P Values Quiz
by Laura King, MA, ELS

Directions: Correctly edit the following sentences regarding *P* values. Refer to sections 20.8.2 and 20.9 of the *AMA Manual of Style*. For further explanation of the correct answers, refer to the cited section of the online or print version of the *AMA Manual of Style*.

---

1. All the tests were 2-sided, with $\alpha = .05$ and $P > .05$ considered statistically significant.

ANSWER:
All the tests were 2-sided, with $\alpha = .05$ and $P < .05$ considered statistically significant.

> **Editor's Note:** The α value is the size of the likelihood acceptable to the investigators that a relationship observed between 2 variables is due to chance (the probability of a type I error); usually $\alpha = .05$. If $\alpha = .05$, $P < .05$, not $P > .05$, will be considered significant. In this example, change $P > .05$ to $P < .05$ and call it to the author's attention (§20.9, Glossary of Statistical Terms, p 852-900 in print; see α [alpha], α level entry).

2. Lower literacy was associated with older age, African American ethnicity, and less education ($P < .0001$ for all).

ANSWER:
Lower literacy was associated with older age, African American ethnicity, and less education ($P < .001$ for all).

> **Editor's Note:** The smallest *P* value that should be expressed is $P < .001$ because additional zeros do not convey useful information; therefore, $P < .0001$ should be rounded to $P < .001$ (§20.8.2, Rounding, pp 851-852 in print).

www.amamanualofstyle.com

3. In the multivariate analysis, combined zidovudine and lamivudine (hazard ratio, 2.71 [95% confidence interval, 1.95-3.83]; $P < .001$), nevirapine (1.95 [1.01-3.81]; $P = .050$), and comedication for an opportunistic infection (2.24 [1.19-4.21]; $P < .01$) were associated with higher rates of treatment modification.

ANSWER:
In the multivariate analysis, combined zidovudine and lamivudine (hazard ratio, 2.71 [95% confidence interval, 1.95-3.83]; $P < .001$), nevirapine (1.95 [1.01-3.81]; $P = .05$), and comedication for an opportunistic infection (2.24 [1.19-4.21]; $P < .01$) **[author query: Please provide exact P value]** were associated with higher rates of treatment modification.

> **Editor's Note:** P values should be expressed to 2 digits to the right of the decimal point (regardless of whether the P value is significant), unless $P < .01$, in which case the P value should be expressed to 3 digits to the right of the decimal point; therefore, in this example, $P < .001$ is correct, $P = .050$ should be rounded to $P = .05$, and the author should be queried for the exact 3-digit P value for $P < .01$ (§**20.8.2**, Rounding, pp 851-852 in print).

4. Cumulative incidence through 6 years was 8.1 per 1000 patients in the tamoxifen group and 11.6 per 1000 patients in the raloxifene group ($P = .048$).

ANSWER:
Cumulative incidence through 6 years was 8.1 per 1000 patients in the tamoxifen group and 11.6 per 1000 patients in the raloxifene group ($P = .048$).

> **Editor's Note:** When rounding P from 3 digits to 2 digits would result in P appearing nonsignificant, such as $P = .046$, expressing the P value to 3 places may be preferred; therefore, $P = .048$ should be left as is (§**20.8.2**, Rounding, pp 851-852 in print).

5. Uterine cancer was less frequent in the raloxifene group, but the difference was not statistically significant (P = NS).

ANSWER:
Uterine cancer was less frequent in the raloxifene group, but the difference was not statistically significant (P = NS) **[author query: Please provide exact *P* value even though nonsignificant]**.

> **Editor's Note:** *P* values should not be listed simply as *not significant* or *NS* because for a meta-analysis the actual values are important and not providing exact *P* values is a form of incomplete reporting (§**20.9**, Glossary of Statistical Terms, p 852-900 in print; see *P* value entry).

6. Baseline mental stress–induced ischemia was associated with significantly higher rates of subsequent cardiac events (odds ratio, 2.8; 95% confidence interval, 1.0-7.7; P < .05).

ANSWER:
Baseline mental stress–induced ischemia was associated with significantly higher rates of subsequent cardiac events (odds ratio, 2.8; 95% confidence interval, 1.0-7.7; P < .05) **[author query: Please provide exact *P* value]**.

> **Editor's Note:** The actual *P* value should be expressed, rather than expressing a statement of inequality (P < .05), unless P < .001 (§**20.9**, Glossary of Statistical Terms, pp 852-900 in print; see *P* value entry).

7. The relative risk for electrocardiography-defined ischemia during exercise testing was 1.9 (95% confidence interval, 0.95-3.96; P = .075), and the relative risk for ambulatory electrocardiography ischemia was 0.75 (95% confidence interval, 0.35-1.64; P = .485).

ANSWER:
The relative risk for electrocardiography-defined ischemia during exercise testing was 1.9 (95% confidence interval, 0.95-3.96; P = .08), and the relative risk for ambulatory electrocardiography ischemia was 0.75 (95% confidence interval, 0.35-1.64; P = .48).

> **Editor's Note:** Because *P* values are reported to 2 digits (unless the first 2 digits are zero), in values with more than 2 digits the digits to the right of the last significant digit are rounded up or down. If the digit immediately to the right of the last significant digit is 5, with either no digits or all zeros after the 5, the last significant digit is rounded up if it is odd and not changed if it is even (§**20.8.2**, Rounding, pp 851-852 in print).

OXFORD
UNIVERSITY PRESS

JAMA
&
ARCHIVES
JOURNALS
American Medical Association

www.amamanualofstyle.com

8. Only 2 patients experienced stent thrombosis; both patients were in the sirolimus-eluting stent group, and the thrombosis occurred after 30 days ($P = 1.0$).

ANSWER:
Only 2 patients experienced stent thrombosis; both patients were in the sirolimus-eluting stent group, and the thrombosis occurred after 30 days ($P > .99$).

> **Editor's Note:** P values should never be rounded up to 1.0 or down to 0. P values may approach infinitely close to these upper and lower bounds but never close enough to establish that the associated observation was either absolutely predestined ($P = 1.0$) or absolutely impossible ($P = 0$). Thus, very large and very small P values should always be expressed as $P > .99$ and $P < .001$, respectively (§**20.8.2**, Rounding, pp 851-852 in print).

9. The mean dysphagia score was 2.81 preoperatively and 0.81 postoperatively ($P = 0.43$).

ANSWER:
The mean dysphagia score was 2.81 preoperatively and 0.81 postoperatively ($P = .43$).

> **Editor's Note:** *JAMA* and the *Archives* Journals do not use a zero to the left of the decimal point because statistically it is not possible to prove or disprove the null hypothesis completely when only a sample of the population is tested (P cannot equal 0 or 1, except by rounding) (§**20.9**, Glossary of Statistical Terms, pp 852-900 in print; see P value entry).

OXFORD
UNIVERSITY PRESS

JAMA
&
ARCHIVES
JOURNALS
American Medical Association

www.amamanualofstyle.com

10. The change in the Cutaneous Lupus Erythematosus Disease Area and Severity Index score was statistically significant for patients who had a meaningful change in their global skin self-ratings ($z = 1.07$; $P = .03$) and approached statistical significance for patients who had a meaningful change in their level of itching ($z = 1.83$; $P = .06$) and their physicians' global skin rating ($z = 1.84$; $P = .06$).

ANSWER:
The change in the Cutaneous Lupus Erythematosus Disease Area and Severity Index score was statistically significant for patients who had a meaningful change in their global skin self-ratings ($z = 1.07$; $P = .03$) and was not statistically significant for patients who had a meaningful change in their level of itching ($z = 1.83$; $P = .06$) and their physicians' global skin rating ($z = 1.84$; $P = .06$).

> **Editor's Note:** If a study does not meet the significance criterion (for example, if the α level was set as < .05, and the *P* value for the finding was .08), authors sometimes describe the findings as "trending toward significance," "having a trend toward significance," "approaching significance," "borderline significant," or "nearly significant." None of these terms is correct. Results do not trend toward significant—they either are or are not statistically significant based on the prespecified study assumptions. Similarly, the results do not include any movement and so cannot "approach" significance. The term *trend* should only be used when reporting the results of statistical tests for trend. Other uses of *trend* or *approaching significance* should be removed and replaced with a simple statement of the findings and the phrase *not statistically significant* (or the equivalent) (Bucking the "Trend" and Approaching "Approaching Significance," Tip of the Month, *AMA Manual of Style* Web site, http://www.amamanualofstyle.com/oso/private/content/jama/subsonly/tipofthemonth.html).