Problems with conclusions from a single field trial, using only one method of assessment, to compare efficacy of normal versus doubly dosed mouse bait.

Media claims have been made that all currently registered zinc phosphide (ZP) mouse baits do not work despite these being used successfully to protect crops in all regions since 1999. Billions of dollars worth of crops would not have been saved if the products, such as Mouseoff®, were not working.

Instead, a doubling of ZP dosing to 50g/kg of grain has been promoted to overcome the occasional need to re-bait. Sometimes re-baiting is needed but doubling the chemical dosing, to twice the highest dose used globally, may not overcome this need. It has also been recommended in seminars that the doubly dosed product can require repeat baiting. The evidence supporting field failures appears to come from South Australia, where products other than our product MOUSEOFF are often used. In 2021, a single field trial compared mouse control using a competitors manufactured baits containing 25 or 50g/kg ZP. The published trial [**Ruscoe WA et al. (2022) Wildlife Research doi:10.1071/WR22009**] used only one assessment method (capture- recapture) which, though a recognized technique, can be subject to incorrect results if there are mouse movements between test sites and if traps are saturated. It is preferred that multiple supporting measures (e.g., talc marked burrow activity, spotlight surveys, census cards, tracking boards and measured grain uptake pre and post baiting) are used, as we have always done in our development work.

Re-baiting is sometimes needed and is provided on the approved labels, but we believe this is not because the normal approved product is not killing mice. Rather, it is mainly because mice can quickly re- infiltrate baited areas or because spreading just 1 kg/ha is difficult to achieve and if, for example, spreaders run out of bait early, some areas thought to have been baited may not have been baited. Farmers may see residual mice at "hot spots" of extreme local mouse numbers, such as around storage bags or fodder storages. Aerial spreaders sometimes spread with wide swathes and thus only stripes of paddocks may be baited in some cases. Finally, as the crop develops mice can infest at several stages (at planting, tillering and pre-harvest).

The data from the published trial [**Ruscoe WA et al. (2022) Wildlife Research doi:10.1071/WR22009**], shown in our ACTA Newsletter #39, is claimed to validate the need to double the ZP dosing rate. However, the paper, and several in-house studies during ACTA development of MOUSEOFF, show very high levels of control with normally dosed bait and there appears to be **no significant advantage** of doubling the dose rate of ZP. Instead in the published trial, a model was used to claim a higher likelihood of >80% control if the results from one trial are multiplied 10,000 times in a computer simulation. If a plague occurs every 5 years a landholder might be 50,000 years old to obtain this benefit.

As we do not have qualified statisticians in-house, we have sought independent expert advice from an international leader in this field who has strongly challenged the methods and analysis used to make the claims in the Wildlife Research paper. We advised the CSIRO, owner of the Wildlife Research journal, of this high-level statistical advice but they have denied that the analysis and claims in the paper are in error.
The independent expert advice is copied below and raises several concerns about the data analysis and interpretation. There is a divergence of opinion here, so we are calling for an independent review of the data and claims being made.

Abstract

They allude to a BACI design in a replicated trial. However, they do not correctly account for the nature of the replication in their analyses.

They also summarize their nonsense about the inferential ability from their simulation modelling.

Introduction

They compare published LD50 estimates between Australia and US (presumably, considering where Rex Marsh worked).

Methods

Experimental sites

They had 9 sites distributed among 2 blocks, with 3 treatments applied in each block and observed before and after application. Thus, 9 x 2 = 18. They have 18 data points or 17 df, no more. Blocking is used to accumulate experimental units that share commonalities. Thus, block effect must be accounted for in analytical designs. One of their blocks had 3 sites, with one replicate of each treatment, and the other had 6 sites, with 2 replicates of each treatment. They have some form of repeated measures design. I believe their hope was to treat all sites as replicates from one large potential pool of sites. If the sites are all treated as random samples of all potential sites, the appropriate ANOVA format would be:

Source	**DF**
Treatment	2
Site(Trt)	6
Time	1
Time x Trt	2
Time x Site(Trt)	6

The error term for treatment is site(treatment), but the only term of real interest is time x treatment which has time x site(trt) as its error term (as does time). So, they have a relatively insensitive 2,6 df test for treatments before and after, IF THIS WAS THE CORRECT ANALYTICAL DESIGN. Main effect of treatment would be tested against site(trt), which is of little interest unless no mice were lost.

However, this really was not their true experimental design, especially since their results showed that there was a difference among their 2 blocks of sites, which were part of their a priori design. Therefore, A BLOCK EFFECT SHOULD HAVE BEEN TAKEN INTO ACCOUNT in their analyses. This ANOVA format would look like:

Source	DF
Treatment	2
Block	1
Trt x Blk	2
Time	1
Time x Blk	1
Time x Treatment	2
Time x Trt x Blk	2
Site (TimexTrtxBlk)	6

The error term for treatment is treatment x block (2,2 df test). The error term for time is time x block (1,1 df test). The error term for the effect of interest, time x treatment, is time x treatment x block (2,2 df test). Clearly, a very weak test on the effect of interest. Site (TimexTrtxBlk) doesn't serve to test any effect.

Their analyses do not appear to consider blocking effect, even though this was part of their a priori design and was well-demonstrated by their data to be an impactful effect.

Experimental treatments
It is interesting to note that the 25mg/kg treatment used does not well-represent current practices since it was not obtained from the supplier that most of Australia turns to.

Population monitoring
The first 2 nights post-treatment were wet. ZnP is vulnerable to wetness. Thus, a differential treatment effect could be due to wetness impacting 25 vs 50mg/kg ZnP baits.

MR population estimation
As for their MR approach, they say they account for heterogeneity, but it does not appear that includes demographic heterogeneity of gender and age. I did not look at their capture data, but I have real problems with MR in general, as shown in the papers I sent you previously. The number known to be alive would probably have been better suited. Not knowing exactly what they did, I am concerned about their modelling of heterogeneity through a Bayesian “DATA AUGMENTATION.” This strikes me as making up numbers to find estimates of heterogeneity, with heterogeneity being the bane of MR estimation. In fact, in the literature I sent you, it is stated that without a sufficient design, captures, and recaptures, the data requirement for valid estimation in the face of real (not augmented data) heterogeneity can be in excess of the total population.

BACI analyses
Much of their analytical methodology is indecipherable without sitting down with them for an explanation. Nevertheless, if you apply an analysis the doesn’t accurately take into account the actual experimental design, then the results are meaningless. If the true design had been accounted for using the correct error terms, then the inferentially weak design would have been hard-pressed to find any differences.

Moreover, it seems to me that if you take a small experiment with a weak design and input those results as a foundation for a large Monte Carlo simulation, you'll just magnify those weak results into a hypothetical output predestined by the weak input. They really do not have 10,000 observations, just 18.

Results

Population size estimation

Re-treatment numbers estimated between 300-500, almost 2-fold.

That population size was uniformly larger on block 1 proves that block should have been a factor in the design and determined the appropriate error terms.

Again, the MR estimates did not account for all potential forms of demographic heterogeneity, and the forms they did account for required "data augmentation."

BACI analysis

All based on a flawed analytical design, so meaningless.

Paragraph beginning "Assuming at least an 80%..." is not really based on data, but magnified, self-fulfilling, simulated data based on a flawed analytical design.

Discussion

I do not believe this experiment merits discussion.