Exposing the P value fallacy to young residents

Piersante Sestini and Stefania Rossi

School of Medicine, University of Siena

5th International Conference of Evidence-Based Health Care Teachers & Developers
Taormina, October 29, 2009
The P value fallacy

The fallacy underlying the widespread use of the P value as a tool for inference in critical appraisal is well known, still little is done in the education system to correct it.


We used a questionnaire to ask 16 residents in respiratory diseases (age range 26-38 yrs, average 30) about their use and knowledge of the P value.

We devised a simple intervention to improve understanding of the p value and to promote the use of the Bayes Factor.

The P value fallacy

- The fallacy underlying the widespread use of the P value as a tool for inference in critical appraisal is well known, still little is done in the education system to correct it.


- We used a questionnaire to ask 16 residents in respiratory diseases (age range 26-38 yrs, average 30) about their use and knowledge of the P value.

- We devised a simple intervention to improve understanding of the p value and to promote the use of the Bayes Factor.

The fallacy underlying the widespread use of the P value as a tool for inference in critical appraisal is well known, still little is done in the education system to correct it.


We used a questionnaire to ask 16 residents in respiratory diseases (age range 26-38 yrs, average 30) about their use and knowledge of the P value.

We devised a simple intervention to improve understanding of the p value and to promote the use of the Bayes Factor.

The P value fallacy

- The fallacy underlying the widespread use of the P value as a tool for inference in critical appraisal is well known, still little is done in the education system to correct it

- We used a questionnaire to ask 16 residents in respiratory diseases (age range 26-38 yrs, average 30) about their use and knowledge of the P value.

- We devised a simple intervention to improve understanding of the p value and to promote the use of the Bayes Factor
### Does the P value matters to you?

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do you know enough about the P value?</td>
<td>93%</td>
</tr>
<tr>
<td>Average patients per day</td>
<td>8 ± 1</td>
</tr>
<tr>
<td>Average articles read per week</td>
<td>4 ± 1</td>
</tr>
<tr>
<td>Do you see the P as the main parameter of a result?</td>
<td>75%</td>
</tr>
<tr>
<td>Average drug reps contacts per week</td>
<td>3.2 ± .3</td>
</tr>
<tr>
<td>Do the drug reps present sponsored research?</td>
<td>100%</td>
</tr>
<tr>
<td>Do the reps present the P value as the main result?</td>
<td>93%</td>
</tr>
</tbody>
</table>
What does a p value of 0.05 mean?

Multiple answers accepted

1. If there is no difference, the probability of getting this or a more extreme result is 5%
2. The probability that there is no difference is 5%
3. The probability that the groups are different is 95%
4. The minimum difference between the groups is 5%
5. If repeating the experiment, the probability of obtaining again the same result is 5%
6. I have no idea
What does a p value of 0.05 mean?

Multiple answers accepted

1. If there is no difference, the probability of getting this or a more extreme result is 5%
2. The probability that there is no difference is 5%
3. The probability that the groups are different is 95%
4. The minimum difference between the groups is 5%
5. If repeating the experiment, the probability of obtaining again the same result is 5%
6. I have no idea
Definition of the P value

The probability of obtaining a test statistic at least as extreme as the one that was actually observed, assuming that the null hypothesis is true.
What does a p value of 0.05 mean?

**Possible answers**

<table>
<thead>
<tr>
<th>Option</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>The probability if there is no difference is 5%</td>
<td>2</td>
</tr>
<tr>
<td>The probability that there is no difference is 5%</td>
<td>7</td>
</tr>
<tr>
<td>The probability that there is a difference is 95%</td>
<td>11</td>
</tr>
<tr>
<td>The minimum difference is 5%</td>
<td>1</td>
</tr>
<tr>
<td>The probability of obtaining again this result is 5%</td>
<td>2</td>
</tr>
<tr>
<td>I have no idea</td>
<td>2</td>
</tr>
</tbody>
</table>
What a P of 0.05 means...

- Imagine that you came home, and the light is on.
- You want to guess whether someone is in when the light is on.
- A P of 0.05 is like knowing that when nobody is in (the null hypothesis), the light is on only 5% of the times.
- Does this mean that there is a 95% probability that somebody is in?
What a P of 0.05 means...

- Imagine that you came home, and the light is on.
- You want to guess whether someone is in when the light is on.
- A P of 0.05 is like knowing that when nobody is in (the null hypothesis), the light is on only 5% of the times.
- Does this mean that there is a 95% probability that somebody is in?
What a P of 0.05 means...

- Imagine that you came home, and the light is on.
- You want to guess whether someone is in when the light is on.
- A P of 0.05 is like knowing that when nobody is in (the null hypothesis), the light is on only 5% of the times.
- Does this mean that there is a 95% probability that somebody is in?
What a P of 0.05 means...

- Imagine that you came home, and the light is on.
- You want to guess whether someone is in when the light is on.
- A P of 0.05 is like knowing that when nobody is in (the null hypothesis), the light is on only 5% of the times.
- Does this mean that there is a 95% probability that somebody is in?
Is somebody in?

<table>
<thead>
<tr>
<th></th>
<th>No</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Off</td>
<td>95</td>
<td>?</td>
</tr>
<tr>
<td>On</td>
<td>5</td>
<td>?</td>
</tr>
<tr>
<td>P →</td>
<td>5%</td>
<td>but you want this↑</td>
</tr>
</tbody>
</table>

The P is the probability of finding the light on when nobody’s in (L|P).

But you want the probability that nobody is in when the light is on (P|L).
Example

<table>
<thead>
<tr>
<th></th>
<th>No</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Off</td>
<td>95</td>
<td>?</td>
</tr>
<tr>
<td>On</td>
<td>5</td>
<td>?</td>
</tr>
<tr>
<td>P</td>
<td>5%</td>
<td>but you want this↑</td>
</tr>
</tbody>
</table>

- The P is the probability of finding the light on when nobody’s in (L|P)
- But you want the probability that nobody is in when the light is on (P|L)
Example

<table>
<thead>
<tr>
<th></th>
<th>No</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Off</td>
<td>95</td>
<td>?</td>
</tr>
<tr>
<td>On</td>
<td>5</td>
<td>?</td>
</tr>
<tr>
<td>$P \rightarrow$</td>
<td>5%</td>
<td>but you want this↑</td>
</tr>
</tbody>
</table>

The $P$ is the probability of finding the light on when nobody’s in ($L|P$)

But you want the probability that nobody is in when the light is on ($P|L$)
Bayes Factor (Likelihood ratio)

Minimum BF = \%TP/\%FP = \text{e}^{Z^2/2} = \text{e}^{1.96^2/2} = 6.8

<table>
<thead>
<tr>
<th></th>
<th>No</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Off</td>
<td>95</td>
<td>100 - 32 = 68</td>
</tr>
<tr>
<td>On</td>
<td>5</td>
<td>5 * 6.8 = 32</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>P</td>
<td>5%</td>
<td>32%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>32/37 = 86%</td>
</tr>
</tbody>
</table>
In many experiments, the statistical power is selected to have a false negative rate ($\beta$) of 20% and a false positive rate ($\alpha$) of 5% (under certain assumptions a priori).
Is somebody in?

<table>
<thead>
<tr>
<th></th>
<th>No</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Off</td>
<td>95</td>
<td>20</td>
</tr>
<tr>
<td>On</td>
<td>5</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>P -&gt;</td>
<td>5%</td>
<td>80%</td>
</tr>
</tbody>
</table>

\[
\frac{80}{85} = 94\%
\]

In many experiments, the statistical power is selected to have a false negative rate \((\beta)\) of 20% and a false positive rate \((\alpha)\) of 5% (under certain assumptions \textit{a priori})
### Example

<table>
<thead>
<tr>
<th></th>
<th>No</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Off</strong></td>
<td>95</td>
<td>2</td>
</tr>
<tr>
<td><strong>On</strong></td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td><strong>P -&gt;</strong></td>
<td>5%</td>
<td>80%</td>
</tr>
</tbody>
</table>

\[
\frac{8}{13} = 62\% 
\]

- \(\alpha\) and \(\beta\) errors alone do not adjust for low pre-test probabilities.
Is somebody in?

<table>
<thead>
<tr>
<th></th>
<th>No</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Off</td>
<td>95</td>
<td>2</td>
</tr>
<tr>
<td>On</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>10</td>
</tr>
</tbody>
</table>

8/13 = 62%

$P \rightarrow 5\% \quad 80\%$

α and β errors alone do not adjust for low pre-test probabilities.
Post-intervention

- Many got confused by the abstraction
- In retrospect, a diagnostic example could have been better (e.g. nausea and pregnancy)
- After the course all except two identified the meaning of the P value
- All except one declared that in the future they would use the Bayes Factor instead.
- After 3 months, no one has ever done it
Results

Post-intervention

- Many got confused by the abstraction
- In retrospect, a diagnostic example could have been better (e.g. nausea and pregnancy)
- After the course all except two identified the meaning of the P value
- All except one declared that in the future they would use the Bayes Factor instead.
- After 3 months, no one has ever done it
Post-intervention

- Many got confused by the abstraction
- In retrospect, a diagnostic example could have been better (e.g. nausea and pregnancy)
- After the course all except two identified the meaning of the P value
- All except one declared that in the future they would use the Bayes Factor instead.
- After 3 months, no one has ever done it
Post-intervention

- Many got confused by the abstraction
- In retrospect, a diagnostic example could have been better (e.g. nausea and pregnancy)
- After the course all except two identified the meaning of the P value
- All except one declared that in the future they would use the Bayes Factor instead.
- After 3 months, no one has ever done it
Results

Post-intervention

- Many got confused by the abstraction
- In retrospect, a diagnostic example could have been better (e.g. nausea and pregnancy)
- After the course all except two identified the meaning of the P value
- All except one declared that in the future they would use the Bayes Factor instead.

*After 3 months, no one has ever done it*
The P value is frequently used by drug reps to promote the results of drug trials and by young residents for interpreting them.

A simple, short intervention in a group with poor understanding of the P value is able to dramatically improve knowledge.

However, such interventions are unlikely to have a significant impact, unless major changes occur in the medical community.
Conclusions

The P value is frequently used by drug reps to promote the results of drug trials and by young residents for interpreting them.

A simple, short intervention in a group with poor understanding of the P value is able to dramatically improve knowledge.

However, such interventions are unlikely to have a significant impact, unless major changes occur in the medical community.
The P value is frequently used by drug reps to promote the results of drug trials and by young residents for interpreting them.

A simple, short intervention in a group with poor understanding of the P value is able to dramatically improve knowledge.

However, such interventions are unlikely to have a significant impact, unless major changes occur in the medical community.