Bioelectromagnetic Research and Statistical Power

John Podd, Wyatt Page, Bruce Rapley, Ivan Beale
Departments of Psychology and Production Technology, Massey University, Palmerston North, Psychology Department, University of Auckland, New Zealand

Abstract: Bioelectromagnetic field research has some way to go before the scientific community at large is convinced that weak, extremely low frequency electromagnetic fields affect biological systems. This paper argues that researchers are doing themselves and the discipline an injustice by running studies where statistical power is very low. The most frequent outcome of low statistical power is a null result, even when there are experimental effects to be detected. It is suggested that future research must address the issue of statistical power, both through increasing it in individual studies and through conducting meta-analytic investigations of existing research.

INTRODUCTION
The effects of weak, environmental-level, extremely low frequency (ELF) electromagnetic fields (EMFs) on biological systems have been studied over many years. Despite several examples of clear, replicable effects, the scientific research community at large remains sceptical that these EMFs do in fact perturb biological systems to the extent claimed. Null results are frequently obtained, and even when an effect is demonstrated it sometimes is difficult to replicate [1,2].

Why are we finding it difficult to establish EMF effects on biological systems that can be easily and consistently replicated? There are many reasons, but one possibility that has received scant attention is that the experimental and analytical techniques employed to detect small EMF effects are not sensitive enough. In other words, there has been insufficient statistical power to reveal effects in the presence of a low signal-to-noise ratio [3].

STATISTICAL POWER
When a decision is made about an experimental outcome, it almost always has to be made in the face of uncertainty. For example, the value of a statistic could have arisen from the hypothesised effect or it could have come about as a result of chance. Under such circumstances, there are four possible outcomes. 1. Accept the experimental hypothesis, \( H_A \), when it is true, \( P(\text{Accept } H_A | H_A) \); 2. Accept the hypothesis no effect, \( H_0 \), when it is true, \( P(\text{Accept } H_0 | H_0) \); 3. Accept \( H_A \) when it is false, \( P(\text{Accept } H_A | H_0) \); and 4. Accept \( H_0 \) when it is false, \( P(\text{Accept } H_0 | H_A) \). Two types of error can be made and these are shown graphically in Figure 1.

The distributions shown represent hypothetical sampling distributions for some statistic, such as \( F \). The bottom panel illustrates the usual state of affairs: values of the statistic do not allow us to choose unequivocally between \( H_0 \) and \( H_A \), resulting in Type 1 and Type 2 errors. The Type 1 error probability is better known as the "significance" level and, by convention, is usually set at 0.05. The type 2 error probability, \( P(\text{Reject } H_A | H_A) \), has played little part in statistical decision making in weak, ELF EMF studies until recently [5,6].

FIGURE 1: Hypothetical \( H_0 \) (top) and \( H_A \) (middle) distributions, shown combined in the bottom panel. Two types of errors are possible, and as the decision cut-off point is moved these can be seen to be inversely related. (After Keppel [4]).

Figure 1 shows that \( P(\text{Type 2 error}) \) and statistical power are inversely related. That is, statistical power = 1 - \( P(\text{Type 2 error}) \). As power drops, there is an increasing likelihood of concluding there was no experimental effect, even when there was one.

STATISTICAL POWER OF ELF EMF STUDIES
Do ELF EMF studies have sufficient power? To our knowledge, there are very few statistical power analyses of EMF research. One such study [5] calculated the statistical power of weak ELF EMF effects on human physiology and performance experiments (Table 1).
TABLE 1. Mean and median statistical power levels for three effect sizes based on 19 experiments reported in the literature. Data taken from Whittington & Podd [5].

<table>
<thead>
<tr>
<th>Effect Size</th>
<th>Small</th>
<th>Medium</th>
<th>Large</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.08</td>
<td>0.23</td>
<td>0.46</td>
</tr>
<tr>
<td>Median</td>
<td>0.07</td>
<td>0.20</td>
<td>0.43</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.01</td>
<td>0.09</td>
<td>0.16</td>
</tr>
<tr>
<td>95% Confidence</td>
<td>0.08 - 0.09</td>
<td>0.19 - 0.27</td>
<td>0.38 - 0.54</td>
</tr>
</tbody>
</table>

The analysis was based on 19 published experiments for three effect sizes (standardised mean differences between experimental and control groups) [3]. Even if it is assumed the effect size was large on average, the average statistical power is still only 0.46. That is, P(Type 2 error) = 1 - 0.46 = 0.54. For small effect sizes, P(Type 2 error) = 1 - 0.08 = 0.92. Put another way, there was an average 92% chance of concluding that there was no experimental effect when in fact there was one! If one compares the Type 1 and Type 2 error probabilities, making the reasonable assumption that P(Type 1 error) = 0.05, then the relative risk of a Type 2 error is more than 18 times as likely (0.92/0.05) as a Type 1 error.

In exploratory research (which includes most ELF EMF experiments, we suggest), such an imbalance between Type 1 and 2 error probabilities is unacceptable. Even if it is assumed (unrealistically) that weak ELF EMF effects are large, the Type 1/Type 2 error ratio is still about 1:11 (0.54/0.05).

The main problem with running with low statistical power studies is that many real experimental effects are being reported as “not significant”, and many (in fact most) replication attempts have the odds heavily stacked against a successful outcome. Obtaining a null result in the presence of low statistical power (all of the values reported in Table 1), confounds two possible conclusions: Either the null hypothesis is true and there is no EMF effect, or there is an EMF effect but the study lacked sufficient statistical power to detect it. Our analysis of the situation suggests that the odds in favour of the latter conclusion are far too high.

It has been argued [6] that in exploratory research Type 1 and 2 error probabilities should be equal, unless there are good reasons to the contrary. When one is simply trying to find out if any EMF effects exist, we cannot see any argument for running the Type 2 error probability between 11 and 18 times the level of the Type 1 error probability.

Increasing statistical power in individual studies can be achieved by reducing error variance in the design of the study, increasing subject numbers, increasing the significance level (the Type 1 error probability), or by a combination of these approaches. However, subject numbers will usually need to be increased by several orders of magnitude to improve statistical power to the recommended level of 0.8 [3]. Furthermore, altering the significance level may be unpalatable to journal editors and researchers. Decreasing the risk of a Type 2 error at the expense of increasing the risk of a Type 1 error may not be acceptable to some. So, what is to be done? Schmidt [7] proposes one solution, suggesting that additional statistical power can be obtained by combining the results of several similar investigations using the technique of meta-analysis. The statistical power of individual studies may then no longer be an issue.

CONCLUSION

ELF bioelectromagnetic research must pay attention to the level of statistical power available in experiments. Cohen [3] suggests a power level of 0.8 as a minimum. At present, at least in some areas of EMF research, statistical power levels are far too low. Most studies have been set up (unwittingly) to have a very high chance of producing a null result, even though an EMF effect exists. At present, one cannot decide with any confidence whether these null results are due to no ELF EMF effects, or due to low levels of statistical power.

REFERENCES