

### Misleading P-Value: Do You Recognise It?

We read the paper by Huak<sup>1</sup> with great interest. The author is to be congratulated for his view regarding the biostatistics. Although his paper is interesting, some considerations should be addressed.

Huak<sup>1</sup> suggested that when statistical significance was reached, the manuscript stood a better chance of getting published (scenario 1 and 3 in Table 3). In contrast, a study with negative (statistically insignificant) results would have a lower possibility of getting published if its probability value (P-value) was of  $>.05$  (scenario 2). A study that lacks both clinical and statistical significance would not merit inclusion in the literature (scenario 4).

Indeed, this doctrine seems to be wrong. It contributes to the so-called ‘*positive outcome bias*’ or ‘*pipeline bias*’, a common form of ‘*publication bias (PB)*’. PB influences the chances of publication and the tendency of investigators, reviewers, and editors to submit or accept manuscripts for publication based on the direction or strength of the findings of quantitative studies. It can occur in any step before a research paper is published. The positive outcome bias decreases the tendency of a manuscript being published when its results are near the null, not statistically significant, or otherwise less interesting.<sup>2,3</sup>

Several studies including Cochrane reviews demonstrated that studies with positive findings were given priority in publication compared with those with inconclusive or invalidating results or with findings contrary to the study hypotheses.<sup>2</sup> This preference can mislead readers about the effectiveness of the reported therapy<sup>2</sup> and inflate the rate of type I (false-positive) error of a meta-analysis.<sup>4</sup>

The Declaration of Helsinki (Article 30) clearly states that ‘*Authors, editors and publishers all have ethical obligations with regard to the publication of*

*the results of research*. Authors have a duty to make the results of their research on human subjects publicly available and are accountable for the completeness and accuracy of their reports. They should adhere to accepted guidelines for ethical reporting. *Negative and inconclusive as well as positive results should be published or otherwise made publicly available*’.<sup>5</sup> Hence, it is the moral responsibility of researchers, fund givers and journals to distribute research findings, regardless of the outcome. Authors and fund providers should not have a preference to submit only studies reporting positive results. Meanwhile, journals should implement the ‘*must have*’ measures to diminish PB from their selection processes. In this way, scientific integrity will be upheld and maintained.<sup>2</sup> Bias in the dissemination of research, publication, interpretation and review of scientific findings is considered as ‘*scientific misconduct*’.<sup>6</sup> For details on publication bias, we refer to our recent publication.<sup>2</sup>

Moreover, the P-value does not provide a good measure of the strength of evidence against the null hypothesis of no difference (no association between a characteristic and an outcome), even though it is often interpreted in this way.<sup>7</sup> A small P-value signifies that the evidence in favour of the null hypothesis is weak and that the likelihood of the observed differences due to chance is so small that the null hypothesis is unlikely to be true.<sup>3,8-10</sup> The rejection of the null hypothesis (when a P-value of  $<.05$ ) must be based on the limitations/assumptions that (1) there is up to 5% chance of a type I error of finding a difference where there is none, (2) there is 50% chance of a type II error of finding no difference where there is one, (3) the data are normally distributed, (4) they follow exactly the same distribution as that of the population from which the sample was taken.<sup>10</sup>

Conversely, a P-value of  $>.05$  only indicates that the evidence is inadequate to reject the null hypothesis, and the alternative hypothesis (the

opposite of the null hypothesis) that the observed differences between the groups is real or not due to chance is not accepted. As a consequence, the study results are unlikely to have occurred by chance.<sup>3,7-11</sup> However, it does not imply that the null hypothesis is true, and that the test treatment and control (e.g. standard treatment, placebo, or baseline) in the study are equivalent.<sup>3</sup> The study itself may have the weakness such as a small sample size to detect a clinically important difference as statistically significant.<sup>9</sup> For example, a P-value of 0.08, albeit not significant, does not mean 'nil'. There is still an 8% chance that the null hypothesis is true.<sup>7</sup> A P-value alone cannot be used to accept or reject the null hypothesis. The cut-off level of 0.05 is purely arbitrary and gives no indication as to the clinical significance of any observed differences.<sup>3,7-9</sup>

There is not much difference between a P-value of 0.055 and a P-value of 0.045. Small changes in sample size can tilt the P-value from one side of the cut-off to the other.<sup>8,9</sup> Any small difference will be statistically significant ( $P < .05$ ) if the sample size is large enough, regardless of the clinical relevance. In contrast, any large difference, no matter how clinically important, will not be statistically significant ( $P > .05$ ). Hence, a low P-value in a small study is more evidential than the same P-value in a large study. However, it will increase the effect of PB.<sup>3,7,9</sup>

Hypothesis testing using a P-value is a binary decision: yes/no, so it is not reliable. If we test 1000 null hypotheses, while only 10% of which are false, at the levels of  $\alpha$  (probability of type I error, which is the probability of rejection of a correct null hypothesis) = 5% and power ( $1 - \beta$ , which is the probability of correctly rejecting the null hypothesis) = 80%, there will be 64% true-positive and 36% false-positive significant results. This means that 36% of the significant P-values will not report the true differences between the 2 treatments.<sup>11</sup>

Until now, some journals have no longer considered P-values, and many prestigious journals such as the *Lancet* and the *British Medical Journal* prefer the effect range estimation (confidence interval: CI) rather than the hypothesis testing.<sup>3,7</sup> The P-value is less informative, can be deduced from the CI and conveys no information on clinical importance. A low or high P-value does not prove anything with regard to the effectiveness of an intervention: a P-value of 0.001 does not reflect a

larger effect than a P-value of 0.04.<sup>8,9</sup> Judgements on the clinical importance of a result should be based on the size of the effect seen rather than the P-value.<sup>9</sup> This concept is in agreement with the International Committee of Medical Journal Editors (ICMJE)'s recommendations, the Consolidated Standards of Reporting Trials (CONSORT) statement, and the Quality of Reporting of Meta-analyses (QUORUM) statement.<sup>3</sup>

CIs are used to infer information of a population based on data obtained from a representative sample from that population. The width of CIs indicates the size and direction of the effect, the amount of random error and the precision of the estimate. The wider the CI, the less is the precision, suggesting more data should be collected before any firm conclusion can be drawn from the results.<sup>8</sup> Appropriate interpretation of P-values, CIs and statistical significance was extensively reviewed by many authors.<sup>3,7,8,10-12</sup>

Taken together, it is a gross misconception that the decision to submit, accept or reject a manuscript for publication relies on the P-value. A study of good quality should be published, regardless of statistical significance.

Poramate Pitak-Arnnop  
Kittipong Dhanuthai  
Alexander Hemprich  
Niels Christian Pausch  
Leipzig, Germany, Bangkok, Thailand.

## REFERENCES

1. Huak CY. Are you a P-value worshipper? *Eur J Dent* 2009;3:161-164.
2. Pitak-Arnnop P, Sader R, Rapidis AD, Dhanuthai K, Bauer U, Hervé C, et al. Publication bias in oral and maxillofacial surgery journals: An observation on published controlled trials. *J Craniomaxillofac Surg* 2010;38:4-10.
3. du Prel JB, Hommel G, Röhrig B, Blettner M. Confidence interval or P-value?: Part 4 of a series on evaluation of scientific publications. *Dtsch Arztebl Int* 2009;106:335-339.
4. Borm GF, Donders AR. Updating meta-analyses leads to larger type I errors than publication bias. *J Clin Epidemiol* 2009;62:825-830.
5. The Declaration of Helsinki. World Medical Association Web site. Available at: <http://www.wma.net/en/30publications/10policies/b3/index.html>. Accessed March 28, 2010.

6. Pitak-Arnnop P, Schouman T, Bertrand JC, Hervé C. Comment éviter la non-conformité de la recherche biomédicale? Recommandations aux chirurgiens. *J Chir (Paris)* 2008;145:534-541.
7. Grunkemeier GL, Wu Y, Furnary AP. What is the value of a P-value? *Ann Thorac Surg* 2009;87:1337-1343.
8. Akobeng AK. Confidence intervals and P-values in clinical decision making. *Acta Paediatr* 2008;97:1004-1007.
9. Young J. Understanding statistical analysis in the surgical literature: Some key concepts. *ANZ J Surg* 2009;79:398-403.
10. Cleophas TJ. Clinical trials: Renewed attention to the interpretation of the P values—Review. *Am J Ther* 2004;11:317-322.
11. Biau DJ, Jolles BM, Porcher R. P value and the theory of hypothesis testing: An explanation for new researchers. *Clin Orthop Relat Res* 2010;468:885-892.
12. Koretz RL. Is statistical significance always significant? *Nutr Clin Pract* 2005;20:303-307.

### Author's response

I totally agree with comments by Pitak-Arnnop et al which emphasizes the importance of clinical relevance (though not statistically significant) and 95% CI confidence interval in today's publication process - this change of mindset had evolved over many years of education on research & biostatistics. I must clarify that the present article is highlighting the well-known bias of significant P-value publication and hopefully to educate the researchers on this aspect - look at the clinical relevance rather than statistical significance. Yes, today, researchers like Pitak-Arnnop et al understood this concept of this poor decision using the P-value and I thank him for adding the important information for the follow-up of this present article (which is constraint by length).

Chan YH  
Singapore

## Microleakage of Flowable Composite Restorations

I read with great interest the article entitled "Effects of different light curing units/modes on the microleakage of flowable composite resins" by A. Ruya Yazici et al which has been published in your esteemed journal (*Eur J Dent* 2008;2:240-246). I want to share few of my thoughts regarding this study. It was a good study comparing different curing units/modes and flowable composite

resins, but the authors can further redefine the study by incorporating few parameters. First, the teeth with class V cavities which were restored with flowable composite resin could have been subjected to occlusal loading prior to microleakage testing. Studies have proven that, the teeth (especially posteriors) are subjected to heavy occlusal stresses at the cervical area during the normal function and parafunction. These stresses tend to flex the teeth. As the teeth flexes, tensile and shear stresses are generated at the cervical region.<sup>1</sup> These stresses may cause debonding of composite resin from the cavity margins and thereby increases the potential for microleakage. Second, dye leakage studies can be conducted under vacuum pressure, since the validity of dye leakage studies has been questioned because of the possible effect of entrapped air on the ingress of the dye solution.<sup>2</sup> Studies have reported that vacuum pressure decreases the volume of entrapped air and allows complete dye penetration.<sup>3</sup> Hence, authors can perform further similar studies using the above mentioned parameters for the better appreciation of the results.

Vasudev Ballal  
India

### REFERENCES

1. Rees JS. The effect of variation in the occlusal loading on the development of abfraction lesions: A finite element study. *J Oral Rehab* 2002;29:188-193.
2. Spradling PM, Senia S. The relative sealing ability of paste type filling materials. *J Endod* 1982;8:543-549.
3. Oliver CM, Abbott PV. Entrapped air and its effect on dye penetration of voids. *Endod Dent Traumatol* 1991;7:135-138.

### Author's response

The aim of this study was to evaluate the effects of different light curing units/modes on the microleakage of flowable composite resins. If the specimens were subjected to load cycling, they might similarly resistant to load cycling as the other parameters were constant. On the other hand the load cycling parameter could be a scope of another study. There are also contradictory findings related with load cycling. Yap reported that mechanical load cycling had no appreciable effects on microleakage at the enamel-cement interface but caused an increase in microleakage at the dentine-cement interface.<sup>1</sup> In another study, it