commercially. However, new and improved free-standing objective monitors have recently been introduced into clinical use in the United States (StimPod NMS450; Xavant Technology Ltd, Pretoria, South Africa, and the TOFscan; IDMed, Marseilles, France), while others (TOFcuff; RGB Medical, Madrid, Spain, and TetraGraph; Sensime, Uppsala, Sweden) have been Conformité Européenne (CE) Mark certified and are available for clinical use outside the United States. Despite this availability, however, their use in the routine clinical setting remains limited. Clinicians who have never used quantitative monitors (erroneously) believe that clinically significant postoperative residual neuromuscular block is a rare event<sup>1</sup> and that such adverse outcomes do not occur in their own practice. Psychologists have long been investigating the reluctance of people to take precautions to protect against low-probability, high-impact negative events.<sup>50,51</sup> For example, until the advent of mandatory seat belts in American automobiles, many people refused to wear such belts due to the low probability of a serious accident in any individual trip. Only when the lifetime risk of not wearing a seat belt was made prominent were people more likely to “buckle up.”<sup>50</sup> Analogously, anesthesiologists might be reluctant to use objective monitoring due to the low perceived probability of residual block in any 1 patient. In fact, the majority of clinicians feel that the incidence of residual block is <1%. However, 15% of clinicians admit that they have seen a patient who had inadequately recovered from neuromuscular block in the PACU.<sup>1</sup>

Other factors contributing to the nonuse of neuromuscular monitoring might include the cost of providing monitors in every operating room, their maintenance, and practitioner resistance to change.

Our observations regarding overconfidence are perhaps even more applicable to the use of conventional PNS devices. There may be mitigating circumstances (eg, availability) for not using an objective monitor. But there are no reasonable excuses for not routinely using a PNS device when managing intraoperative neuromuscular block.

## LIMITATIONS

As we are unable to provide a precise response rate, the question may arise as to whether our respondents comprise a representative sample of anesthesiologists. Perhaps only those who were confident in the accuracy of their answers or interested in neuromuscular monitoring would attempt the survey. Would this “self-selection” foster the large magnitude of overconfidence we found? We think not, for a couple of reasons. A calibration analysis does not in any way

penalize respondents who are highly confident if their confidence is warranted by a high level of accuracy. What we do know from our analysis is that the confidence expressed by this large group of anesthesiologists was not warranted. We also know that their amount of clinical experience had no influence on calibration, which suggests that if either novice or highly experienced anesthesiologists selectively refrained from participating in our survey the results would likely not have been different.

A similar argument pertains to slope. An analysis of slope will not necessarily be unfavorable if that person had a high overall confidence level. One can still manifest a large difference between confidence levels assigned to correct versus incorrect answers despite being overconfident. We suggest that even if our survey had been differentially attractive to confident respondents, this would not explain their very poor calibration and slope.

This survey could also be criticized by the low response rate. Perhaps the biggest surveys of anesthesiologists are the annual surveys on salary and related issues (www.LocumTenens.com; www.medscape.com). Although these 2 surveys are not research experiments and do not ask the respondents to take a quiz in which their knowledge and confidence are being assessed, the 2017 response rate of these 2 surveys among American anesthesiologists was only 2% and 6%, respectively. We suspect that anesthesiologists are reticent about volunteering to have their professional knowledge assessed; we were gratified that 1629 anesthesiologists graciously did so. In fact, our study had a sample size far in excess of the number of participants ( $n = 24\text{--}144$ ) in other studies that assessed confidence.<sup>31–33,35–37</sup> Furthermore, Dawson et al<sup>23</sup> and Arkes et al<sup>26</sup> found very poor calibration among physicians in a range of specialties. Therefore, we think that the overconfidence we documented is not an anomalous finding caused by a nonrepresentative sampling procedure.

However, we do acknowledge that the number of our respondents from each country represents only a very small fraction of anesthesiologists practicing in that country. Although our sample size greatly exceeded that employed in other studies of calibration of medical knowledge, we have not tested the vast majority of anesthesiologists in any 1 country, and the survey results may not be an accurate representation of practicing anesthesiologists worldwide.

Another limitation of the current survey may be considered to be that many of the survey questions did not ask specifically how to monitor neuromuscular function or how to dose the reversal agents based on the results of monitoring. While this may seem a relevant limitation, it would be insufficient to simply ask clinicians about monitoring. For without knowledge of the neuromuscular blocking drugs and antagonists whose effects they are monitoring, would one expect the clinician to monitor appropriately, or even monitor at all? We think not. We therefore suggest that our questions were all appropriate.

Finally, another criticism might be that our survey did not specifically ask the respondents whether they were using neuromuscular (objective) monitoring in their practice, and perhaps those who do use such monitoring would have provided different data than those who do not. Thus,

the more users of neuromuscular monitoring there were among our survey participants, the more surprising would be the substantial overconfidence we did find among our respondents. However, our survey design does not allow us to ascertain if users and nonusers would have provided different data.

## IMPLICATIONS

The respondents to our survey answered an average of only 57% of the questions correctly. This was only slightly better than that might be achieved by pure guesswork. Even when respondents were absolutely certain of the correctness of their answer (100% confidence), they were correct only 66% of the time. Only 14 respondents (0.9%) were perfectly calibrated. Thus, the clinicians surveyed were considerably less knowledgeable regarding perioperative neuromuscular monitoring than they believed. Almost 2 decades ago, 2 social scientists described a similar phenomenon, which has subsequently become known as the Kruger–Dunning Effect.<sup>52</sup> This is a cognitive bias wherein unskilled individuals suffer from illusory superiority, mistakenly rating their ability much higher than warranted; this was attributed to a metacognitive inability of the unskilled to recognize deficits in their competence or knowledge. Our results are more remarkable than those of Kruger and Dunning<sup>52</sup> in that our subjects were, in contrast, very highly skilled professionals answering questions in their domain of expertise, not undergraduates rating such domains as the funniness of jokes and the grammatical quality of sentences. We suggest that the anesthesiologists' overconfidence and the inability to discriminate correct from incorrect answers are at least partially responsible for the widespread failure to adopt quantitative perioperative neuromuscular monitoring. In addition, the findings could have more general clinical relevance. When clinicians are highly confident that they are knowledgeable about a procedure, they are less likely to modify their clinical practice or seek further guidance or knowledge. It is hoped that the results of this survey and the discussion generated by the discrepancy between clinicians' knowledge and confidence will serve as an impetus for a new evaluation of the role of neuromuscular and specifically quantitative monitoring in routine anesthetic practice. ■■

## ACKNOWLEDGMENTS

The authors thank Francois Donati, MD, Anke Bellinger, MD, Klaus D. Torp, MD, Mauricio Perilla, MD, Jose L. Diaz-Gomez, MD, Daniela Ionescu, MD, PhD, and Ricardo Carlos Vieira, MD, for their help with language translations.

## DISCLOSURES

**Name:** Mohamed Naguib, MD, MSc, FCARCSI.

**Contribution:** This author helped design the study, analyze the data, and write the manuscript.

**Conflicts of Interest:** M. Naguib has served as a consultant for GE Healthcare in 2018.

**Name:** Sorin J. Brull, MD, FCARCSI (Hon).

**Contribution:** This author helped design the study and write the manuscript.

**Conflicts of Interest:** S. J. Brull is a principal and shareholder in Senzime AB (publ) (Uppsala, Sweden); and a member of the Scientific Advisory Boards for ClearLine MD (Woburn, MA),

The Doctors Company (Napa, CA), and NMD Pharma (Aarhus, Denmark).

**Name:** Jennifer M. Hunter, MBE, MBChB, PhD, FRCA, FCARCSI (Hon).

**Contribution:** This author helped design the study and write the manuscript.

**Conflicts of Interest:** None.

**Name:** Aaron F. Kopman, MD.

**Contribution:** This author helped design the study and write the manuscript.

**Conflicts of Interest:** None.

**Name:** Béla Fülesdi, MD, PhD, DSci.

**Contribution:** This author helped design the study and write the manuscript.

**Conflicts of Interest:** None.

**Name:** Ken B. Johnson, MD.

**Contribution:** This author helped design the study and write the manuscript.

**Conflicts of Interest:** None.

**Name:** Hal R. Arkes, BA, PhD.

**Contribution:** This author helped design the study, analyze the data, and write the manuscript.

**Conflicts of Interest:** None.

**This manuscript was handled by:** Tong J. Gan, MD.

## REFERENCES

1. Naguib M, Kopman AF, Lien CA, Hunter JM, Lopez A, Brull SJ. A survey of current management of neuromuscular block in the United States and Europe. *Anesth Analg*. 2010;111:110–119.
2. Naguib M, Kopman AF, Ensor JE. Neuromuscular monitoring and postoperative residual curarisation: a meta-analysis. *Br J Anaesth*. 2007;98:302–316.
3. Brull SJ, Murphy GS. Residual neuromuscular block: lessons unlearned. Part II: methods to reduce the risk of residual weakness. *Anesth Analg*. 2010;111:129–140.
4. Kopman AF, Yee PS, Neuman GG. Relationship of the train-of-four fade ratio to clinical signs and symptoms of residual paralysis in awake volunteers. *Anesthesiology*. 1997;86:765–771.
5. Todd MM, Hindman BJ, King BJ. The implementation of quantitative electromyographic neuromuscular monitoring in an academic anesthesia department. *Anesth Analg*. 2014;119:323–331.
6. Fortier LP, McKeen D, Turner K, et al. The RECITE study: a Canadian prospective, multicenter study of the incidence and severity of residual neuromuscular blockade. *Anesth Analg*. 2015;121:366–372.
7. Brull SJ, Naguib M, Miller RD. Residual neuromuscular block: rediscovering the obvious. *Anesth Analg*. 2008;107:11–14.
8. Murphy GS, Szokol JW, Marymont JH, Greenberg SB, Avram MJ, Vender JS. Residual neuromuscular blockade and critical respiratory events in the postanesthesia care unit. *Anesth Analg*. 2008;107:130–137.
9. Bulka CM, Terekhov MA, Martin BJ, Dmochowski RR, Hayes RM, Ehrenfeld JM. Nondepolarizing neuromuscular blocking agents, reversal, and risk of postoperative pneumonia. *Anesthesiology*. 2016;125:647–655.
10. Bronsart MR, Henderson WG, Monk TG, et al. Intermediate-acting nondepolarizing neuromuscular blocking agents and risk of postoperative 30-day morbidity and mortality, and long-term survival. *Anesth Analg*. 2017;124:1476–1483.
11. Mathew JP, Rosenbaum SH, O'Connor T, Barash PG. Emergency tracheal intubation in the postanesthesia care unit: physician error or patient disease? *Anesth Analg*. 1990;71:691–697.
12. Rose DK, Cohen MM, Wigglesworth DE, DeBoer DP. Critical respiratory events in the postanesthesia care unit. Patient, surgical, and anesthetic factors. *Anesthesiology*. 1994;81:410–418.
13. Lee PJ, MacLennan A, Naughton NN, O'Reilly M. An analysis of reintubations from a quality assurance database of 152,000 cases. *J Clin Anesth*. 2003;15:575–581.
14. Epstein RH, Dexter F, Lopez MG, Ehrenfeld JM. Anesthesiologist staffing considerations consequent to the temporal distribution of hypoxic episodes in the postanesthesia care unit. *Anesth Analg*. 2014;119:1322–1333.
15. Kopman AF. Managing neuromuscular block: where are the guidelines? *Anesth Analg*. 2010;111:9–10.



16. Miller RD, Ward TA. Monitoring and pharmacologic reversal of a nondepolarizing neuromuscular blockade should be routine. *Anesth Analg*. 2010;111:3–5.
17. Donati F. Neuromuscular monitoring: what evidence do we need to be convinced? *Anesth Analg*. 2010;111:6–8.
18. Brull SJ, Kopman AF. Current status of neuromuscular reversal and monitoring: challenges and opportunities. *Anesthesiology*. 2017;126:173–190.
19. Baillard C, Gehan G, Reboul-Marty J, Larmignat P, Samama CM, Cupa M. Residual curarization in the recovery room after vecuronium. *Br J Anaesth*. 2000;84:394–395.
20. Kotake Y, Ochiai R, Suzuki T, et al. Reversal with sugammadex in the absence of monitoring did not preclude residual neuromuscular block. *Anesth Analg*. 2013;117:345–351.
21. Arkes HR, Dawes RM, Christensen C. Factors influencing the use of a decision rule in a probabilistic task. *Organ Behav Hum Decis Process*. 1986;37:93–110.
22. Sieck WR, Arkes HR. The recalcitrance of overconfidence and its contribution to decision aid neglect. *J Behav Decis Mak*. 2005;18:29–53.
23. Dawson NV, Connors AF Jr, Speroff T, Kemka A, Shaw P, Arkes HR. Hemodynamic assessment in managing the critically ill: is physician confidence warranted? *Med Decis Making*. 1993;13:258–266.
24. Stiegler MP, Neelankavil JP, Canales C, Dhillon A. Cognitive errors detected in anaesthesiology: a literature review and pilot study. *Br J Anaesth*. 2012;108:229–235.
25. Stiegler MP, Tung A. Cognitive processes in anesthesiology decision making. *Anesthesiology*. 2014;120:204–217.
26. Arkes HR, Dawson NV, Speroff T, et al. The covariance decomposition of the probability score and its use in evaluating prognostic estimates. SUPPORT Investigators. *Med Decis Making*. 1995;15:120–131.
27. Koriati A, Lichtenstein S, Fischhoff B. Reasons for confidence. *J Exp Psychol Hum Learn*. 1980;6:107–118.
28. Lichtenstein S, Fischhoff B, Phillips LD. Calibration of probabilities: the state of the art to 1980. In: Kahneman D, Slovic P, Tversky A, eds. *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge, United Kingdom: Cambridge University Press, 1982:306–334.
29. Juslin P, Winman A, Olsson H. Naive empiricism and dogmatism in confidence research: a critical examination of the hard-easy effect. *Psychol Rev*. 2000;107:384–396.
30. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
31. Barnsley L, Lyon PM, Ralston SJ, et al. Clinical skills in junior medical officers: a comparison of self-reported confidence and observed competence. *Med Educ*. 2004;38:358–367.
32. Hodges B, Regehr G, Martin D. Difficulties in recognizing one's own incompetence: novice physicians who are unskilled and unaware of it. *Acad Med*. 2001;76:S87–S89.
33. Davis DA, Mazmanian PE, Fordis M, Van Harrison R, Thorpe KE, Perrier L. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *JAMA*. 2006;296:1094–1102.
34. Gordon MJ. A review of the validity and accuracy of self-assessments in health professions training. *Acad Med*. 1991;66:762–769.
35. Meyer AN, Payne VL, Meeks DW, Rao R, Singh H. Physicians' diagnostic accuracy, confidence, and resource requests: a vignette study. *JAMA Intern Med*. 2013;173:1952–1958.
36. Morgan PJ, Cleave-Hogg D. Comparison between medical students' experience, confidence and competence. *Med Educ*. 2002;36:534–539.
37. Risucci DA, Tortolani AJ, Ward RJ. Ratings of surgical residents by self, supervisors and peers. *Surg Gynecol Obstet*. 1989;169:519–526.
38. Hoerl AE, Fallin HK. Reliability of subjective evaluations in a high incentive situation. *J Roy Stat Soc A*. 1974;137:227–230.
39. Keren G. Facing uncertainty in the game of bridge: a calibration study. *Organ Behav Hum Dec*. 1987;39:98–114.
40. Murphy AH, Winkler RL. Subjective probability forecasting experiments in meteorology: some preliminary results. *Am Meteorol Soc*. 1995;15:120–131.
41. Lichtenstein S, Fischhoff B. Do those who know more also know more about how much they know? *Organ Behav Hum Perform*. 1977;20:159–183.
42. Kee F, Owen T, Leatham R. Offering a prognosis in lung cancer: when is a team of experts an expert team? *J Epidemiol Community Health*. 2007;61:308–313.
43. Alicke MD, Klotz ML, Breitenbecher DL, Yurak TJ, Vredenburg DS. Personal contact, individuation, and the better-than-average effect. *J Pers Soc Psychol*. 1995;68:804–825.
44. Svenson O. Are we all less risk and more skillful than our fellow drivers? *Acta Psychol*. 1981;94:143–148.
45. Grove WM, Zald DH, Lebow BS, Snitz BE, Nelson C. Clinical versus mechanical prediction: a meta-analysis. *Psychol Assess*. 2000;12:19–30.
46. Dalrymple DJ. Sales forecasting practices: results of a United States survey. *Int J Forecast*. 1987;3:379–392.
47. Corey GA, Merenstein JH. Applying the acute ischemic heart disease predictive instrument. *J Fam Pract*. 1987;25:127–133.
48. Ridderikhoff J, van Herk B. Who is afraid of the system? Doctors' attitude towards diagnostic systems. *Int J Med Inform*. 1999;53:91–100.
49. Dawes RM, Faust D, Meehl PE. Clinical versus actuarial judgment. *Science*. 1989;243:1668–1674.
50. Slovic P, Fischhoff B, Lichtenstein S. Accident probabilities and seat belt usage: a psychological perspective. *Accid Anal Prev*. 1978;10:281–285.
51. Henrich L, McClure J, Crozier M. Effects of risk framing on earthquake risk perception: life-time frequencies enhance recognition of the risk. *Int J Disaster Risk Reduct*. 2015;13:145–150.
52. Kruger J, Dunning D. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *J Pers Soc Psychol*. 1999;77:1121–1134.