Landcare Research, P.O. Box 69, Lincoln, New Zealand.

ASSESSING THE EFFECT OF POISONING PROGRAMS ON THE DENSITY OF NON-TARGET FAUNA: DESIGN AND INTERPRETATION

Summary: To establish whether poisoning programs affect non-target density, the null hypothesis that density does not decline on poisoned sites needs to be tested. However, where no statistically significant reduction in density is found, there is some probability that a biologically significant reduction has been overlooked. The probability that such an error has occurred (a Type 2 error) depends on the effect poisoning has on non-target density, the precision with which the reduction is assessed, and the number of poisoning operations sampled. Prospective power analysis can identify minimum sample sizes that reduce the probability of a Type 2 error to acceptable levels. Equivalence tests require a priori identification of the minimum change in non-target density that can be safely overlooked and the acceptable probability of doing so. As such, they explicitly link the statistical and biological significance of non-target poisoning assessments. We illustrate these principles using an experimental assessment of the effect rabbit poisoning has on the density of large kangaroo populations in Australia. A rule-of-thumb guide was used to estimate appropriate levels of power (0.85) and reductions in kangaroo density (r = -0.12) for the assessment, and a pilot study conducted to estimate the between-sample standard deviation for estimates of change in kangaroo density (s = 0.089). Prospective power analysis based on these estimates indicated that 6 poisoning programs would provide a robust assessment of the effect of poisoning on kangaroos. However, because the between-sample standard deviation was underestimated, a subsequent assessment based on 6 samples had insufficient power to usefully estimate the effect poisoning had on kangaroos. Retrospective power analysis indicated that at 0.85 power, reductions in kangaroo density as high as r = -2.2 may have been overlooked. Using the between-sample standard deviation from this assessment, changes in kangaroo density would have to be estimated for 17-19 poisoning programs if a subsequent experiment was to achieve a biologically as well as statistically robust result.

Keywords: Poisoning; non-target; power; hypothesis testing; *Macropus* spp.; equivalence testing; rabbit control.

Introduction

The potential that poisoning programs have to affect the viability of non-target wildlife populations continues to cause public concern about the acceptability of poisoning as a broad-acre pest management technique (Livingstone, 1994). These misgivings have prompted a considerable number of studies on the effect of poisoning on non-target fauna. These studies concerned either the direct effect of poisons (i.e., where responses are measured after individual representatives of non-target species have been dosed with the poison of interest), or the effect of poisoning programs (i.e., where responses are measured after the non-target species is exposed to poisoning programs conducted under field conditions). Studies on non-target species can be further divided into those that estimate (1) susceptibility or exposure of non-target species to poison, (2) mortality of non-target species by determining differential mortality of non-target individuals exposed or not exposed to poisoning programs, and (3) changes in the density of non-target species due to a poisoning program. A review of relevant papers in the *Journal of Wildlife Management* between 1967 and 1997 reveals that studies of the direct effect of poisoning on non-target species have been far more numerous than studies of the effects of poisoning programs, and for the latter, most studies have estimated susceptibility or mortality of non-target species with very few focussing on changes in non-target species density (Table 1).

Studies that assess the direct effects of poisons provide incomplete information on non-target impacts because they cannot account for (1) the probability that individuals will be poisoned during a program, or (2) the potential for compensatory demographic or dispersal responses to any mortality that results from a poisoning program. For example,

Table 1: *Numbers of papers concerning the effects of poisoning on non-target species in* The Journal of Wildlife Management *between 1967 and 1997, divided according to whether they assessed the direct effect of poisons or the effect of poisoning programs, and whether they were based on estimates of susceptibility, mortality or change in density. Percentages for each class are given in parentheses.*

Estimated parameter	Direct assessment of poisons	Assessment of poisoning programs
Susceptibility	10 (21)	4 (22)
Mortality	38 (79)	12 (67)
Change in density	-	2 (11)
Total (%)	48 (100)	18 (100)

lethal dose studies are commonly used to gauge the potential effect poisoning may have on non-target species by estimating their susceptibility to a poison (e.g., McIlroy, 1985; Twigg and King, 1989). However, susceptibility alone provides little information on the effect that exposure to a poisoning program may have on the density of non-target species. While studies on non-target susceptibility or exposure during poisoning programs indicate the potential for reductions in non-target density, they provide no information on whether or not reductions actually occur. Estimates of non-target mortality by searches for dead animals or the rate at which radio-tagged individuals die are also incomplete because they cannot account for compensatory demographic or dispersal responses Such responses include enhanced survival and/or reproduction (Sinclair, 1989), or higher rates of dispersal into the poisoned area (Schieck and Millar, 1987; Nakata and Satoh, 1994). Hence, an apparent increase in non-target mortality on poisoned sites may reflect a transient demographic response rather than a longer-term reduction in population density.

Results

The effect of poisoning programs on non-target species: hypothesis testing

A reduction in the density of a non-target population that can be directly linked to a poisoning program represents the most unequivocal evidence that the program has had an effect. This requires comparisons to be set in a hypothesis testing framework, where appropriate null (H_o) and alternative (H_a) hypotheses are identified and assessed using replicated experiments. One set of null and alternative hypotheses that could be used to determine whether poisoning reduced the density of a non-target species to unacceptable levels are:

$$H_o: (r_p - r_u) \geq r_g \text{ and } H_a: (r_p - r_u) < r_g$$

where r_p is the change in the density of the non-target population resulting from a poisoning program, r_u is the change in the density for a non-target population that is not exposed to the poisoning program and r_g is the reduction in the density of the non-target population that is considered acceptable. One-tailed hypotheses are used because instances where poisoning leads to significant increases in non-target density generally cause little concern.

By contrasting the estimated change in the density of the non-target population with a pre-determined acceptable level, this set of hypotheses allows for the fact that many non-target species are able to sustain some reduction in their abundance without any significant risk to their longer term viability (i.e., $r_g < 0$). Specifying a biologically acceptable level of reduction in non-target density will be particularly appropriate for species that realise a longer-term benefit from the reduction in pest density that a poisoning program aims to achieve (e.g., see Powlesland, Knegtmans and Marshall, 1999). For non-target species for which any reduction in density is considered unacceptable, r_g will always be 0. This approach to hypothesis evaluation is called equivalence testing and is specifically focussed on the links between statistical significance and biological significance (Dixon and Garrett, 1994).

A useful measure of changes in the density of non-target species inhabiting poisoned and non-poisoned areas (r_p and r_u) would be their average instantaneous rate of change over the period of the assessment (t). The instantaneous rate of change in non-target density for poisoned and non-poisoned sites (r_i) can be estimated from:

$$r_i = Ln \left[\frac{(N_{ai} - N_{bi})}{t_i} \right] \qquad \text{(Eq. 1)}$$

where N_{bi} and N_{ai} are the density of the non-target population on site i before and after a poisoning program and t_i is the period between N_{bi} and N_{ai}. While r is not the only measure that could be used to

assess the effects of poisoning on non-target species, it is a particularly useful measure because it is centered at 0. This means that incremental declines in non-target abundance will have the same *r* value as equivalent increases, apart from the reversal of sign (Caughley, 1977). In some instances, it may be acceptable to estimate r_u from existing data rather than from the change in density on simultaneous non-poisoned control sites. For example, if there were no reason to expect non-target density to vary independently of the effects of poisoning over the period of between estimates of non-target density on poisoned sites, r_u could be assumed to average 0. However, where such an assumption is dubious, r_u should be estimated from simultaneous control sites.

An experimental result which indicates that the null hypothesis $(r_p - r_u) \geq r_g$ cannot be rejected could arise because (1) differences in the rate of decline in non-target density between poisoned and non-poisoned sites was no less than that specified as the minimum biologically significant rate, or (2) this difference was less than the minimum biologically significant rate but the experiment failed to detect this effect. Experimental failures of this kind are termed Type 2 errors and the probability of them occurring for a given experiment is the Type 2 error rate, β. It follows that the probability of an experiment being able to successfully identify a decline in non-target density related to poisoning is 1 - β, the so-called statistical power of the experiment (Sokal and Rohlf, 1995; p. 159). Type 2 errors will be made in experimental assessments of the effect of poisoning on non-target species density because limits to our capacity to measure wildlife populations mean that estimates of r_p and r_u are subject to sampling error. Replicated measures are used to increase the probability that the effect observed is accurate. The sample size used will largely determine the probability that the assessment will avoid a Type 2 error (the power of the experiment) and correctly identify reductions in non-target species density. Hence, an understanding of how sample size influences experimental power is essential. Similarly, retrospective analysis of experimental power can aid interpretation of experimental results where the null hypothesis (that poisoning has not affected non-target species density) cannot be rejected.

Cohen (1988) gives a comprehensive overview of power analysis and Steidl, Hayes and Schauber (1997) provide a very useful summary of its application to the design and interpretation of wildlife experiments. Most commercially available statistics packages provide basic capabilities for power analysis, while several specialty packages provide facilities to undertake power analysis of more complex experimental designs. Rather than restate this material, we will illustrate how factors which determine experimental power specifically influence aspects of simple designs for assessment of the non-target effects of poisoning programs. The simple designs discussed assume that the potential for temporal and spatial confounding of poisoning effects on non-target species can be avoided through unrestricted selection of sites and random allocation of treatments and controls to these sites. Where this is not possible, more complex designs will be necessary (Underwood, 1993).

Power and sample size

Sample sizes used to estimate the effect of poisoning on non-target density (i.e., those used to estimate r_p and r_u) influence the statistical power of the experiment by affecting the precision with which changes in density can be assessed. For example, the first data point in Fig. 1 shows a 95% confidence interval for an experiment which yielded an average $(r_p - r_u)$ of -0.3 and a between-sample standard deviation of $s = 0.2$ from a sample size of $n = 2$. The sample standard deviation would have been a pooled estimate across both samples used to estimate r_p and r_u when the latter is estimated from control site data, or calculated directly from the sample used to estimate r_p if r_u was assigned some assumed value. If the minimum biologically significant effect of poisoning (r_g) was assigned a value of 0, $(r_p - r_u)$ clearly would not be significantly lower than r_g, and the null hypothesis $(r_p - r_u) \geq r_g$ would not be

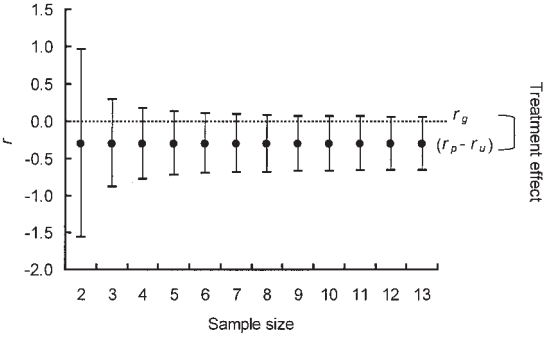

Figure 1: *A sequence of 95% confidence intervals for a hypothetical experiment assessing the effect of poisoning on a non-target species. The average rate of reduction in non-target density over the course of the poisoning program* ($r_p - r_u$) *is -0.3, with a between-sample standard deviation of* $s = 0.2$, *obtained from a range of sample sizes. A nominal value for the rate of reduction in non-target density deemed biologically significant* (r_g) *is shown along with the treatment effect detected by the experiment.*

rejected. The experiment which yielded the first confidence interval shown in Fig. 1 has a statistical power of 0.26, suggesting that the probability of this experiment overlooking a biologically significant reduction in non-target species density is high ($\beta = 0.74$).

Both the size of the confidence interval and its placement relative to r_g contribute to the probability of making a Type 2 error. The size of the 95% confidence interval is dependent on the standard deviation (s) and size (n) of the sample from which ($r_p - r_u$) is derived. The standard deviation and size of the sample contribute to the size of the confidence interval because both influence the probability that ($r_p - r_u$) is a good estimate of the underlying population mean. As such, the precision with which ($r_p - r_u$) is estimated will largely determine the power of an experiment to evaluate the effect of poisoning programs on non-target species density. For example, the sequence of data points in Fig. 1 show how increasing sample size dramatically reduces the width of a confidence interval, despite the other parameters used to calculate the interval remaining constant. Fig. 2 shows how statistical power increases with sample size for the experiments shown in Fig. 1.

The trade-off between power and sample size apparent in Fig. 2 underlies one of the important decisions that must be made when determining the specifications for an experiment to evaluate the effect of poisoning on non-target species density. Lower sample sizes are associated with reduced experimental power leading to a concomitant increase in the probability that a biologically significant reduction in non-target density will be overlooked. As such, the minimum number of

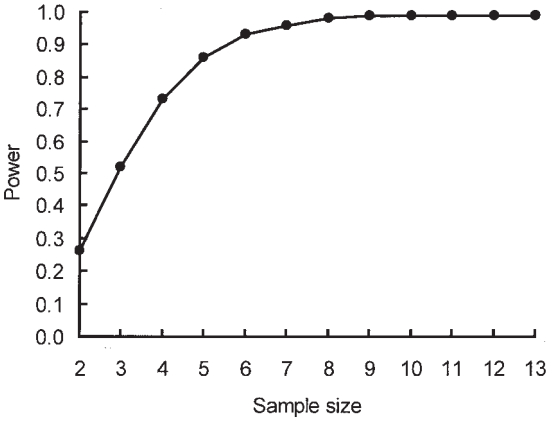

Figure 2: *The effect of increasing sample size shown in Fig. 1 on the power of the experiment.*

poisoning programs that must be sampled in a non-target assessment should be dictated by the minimum level of experimental power considered acceptable. While there are no universal rules dictating acceptable levels of power for such assessments, most biological experiments would be considered of limited value if their power fell below 0.8 (Sokal and Rohlf, 1995). However, given that failures to detect significant reductions in non-target species density can threaten entire populations, high levels of experimental power are desirable.

Non-target species resilience and sample size

The resilience of a non-target species to reductions in its density due to poisoning will determine the appropriate value of r_g to use. The relative resilience of a species will reflect its capacity to compensate for density-independent perturbations in abundance. If poisoning reduces the density of a non-target species, this capacity will determine how long the species takes to recover and whether or not the effect of poisoning will influence the species long-term viability at that location (Sinclair, 1989; Caughley and Sinclair, 1994; Caughley and Gunn, 1997). Because the value of r_g applied in a non-target assessment of poisoning should reflect the perceived resilience of the non-target species, resiliency will have implications for the experimental power of the assessment and hence the sample size that should be used. The difference between the observed rate of reduction in non-target density ($r_p - r_u$) and the acceptable rates (r_g) is termed the treatment effect (Fig. 1). All other things being equal, the power associated with statistical comparisons between ($r_p - r_u$) and r_g declines as the treatment effect decreases. For example, Fig. 3 shows how decreases in the size of a treatment effect progressively diminishes the power achieved by an experiment in which r_g is set at -0.15, $s = 0.2$ and 6 samples are used to estimate ($r_p - r_u$). As treatment effect varies, sample sizes required to maintain acceptably high levels of experimental power will also vary. Because of the interdependent variation in experimental power and the minimum rate of reduction in non-target density deemed biologically significant, their implications for sample size in assessments of the non-target effects of poisoning cannot be considered separately.

Determining sample size for a non-target assessment

Ideally the levels of experimental power and minimum biologically significant change in density should be set with reference to the biology of the

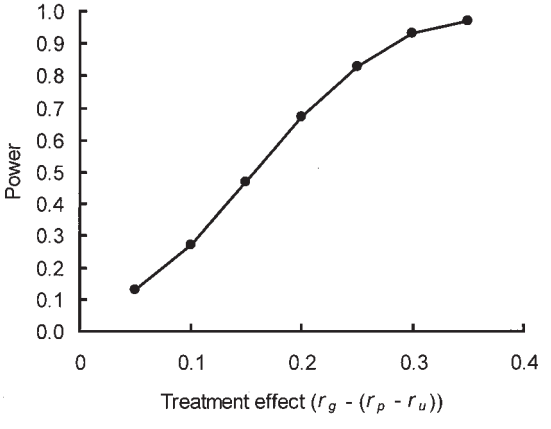

Figure 3: *The relationship between the size of treatment effect for a hypothetical experiment assessing the effect of poisoning on non-target species density and the power of the experiment. Variation in the size of the treatment effect reflects variation in* (r_p - r_u) *with* r_g *set at a constant value of -0.15. The between-sample standard deviation is* s = *0.2 and sample size used to estimate* (r_p - r_u) *is* n = *6.*

non-target species. Only after appropriate values for these parameters have been identified should minimum sample sizes be set. Using appropriate levels for these factors to determine sample size means that when sample size is decreased to fit an available budget, the consequence for the capacity of the experiment to provide useful information on the non-target effects of poisoning programs is made clear.

Both the choice of power and the minimum biologically significant change in density should reflect the perceived value and resilience of the non-target species. Assessments for species of higher intrinsic value or lower resilience should employ higher levels of power and more conservative levels of minimum change in density than assessment for species of lower value or higher resilience. For example, Table 2 associates some indicative rule-of-thumb measures for power and r_g with three broad classes of relative species value and resilience. Relative value will of course be highly dependent on the species assemblage and situation. For example, hunters may place more value on introduced game animals than endangered invertebrates. In contrast, managers interested in ecosystem function may value invertebrates very highly and introduced species not at all. Regardless of questions of value, the level of power for an assessment of the non-target effects of poisoning should not fall below 0.8 and may be set as high as 0.99 for species that are particularly important or valuable.

Resilience in this context reflects the capacity of the non-target species to recover from the mortality caused by poisoning. Species with high intrinsic rates of increase and strong density-dependent links between their demographic rates and factors that regulate their abundance will typically be more resilient to density-independent perturbations than species with lower intrinsic rates of increase and/or highly stochastic population dynamics. The values of r_g given in Table 2 are expressed as a proportion of a species maximum or intrinsic rate of population increase (r_m). For non-target species with progressively lower resilience, the rate of reduction in density that should be considered biologically significant will increase toward 0, constraining the acceptable impact of poisoning to very low level reductions, or to no reduction at all. In contrast, species for which there is clear evidence of a high intrinsic capacity for increase and strong density-dependence in their dynamics should be able to sustain higher levels of reduction due to poisoning without any undue threat to their long-term viability. For these species, Table 2 indicates that reductions equivalent to 0.6 of their annual intrinsic rate of population increase could be overlooked with a reasonable degree of confidence.

Table 2: *Indicative levels of power and* r_g *to apply to assessments of the effect of poisoning on non-target species, classified in terms of the relative value of the non-target population and its resilience to density independent perturbation. Levels of* r_g *are expressed as proportions of the species intrinsic annual rate of population increase* r_m. *Where* r_m *cannot be estimated, resilience should be considered to be low.*

	Resilience					
	High		Moderate		Low	
Intrinsic value	Power	r_g (proportion of r_m)	Power	r_g (proportion of r_m)	Power	r_g (proportion of r_m)
High	0.90	0	0.95	0	0.99	0
Moderate	0.85	0.3	0.90	0.2	0.95	0.1
Low	0.80	0.6	0.85	0.4	0.90	0.2

An example: the effect of rabbit poisoning on kangaroo density

Using the values summarised in Table 2 or some equivalent levels for acceptable power and r_g, and an estimate of the between sample standard deviation derived from experience, other studies or a pilot survey program, the number of samples of (r_p - r_u) necessary to undertake a robust assessment of the non-target effects of poisoning can be identified. For example, in a study of the efficacy of various forms of rabbit (*Oryctalagus cuniculus*) control, Choquenot *et al.* (1998) conducted line-transect spotlight counts of rabbits and kangaroos (*Macropus* spp.) from a vehicle on a series of sites in the central tablelands of New South Wales, Australia. The perpendicular distance of animals observed from the line of travel was measured, allowing their density to be estimated using distance sampling methods (Buckland *et al.*, 1993). One of the rabbit control techniques evaluated was 1080 poisoning, providing an opportunity to experimentally assess the effect this form of control had on kangaroo density. To evaluate the relationship between sample size and experimental power for such an assessment, an estimate of the standard deviation around the instantaneous rate of change in kangaroo density was obtained from a series of sequential estimates of kangaroo density on 7 sites (Table 3, D. Choquenot, *unpubl. data*; Landcare Research, Lincoln, N.Z.). The average instantaneous rate of increase was not significantly different from 0 (t = 1.261, P = 0.254), and indicated a between site standard deviation of s = 0.089.

While large kangaroos are common in the central tablelands of New South Wales, they probably play an important role in structuring plant communities, particularly in reserves. For this reason, they would generally be considered of moderate value. To assess their resilience to poisoning, we have to use maximum annual instantaneous rates of increase (r_m) for other species of large kangaroos (0.4; Bayliss, 1987). Rates of change in large kangaroo abundance appear to be regulated through density dependent variation in food resources (Caughley, 1987), indicating that they have a moderate to high capacity to recover from perturbations in their abundance. Referring to Table 2, moderate species value in combination with moderate to high resilience indicates that an assessment of the effect of 1080 rabbit poisoning on large kangaroo density should employ experimental power of 0.85-0.9, and set the minimum biologically significant rate of reduction in kangaroo density that could be overlooked (r_g) at 0.2-0.3 of the species r_m. Assuming that r_m = 0.4 as above, this corresponds to an r_g of -0.08 and -0.12. Table 4 summarises sample sizes corresponding to this range of values, assuming the between-sample standard deviation calculated above (s = 0.089). The sample sizes given are the number of rabbit poisoning operations that would need to be evaluated to estimate r_p if average r_u = 0. If this assumption was questionable because the interval between pre- and post-poisoning kangaroo surveys was too long, the duration of the poisoning programs was excessive or some non-random change in kangaroo density unrelated to poisoning was suspected, r_u would have to be estimated from the same number of non-poisoned sites. Note that we have to assume s was similar between poisoned and non-poisoned sites.

Retrospective power analysis

Because power analysis allows the two possible interpretations of a non-significant result to be considered in terms of their probability (i.e., the relative probability that (r_p - r_u) is not significantly lower than r_g reflects the fact that (1) (r_p - r_u) was not any lower than r_g, or (2) (r_p - r_u) was lower than r_g but the experiment failed to detect this difference), it can potentially be used to "quality control" the results

Table 3: *Initial and final kangaroo densities, the period over which change in density was assessed and the annual instantaneous rate of change in density for 7 sites in the central tablelands of New South Wales.*

Site	Initial density (kangaroos km^{-2})	Final density (kangaroos km^{-2})	Period (days)	Instantaneous increase (r)
1	20.0	21.0	119	0.150
2	19.8	20.5	90	0.141
3	17.2	17.0	94	-0.045
4	20.1	19.8	111	-0.049
5	18.7	19.1	111	0.070
6	19.1	18.8	121	-0.048
7	17.8	18.2	102	0.080
Average				0.042
Standard deviation				0.089

of experiments that fail to reject the null hypothesis of no biologically significant reduction in non-target species density. However, because the power of a test that rejects a null hypothesis is confounded with the *P*-value upon which the rejection is based (both being dependent on the same sample size, α-level, *s* and treatment effect), retrospective assessment of this result is meaningless (Steidl *et al*., 1997). Valid retrospective power analysis is restricted to evaluation of alternative hypotheses (i.e., untested hypotheses), with treatment effects that differ from that observed from the experimental data; or identification of the minimum effect size that could have been detected at a given level of power. In the current context, this means that power analysis cannot be performed on the null hypothesis $(r_p - r_u) \geq r_g$, although it can be estimated for values of $(r_p - r_u)$ other than that obtained in the experiment. By assessing the power of alternative experimental results, the capacity of an experiment to have detected treatment effects of specific size can be evaluated. Retrospective power analysis can also be used to identify the minimum reduction in non-target density that could have gone undetected by the experiment at given levels of power.

For example, the prospective power analysis summarised in Table 4 indicated that if $r_g = -0.12$ was considered a reasonable estimate of the minimum biologically significant rate of reduction in kangaroo density, and 0.85 an acceptable probability of detecting such an effect, 6 estimates of $(r_p - r_u)$ were required to evaluate the effect of rabbit poisoning on kangaroo density. Because the short duration of rabbit poisoning programs (10-12 days) would make it unlikely that the density of kangaroo populations would vary in a non-random way, r_u was assumed to be 0 and 6 poisoning programs were sampled in order to estimate average $(r_p - r_u)$. The programs used were a subset of a larger number used to assess the efficacy of a range of rabbit control techniques (Choquenot *et al*., 1998). For each poisoning program, kangaroo density was estimated as the average of 3 consecutive surveys during the weeks prior to and immediately after poisoning. Techniques were the same as those used for the preliminary surveys described above. Post-poisoning spotlight line-transect surveys commenced 3 days after poisoned bait had been distributed. Details of the sites, surveys and poisoning programs, and the results of the pre- and post-poisoning surveys and associated instantaneous rates of change in kangaroo abundance are given in Table 5. While kangaroo density declined in 4 of 6 programs, the average rate of change over the period of poisoning was not significantly lower than that specified as the minimum biologically significant rate (Fig. 4).

The prospective power analysis summarised in Table 4 indicated that a sample of 6 poisoning programs would provide sufficient statistical power to have a 0.85 probability of detecting an instantaneous rate of reduction in kangaroo density exceeding -0.12. However, the standard deviation associated with the

Table 4: *Sample sizes required to achieve the extremes of a range of experimental power and values for* r_g *deemed appropriate for an assessment of the effect of rabbit poisoning on large kangaroo density in the central tablelands of New South Wales. The analysis assumes a between-sample standard deviation of* s = *0.089, derived from an independent study of kangaroo populations in the region.*

Power	r_g	Sample size
85	-0.08	11
85	-0.12	6
90	-0.08	13
90	-0.12	7

Table 5: *Details of study sites used to assess the effect of rabbit poisoning on kangaroo density in the central tablelands of New South Wales, Australia, and results of pre- and post-poisoning kangaroo surveys. Poison density is estimated assuming an average of 6 kg of poisoned bait was distributed for every km of bait trail, which was the nominal rate used.*

Site	Area (ha)	Spotlight transect (km)	Bait trail (km)	Poison density (kg ha^{-1})	Pre-poisoning kangaroo density (kangaroos km^2)	Post-poisoning kangaroo density (kangaroos km^2)	Days between surveys	Instantaneous increase (r_p)
1	102	7.7	20	1.18	21.7	19.8	11	-3.040
2	122	5.7	26	1.28	17.1	16.3	12	-1.457
3	116	7.7	24	1.24	14.3	13.0	10	-3.479
4	125	4.9	27	1.30	10.8	11.0	11	0.609
5	220	11.0	45	1.23	20.0	19.1	11	-1.528
6	217	10.2	37	1.02	15.5	16.1	12	1.155
Average				1.21				-1.2
Standard deviation								1.709

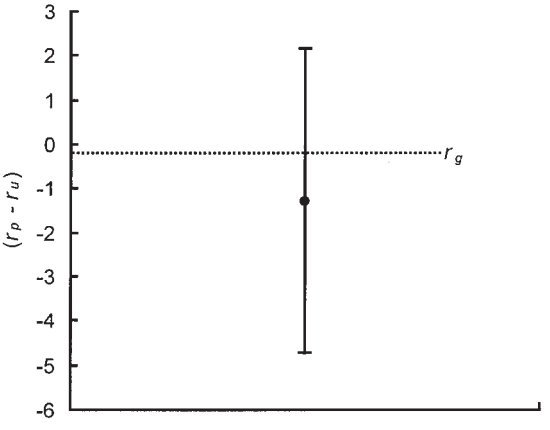

Figure 4: *Average rate of change in kangaroo density and its associated 95% confidence interval for 6 rabbit poisoning programs conducted in the central tablelands of New South Wales. Details of the poisoning programs and results of each kangaroo survey are given in table 5. The value of* r_g *indicated on the Fig. is the minimum rate of reduction in kangaroo density deemed biologically significant (-0.12).*

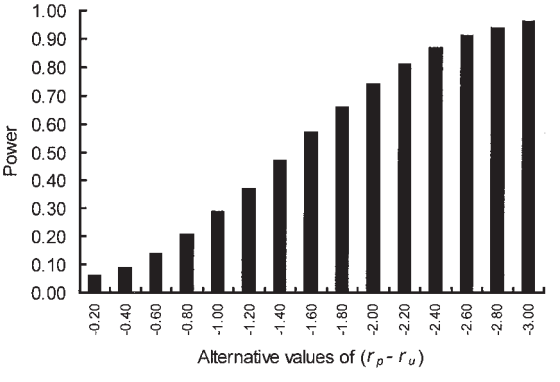

Figure 5: *Power associated with tests of alternative values of* (r_p - r_u) *for the experiment summarised in Fig. 4. Tests are 1-tailed confidence interval comparisons between* (r_p - r_u) *and* r_g *with* s = *1.709 and a sample size of 6.*

sample of $(r_p - r_u)$ obtained in the experiment (s = 1.709) was much higher than that used for the prospective power analysis (s = 0.089). The higher between-sample standard deviation for the experimental assessment was probably due to the much shorter period over which r was estimated in the experimental evaluation (average t = 11 days) compared with the preliminary surveys (average t = 107 days), the lower sample size used and sampling error. To examine the effect of the increased between-sample standard deviation on the experimental assessment, power was estimated for tests of a range of alternative treatment effects (values of $(r_p - r_u)$) lower than that estimated in the experiment (Fig. 5). Levels of power associated with alternative treatment effects did not exceed 0.85 until values of $(r_p - r_u)$ fell below -2.2, a rate of decline almost 20 times lower than the rate of reduction considered biologically significant for large kangaroos. This suggests that there was a >0.9 probability that unacceptably large decreases in kangaroo density went undetected. Based on this result, it would seem prudent to undertake another experiment to assess the effect of rabbit poisoning on kangaroo density, utilising a larger sample size. Using the standard deviation obtained in this experiment and maintaining the minimum treatment effect at -0.12, 17 programs would need to be sampled to reduce the probability of a Type 2 error to 0.15 and 19 programs to reduce this probability to 0.1.

Conclusions

Poisoning operations always seek to minimise non-target mortality because the failure to detect a significant decline in the abundance of some species can have disastrous ecological consequences. Given this, equivalence testing and power analysis should be viewed as integral to the design, analysis and interpretation of experiments to estimate non-target poisoning effects. Prospective power analysis should be used to determine the minimum number of poisoning programs that need to be assessed in order to minimise such errors. It is mandatory that a minimum rate of reduction deemed biologically significant be specified for non-target species. Both this rate and experimental power should be set more conservatively for species with lower intrinsic rates of increase and/or higher stochastic variation in their demographic rates. Prospective power analysis requires an estimate of the standard deviation associated with the measure used to assess change in non-target species density. This estimate may be available from existing data or may require a pilot study. If the existing data used to estimate the relevant standard deviation were collected over a substantially different spatial or temporal scale, or under different conditions than those likely to be encountered (i.e., different survey techniques, population densities, environmental conditions, times of the year or duration of sampling periods), a pilot study is desirable. If the null hypothesis of "no significant reduction in non-target species density" cannot be rejected, retrospective power analysis and examination of confidence intervals can be used to determine the probability that the experiment may

have overlooked biologically significant reductions in non-target density. Both approaches can also be used to "quality control" non-significant results of such experiments.

Acknowledgements

We would like to sincerely thank Gábor Lövei, Ian Henderson, and an anonymous referee for significantly improving an earlier draft, and Clare Veltman for her patience while these improvements were undertaken.

References

Bayliss, P. 1987. Kangaroo dynamics. In: Caughley, G.; Shepherd, N.; Short, J. (Editors), *Kangaroos: their ecology and management in the sheep rangelands of Australia*, pp. 119-134. Cambridge University Press, Cambridge, U.K.

Buckland, S.T.; Anderson, D.R.; Burnham, K.P.; Laake, J.L. 1993. *Distance sampling: estimating abundance of biological populations*. Chapman and Hall, London.

Caughley, G. 1977. *Analysis of vertebrate populations*. John Wiley and Sons, New York, U.S.A. 234 pp.

Caughley, G. 1987. Ecological relationships. *In:* Caughley, G.; Shepherd, N.; Short, J. (Editors), *Kangaroos: their ecology and management in the sheep rangelands of Australia*, pp. 159-187. Cambridge University Press, Cambridge, U.K.

Caughley, G.; Sinclair, A.R.E. 1994. *Wildlife ecology and management*. Blackwell Scientific Publications, Oxford. 334 pp.

Caughley, G.; Gunn, A. 1997. *Conservation biology in theory and practice*. Blackwell Scientific Publications, Oxford.

Choquenot, D.; Druhan, J.; Lukins, B.; Packwood, R.; Saunders, G. 1998. Managing the impact of rabbits on wool production systems in the central tablelands of New South Wales: an experimental study and bioeconomic analysis. *Proceedings of the 11th Australian vertebrate pest conference*, pp. 367-374. Agriculture Western Australia, Forrestfield, Western Australia.

Cohen, J. 1988. *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum Assoc., Hillsdale, New Jersey, U.S.A.

Dixon, P.M.; Garrett, K.A. 1994. Statistical issues for field experimenters. *In:* Kendall, R.J.; Lacher Jr., T.E. (Editors), *Wildlide toxicology and population modeling: integrated studies of agroecosystems*, pp. 439-450. CRC Press, Boca Raton, Florida, U.S.A.

Livingstone, P.G. 1994. The use of 1080 in New Zealand. *In:* Seawright, A.A.; Eason, C.T. (Editors), *The Royal Society of New Zealand, Miscellaneous series 28:* 1-9.

McIlroy, J.C. 1985. Observations on the sensitivity of some Australian birds and the feral pig to the organophosphorus insecticide, fenthion ethyl. *Australian Wildlife Research 12:* 331-335.

Nakata, K.; Satoh, K. 1994. Distance of movements of voles after removal in *Clethrionomys rufocanus bedfordiae* (Rodentia: Cricetidae). *Applied Entomology and Zoology 29:* 97-103.

Powlesland, R.G.; Knegtmans, J.W.; Marshall, I.S.J. 1999. Costs and benefits of aerial 1080 possum control operations using carrot baits to North Island robins (*Petroica australis longipes*), Pureora Forest Park. *New Zealand Journal of Ecology 23:* 149-159.

Schieck, J.O.; Millar, J.S. 1987. Can removal areas be used to assess dispersal of red-backed voles? *Canadian Journal of Zoology 65:* 2575-2578.

Sinclair, A.R.E. 1989. Population regulation in animals. *In:* Cherret, J.M (Editor), *Ecological concepts: the contribution of ecology to an understanding of the natural world*, pp. 197-241. Blackwell Scientific Publications, Oxford.

Sokal, R.R.; Rohlf, F.J. 1995. *Biometry. 3rd ed.* Freeman and Co., San Francisco, California, U.S.A.

Steidl, R.J.; Hayes, J.P.; Schauber, E. 1997. Statistical power analysis in wildlife research. *Journal of Wildlife Management 61:* 270-279.

Twigg, L.E.; King, D.R. 1989. Tolerance to sodium fluoroacetate in some Australian birds. *Australian Wildlife Research 16:* 49-62.

Underwood, A.J. 1993. The mechanics of spatially replicated sampling programmes to detect environmental impacts in a variable world. *Australian Journal of Ecology 18:* 99-116.