Response to Parodi S. and Haupt R. "A short defence of p-value and statistical significance"

We thank Stefano Parodi and Riccardo Haupt (3) for their insightful comment on our paper (1). They agree with us in strongly discouraging the abuse of P-value, but also conclude that P-value "has important uses, and should remain an important tool for inference in epidemiology" (3). Overall, we partly agree with this latter statement: almost every day we calculate, value, and interpret P-values. What we are suggesting, however, is to avoid the rigid dichotomization of results into "statistically significant" and "statistically not significant" based on the arbitrary conventional threshold (alpha error, or false positive rate) of 0.05 (two-tailed) (6). Instead, when possible and relevant, we try to report and discuss effect estimates (e.g., risks, risk ratios or risk differences) and their associated confidence intervals.

We have some notes on the example that Parodi and Haupt provide to support their second conclusion. They presented a table with results of a hypothetical case-control study and a polytomous, unordered, three level (A, B, and

C) exposure variable. They show that two different pictures emerged when changing the reference category. Taking A as reference, they commented that "no clear association was found", based on the fact that the odds ratio (OR) for category B was 0.72 (95% confidence interval (CI): 0.54-0.95) and OR for category B was 1.2 (95% CI: 0.92-1.7). Taking B as reference, instead, their comment is that a clear association between exposure and disease appears, with an OR=1.4 (95% CI: 1.1-1.8) for category A, and an OR=1.7 (95% CI: 1.2-2.4) for category C. Parodi and Haupt also calculated a P-value of 0.005 (χ^2 with 2 degrees of freedom=10.47, from a likelihood ratio test) for the overall exposure-disease association (i.e., considering all three categories simultaneously). (For less experienced readers, we note that a simple Pearson χ^2 of 10.44, yielding P=0.005, could be calculated using counts of cases and controls).

The last part of the letter by Parodi and Haupt is an interesting critique of confidence intervals (CI) computed in single comparisons, in favor of simultaneous analyses (6). We partly disagree with these statements.

First, we note that simultaneous (joint) and single (separate) comparisons address two different scientific hypotheses (6). If we were interested in the question "Is there an association between exposure and disease?", then we agree that the overall (joint) χ^2 is appropriate. Conversely, if our interest was directed, for some reason, only to category B (e.g., a specific job title that had been analyzed in previous studies), the scientific question would be "Is the risk in category B different from the risk in category A?". In this case, we would calculate and report the single ORs and CIs calculated separately, no matter what the value of P is. (Of course, the appropriateness of category A should be evaluated in some way; often we do have enough *a priori* information for choosing the most appropriate reference category without relying on results of statistical analyses).

Therefore, in the presented example we would start, like Parodi and Haupt did, by calculating a simultaneous statistic (the χ^2) to have an idea of the overall exposure-disease association. However, since the P-value is a bad measure because it mixes effect size and sample sizes (2), in most situations we would proceed by calculating CIs even if the overall P-value was, say, 0.06, or 0.12, or 0.50. This of course depends on our *a priori* view (4): if we had a strong interest for these exposure categories, we would give less importance to the overall P-value. For example, imagine we had only one tenth of cases and controls. The ORs would be almost identical, but the overall χ^2 would be 1.16, with an associated P-value of 0.56. In this case, notwithstanding the high value of P, we would proceed by calculating (and reporting) ORs and Cis that could be possibly used in future systematic reviews and meta-analyses. On the contrary, if this was just one of the many variables analyzed, without a specific interest, we would value the overall P-value more and perhaps we would not go on by calculating CIs.

Second, in our view an exposure-disease association was already apparent in the analysis that used category A as reference. In fact, the OR for category B was lower than the null value and category C showed a 20% increased risk, albeit with wide CIs.

Third, the P-value does not seem to us the necessary way for choosing the reference category. Assuming the reference category is indeed exchangeable, we could have simply looked at the odds (case/control ratio), which were 1.04 (=310/297) for category A, 0.75 (=126/168) for category B, and 1.29 (= 147/114) for category C, and then we could have chosen the one with the lowest odds (category B) as reference.

Fourth, when evaluating dose-response, we think we would maintain category A (assuming it refers to no- or low-exposure) as reference.

In conclusion, we agree with Parodi and Haupt that the abuse of P-value should strongly discouraged (3). We also agree that P-value can sometimes be a useful tool in epidemiology. However, we feel its importance has been largely overestimated. We continue to think that in most situations we should pay much less attention to the value of P and much more to effect estimates and confidence intervals (2, 4, 6).

Dario Consonni
Epidemiology Unit, Fondazione IRCCS Ca' Granda
Ospedale Maggiore Policlinico, Milan, Italy

Pier Alberto Bertazzi
Post-Graduate School of Occupational Medicine,
University of Milan, Milan, Italy

References

1. Consonni D, Bertazzi PA: Health significance and statistical uncertainty. The value of P-value. Med Lav 2017; 108: 327-331
2. Lang JM, Rothman KJ, Cann CI: That confounded P-value. Epidemiology 1998; 9: 7-8
3. Parodi S, Haupt R: A short defence of p-value and statistical significance (Letter). Med Lav 2018; 109: 157-158
4. Rosendaal FR: The p-value: A clinician's disease? Eur J Int Med 2016; 35: 20-23
5. Rothman KJ: A show of confidence. N Engl J Med 1978; 299: 1362-1363
6. Rothman KJ, Greenland S, Lash TL: Modern Epidemiology, 3rd Edition. Philadelphia (PA): Lippincott Williams & Wilkins, 2008: 163; 305-307; 323; 326-327