Ecotoxicology and Environmental Safety 102 (2014) 142–146

Contents lists available at ScienceDirect

Ecotoxicology and Environmental Safety

journal homepage: www.elsevier.com/locate/ecoenv

Power and control choice in aquatic experiments with solvents

John W. Green *

Principal Consultant: Biostatistics, DuPont Applied Statistics, Stine Haskell Research Center, S315/1361, 1090 Elkton Road, Newark, DE 19714-0050, USA

ARTICLE INFO

Article history:
Received 26 September 2013
Received in revised form
13 January 2014
Accepted 16 January 2014
Available online 14 February 2014

Keywords:
Solvents
Control choice
Power
Error rate

ABSTRACT

Aquatic toxicology experiments to determine the effects of chemicals sometimes require the use of a carrier solvent. Such experiments typically include both a negative (water) control group and a solvent control group. False positive rates and power to detect treatment effects in such experiments are compared for six possible strategies for deciding the appropriate control or controls for comparison. The main purpose of the present study is to determine the best use of the two controls in statistical analysis. A secondary purpose is to determine purely on statistical grounds whether both controls are actually needed. The evidence supports using either the solvent control only in all cases or a sequential strategy of combining the water and solvent controls unless the two controls are found to be statistically significantly different, in which case only the solvent control should be used. These results extend, and in some ways contradict, a recently published simulation study.

1. Introduction

1.1. Why solvents are used

Pesticides, herbicides and other chemicals with environmental contamination potential must be tested in aquatic experiments before they can be approved for sale. The experimental design generally calls for a single species to be exposed to a water control and one or more concentrations of the test substance. It is common with difficult chemicals to use a solvent to make preparation of stock solutions easier and to facilitate maintenance of the concentrations of the test substance. When a solvent is used, it is customary for the experiment to include both a water control and a solvent control and to use a solvent that, based on prior experience, is not expected to affect the test organism. Approximately the same concentration of the solvent is used in all concentrations of the test substance and in the solvent control. A great deal of experience with these solvents has found little evidence that they affect the test organisms at the solvent concentrations recommended in the applicable test guidelines (Hutchinson et al., 2006). When a solvent is used, the relevant test guidelines generally require the water and solvent control to be compared statistically to determine their equivalence. The main purpose of the present study is to determine the best use of the two controls in statistical analysis. A secondary purpose is to determine purely on statistical grounds whether both controls are actually needed.

1.2. Uses of solvent data

In an evaluation of a large number of studies that utilized dilution water and solvent controls, there was a statistically significant difference in the mean response of the two controls for approximately 30 percent of the responses evaluated over all studies (Green and Wheeler, 2013). For the 30 percent of responses that differed, roughly half of the mean solvent control responses exceeded the mean water control response with the reverse true in the other cases. The question in these situations is which control to use to determine the NOEC or ECx for the test substance. There are six methods for potentially addressing these situations that have been employed by, or at the request of, some regulatory agency in order to obtain approval for sale of a crop protection chemical. (1) *Repeat the experiment* on the grounds that a significant difference between the two controls indicates a problem with the experiment. (2) *Water only method*, i.e., ignore the solvent control and determine the NOEC or ECx using only the water control on the grounds that it is the only "true" control. (3) *Solvent only method*, i.e. ignore the water control and determine the NOEC or ECx using only the solvent control on the grounds that at recommended solvent concentrations no effects are expected but the solvent effects, if any, are almost always additive with those of the test substance. Therefore, comparing the test substance plus solvent effects to the solvent effects "subtracts" out any solvent effect to reveal the treatment effect. (4) *Always pool method*, i.e., pool the data for the two controls since this increases the power of the statistical tests for treatment effects, uses all the data, and solvents have not typically demonstrated effects at the recommended concentrations used in these experiments. Conceptually, if the solvent has no real effect on the test organisms, then the combined controls provide the best indication of the undisturbed

* Corresponding author. Fax: +1 302 731 7276.
E-mail address: John.W.Green@DuPont.com

0147-6513/$ - see front matter © 2014 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ecoenv.2014.01.024

population. (5) *Protocol method*, i.e., pool the two controls unless a statistically significant difference is found between them, in which case only the solvent control is used for subsequent analyses. (6) *Separate analyses method*, i.e., compare the treatment groups to each control independently and declare a treatment effect if either comparison is statistically significant. This article will help sort out the implications and relative merits of these six methods.

1.3. Previous work on use of solvent data

The statistical protocol method is recommended in several publications (e.g., OECD, 2006). A variation of this method is recommended in some OECD Test Guidelines (e.g., OECD TG 231), where the water control is used instead of the solvent control when there is a significant difference between the controls. In van der Hoeven (2010), calculations were presented to address which of methods 2–5 have desirable statistical properties, with an emphasis on false positive rates. That paper preferred the solvent only method for data analysis, while considering the always pooling method an acceptable alternative. It rejected the water only and protocol methods. The objection to the protocol method was that it inflates the false positive rate. For example, if there is no difference between the water and solvent control, the false positive rate under the protocol method was reported as 0.055 instead of the nominal 0.05. It was further argued that the conditional probability of rejecting the null hypothesis of no treatment effect when there is a significant difference between the two controls is as high as 0.15. This second argument is irrelevant since only the false positive rate of the decision process applies. That rate, 0.055 instead of 0.05 would be a small price to pay if there were a substantial increase in statistical power associated with the sequential method compared to the solvent only or always pool methods. To describe a test as invalid, as was done in van der Hoeven (2010), merely because its false positive rate is not exactly 0.05 is curious, as this is merely a convention, not a requirement of a test. It is unfortunate that van der Hoeven (2010) did not consider carefully how the power to find a treatment effect is influenced by the choice of control. It is a purpose of this paper to explore both false positive rates of all six methods and the power associated with methods 2–6 when there is a treatment effect. Both one- and two-sided tests for treatment effects are examined.

2. Materials and methods

2.1. Description of Previous Simulation Study

Van der Hoeven (2010) conducted a computer simulation study to explore the false positive rate associated with the control used. Using the definitions BC=blank control, SC=solvent control, T=treatment, E. mean=expected mean, and SD=standard deviation, the article states that "Since the probability of wrongly rejecting the hypothesis that the treatment had no effect was investigated, simulations were run with the same expected mean for SC and T, E. mean (SC). The mean of the distribution of the BC, E. mean (BC), was in the range of E. mean (SC) $\pm$ 1.2 SD."

The assumption of van der Hoeven (2010) was that if the treatment has no effect, then the solvent control mean and the mean of the group containing solvent and treatment would be the same. The analysis was only concerned with a one-sided test to determine whether a treatment effect exists. This is sometimes not appropriate, since for some endpoints, an increased response is an indication of effect, whereas for others it is a decreased response that indicates an effect and it is not always known in advance what direction the effect, if any, will take. Therefore, two-sided tests are often used for analysis. More importantly, a simulated solvent effect of 1.2 SD is extremely large in the examples used; with the SD=25 and the mean=125, a 1.2 SD effect is a 24 percent effect level. An experiment would most likely be invalidated, with good reason, were such a difference observed between controls. In Green and Wheeler (2013), 91 percent of the aquatic tests evaluated had differences between the water and solvent controls less than 5 percent and most differences were less than 3 percent. As will be shown below, when there is a 20 percent solvent effect *in the same direction* as the treatment effect, then indeed, the power to detect a treatment effect using the protocol method is lower than under water only or always pool method but higher than under the solvent only method. For smaller solvent effects or solvent effects that are in the opposite direction from treatment effects, the protocol method has distinct advantages.

2.2. Description of current simulation study

Following the general methodology used by van der Hoeven (2010), computer simulations were done with water and solvent controls and one treatment containing both solvent and test chemical. The solvent and treatment effects were simulated to be additive and either synergistic or antagonistic. The data were simulated to be normally distributed with homogeneous variances. Solvent effects were simulated at 0, $\pm$5, $\pm$10, and $\pm$20 percent and treatment effects were simulated at 0, 10, and 20 percent, where –*x* percent means the solvent effect is in the opposite direction from the treatment effect and of magnitude *x* percent. Both one- and two-sided tests for treatment effects were evaluated. Each experiment was simulated 100,000 times with 10 replicates per control and treatment. This approach insures that the estimated power is within 0.3 percent of the true value with 95 percent confidence. A 10-fold increase in the number of simulations to 1,000,000 would make the estimated power within 0.1 percent of the true value with 95 percent confidence. This higher number of simulations was done only for the null case of 0 percent solvent effect and 0 percent treatment effect. All simulations were carried out using SAS version 9.2. Similar simulations were done for 5 and 15 reps per control and treatment and yielded similar findings for power and false positive rates.

3. Results and discussion

3.1. Discussion of the repeat experiment method

Method 1, of repeating the experiment if there is a statistically significance difference between the water and solvent controls for even one response, is unacceptable in terms of the time, costs, and animal welfare concerns (unnecessary use of additional test organisms) associated with repeating tests. Furthermore, given that numerous responses (i.e., endpoints) are tested in each experiment and a 5 percent false positive rate is used on each response for the comparison with the controls, there is a substantial probability of repeating the experiment even when there is no real difference between the two controls. This risk is quantified in Table 1 which illustrate that when there are four or more responses to be analyzed, the risk of repeating the experiment unnecessarily under this method is over 18 percent and that increases if there are more endpoints to be evaluated. Table 1 also shows the false rejection rate if the criterion is changed to repeating the experiment only if two or more responses have significantly different control means. That would be a more defensible strategy on statistical grounds alone. However, both types of estimated rejection rates may be low, since they assume tests on different responses are independent, when in fact some response, such as length and weight, or biomass and growth rate, may be highly correlated so that a significant control difference for one response may well lead to a significant control difference for one or more additional responses. The ethical considerations

Table 1
False rejection rate for control comparisons.

K[a]	N[b]	Prob[c]	K	N	Prob
1	1	0.05	2	1	0
1	2	0.0975	2	2	0.0025
1	3	0.14263	2	3	0.00725
1	4	0.18549	2	4	0.014019
1	5	0.22622	2	5	0.022593

[a] *K* is the number of responses with significantly different control means in one experiment.
[b] *N* is the number of responses evaluated for the experiment.
[c] Prob is the probability of at least *K* significant differences among the *N* control comparisons assuming independent tests.