Population monitoring programme for Archey's frog (*Leiopelma archeyi*): pilot studies, monitoring design and data analysis

Amanda Haigh, Shirley Pledger and Sebastian (Avi) Holzapfel

DOC RESEARCH & DEVELOPMENT SERIES 278

Published by
Science & Technical Publishing
Department of Conservation
PO Box 10420, The Terrace
Wellington 6143, New Zealand

DOC Research & Development Series is a published record of scientific research carried out, or advice given, by Department of Conservation staff or external contractors funded by DOC. It comprises reports and short communications that are peer-reviewed.
Individual contributions to the series are first released on the departmental website in pdf form. Hardcopy is printed, bound, and distributed at regular intervals. Titles are also listed in our catalogue on the website, refer www.doc.govt.nz under *Publications*, then *Science & technical*.

ISSN 1176-8886 (hardcopy)
ISSN 1177-9306 (web PDF)
ISBN 978-0-478-14281-5 (hardcopy)
ISBN 978-0-478-14282-2 (web PDF)

This is a client report commissioned by Waikato Conservancy and funded from the Science Advice Fund. It was prepared for publication by Science & Technical Publishing; editing by Ian MacKenzie and layout by Sue Hallas. Publication was approved by the Chief Scientist (Research, Development & Improvement Division), Department of Conservation, Wellington, New Zealand.

In the interest of forest conservation, we support paperless electronic publishing. When printing, recycled paper is used wherever possible.

CONTENTS

Population monitoring programme for Archey's frog (*Leiopelma archeyi*): pilot studies, monitoring design and data analysis

Amanda Haigh[1], Shirley Pledger[2] and Sebastian (Avi) Holzapfel[3]

[1] Waikato Conservancy, Department of Conservation, Private Bag 3072, Hamilton 3204, New Zealand. Email: asmale@doc.govt.nz

[2] School of Mathematics, Statistics and Computer Science, Victoria University of Wellington, Wellington 6140, New Zealand

[3] Northern Regional Office, Department of Conservation, PO Box 112, Hamilton 3240, New Zealand

ABSTRACT

Archey's frog (*Leiopelma archeyi*) is an endangered terrestrial species that occurs in two geographically distinct populations within the Waikato Region, New Zealand. There is concern about a possible decline in abundance in the Whareorino population because of predation and disease threats. However, knowledge of current trends in abundance within this population is limited. Since variable detection rates make abundance estimates from simple counts unreliable, a capture-recapture monitoring programme is needed. This report gives recommendations for the design and analysis of the monitoring programme. Capture-recapture pilot studies were carried out on five separate occasions in 2004–2005, with 2, 3 or 4 nights of sampling per trip, and nightly weather and search effort variables were recorded. Individual frogs were identified by their unique natural markings using a single digital photograph of the frog on a stage surrounded by mirrors. Capture-recapture analysis of these data gave preliminary estimates of abundance and capture probabilities. Using this information, power analyses were completed and the power to detect different types of abundance trends over time was tabled and graphed. The information presented will allow recommendations to be made about the number and size of grids and the number of nights of sampling per trip that will be required to detect a specified drop in abundance with confidence.

Keywords: Archey's frog, *Leiopelma archeyi*, amphibian, population monitoring, abundance, capture-recapture, power analysis, photographic identification, chytrid fungus, New Zealand

Haigh, A.; Pledger, S.; Holzapfel, A. 2007: Population monitoring programme for Archey's frog (*Leiopelma archeyi*): pilot studies, monitoring design and data analysis. *DOC Research & Development Series 278*. Department of Conservation, Wellington. 25 p.

1. Introduction

1.1 ARCHEY'S FROG

Archey's frog (*Leiopelma archeyi*) is one of three surviving terrestrial *Leiopelma* species in New Zealand and is currently ranked as 'Nationally Critical', the highest threat classification in New Zealand (Hitchmough et al. 2007).

The Whareorino Forest population was discovered in 1991 (Thurley & Bell 1994; Thurley 1996). This population extends across a 600-ha area of cloud forest in the Herangi Range, which ranges in altitude from approximately 400 m to 800 m above sea level. The current geographic extent of this population has been defined by various surveys conducted since its discovery (Thurley & Bell 1994; Thurley 1996; Sutton 2005).

There is ongoing concern about the stability of this (and other) Archey's frog populations, largely because of threats posed by disease, such as amphibian chytrid fungus (*Batrachochytrium dendrobatidis*), and introduced mammalian predators. Our current lack of knowledge of population trends means a more intensive monitoring programme with high sensitivity to detect changes in these trends is needed.

Several techniques have been employed to monitor the Whareorino population. Methods have included random daytime transect searches, trials with artificial cover objects and more recently night emergence counts (Thurley & Bell 1994; Thurley 1996; Eggers 1998; Thorsen 1998, 1999; Thurley 2000; Wakeland et al. 2003; N. Webster, Waikato Conservancy, Department of Conservation (DOC), pers. comm. 2003). Monitoring Archey's frog (and other *Leiopelma* species) using random transects at most provides information on frog presence/absence and, possibly, gross changes in abundance, due to the high variation in the detectability of this species. Recent studies by N. Webster confirmed that counting emerged frogs at night was an inappropriate monitoring technique, again owing to large variation in detectability (N. Webster, DOC, pers. comm. 2003). Consequently, this programme is focused on using capture-recapture methods to monitor this frog population and the effectiveness of current conservation management.

1.2 CURRENT MANAGEMENT

An intensive ground-based rodent control programme was established at Whareorino Forest in August of 2003 following a spate of rat predation events on Archey's and Hochstetter's frogs (*Leioplema archeyi* and *L. hochstetteri*, respectively) (Fitzgerald & Campbell 2003; Thurley 2003). Rodent control was established over the northern 300-ha block of Archey's frog habitat; this block is hereafter referred to as the 'treatment block'. The southern block of frog habitat has received no rodent control and is referred to as the ‘non-treatment block’. Frog habitat was divided into the treatment and non-treatment blocks for the following reasons:

- To limit the risks posed by rodent control to only half of the Archey’s frog population in case rodent control has a negative effect on the population.
- To enable assessment of the effectiveness of the rodent control programme as a management tool for protecting Archey’s frogs.

The rodent control programme was set up as an emergency response to rat predation; consequently, there was no opportunity to monitor the frog population prior to implementing rodent control. This programme is an ongoing operation and is funded, along with the Archey’s frog monitoring programme, until July 2009.

Strict hygiene protocols are implemented by staff working in Archey’s frog habitat at Whareorino Forest to minimise the introduction and/or spread of chytrid fungus and other unidentified amphibian diseases.

1.3 PROJECT AIM

To establish and implement a capture-recapture monitoring programme for the Whareorino Forest Archey’s frog population that has high sensitivity to detect small changes in population trends.

1.4 PROJECT OBJECTIVES

1. To monitor long-term trends in population size, survival and recruitment to determine whether the population is stable, increasing or in decline.
2. To detect a population decline of 20% or greater within 2–8 months of a decline occurring.
3. To test whether the rodent control programme is influencing trends in population abundance, survival and recruitment.
4. To document and model environmental variables that influence frog emergence and population trends in abundance, survival and recruitment.
5. To investigate the population ecology of Archey’s frog, including movement behaviour (home range, territory) and habitat use. (This objective is not addressed in this report.)

1.5 HYPOTHESIS TEST (OBJECTIVE 3 ONLY)

Null hypothesis: That there will be no difference in population abundance, survival rates or recruitment rates between sampling units in the treatment and non-treatment blocks from 2005 to 2009.

Alternative hypothesis: That trends in population abundance, survival rates, and recruitment rates will be significantly different between sampling units in the treatment block and in the non-treatment block from 2005 to 2009.

'Population' is defined as the frog population within the sampling units (i.e. 10 m × 10 m grids) used in this monitoring programme. There are two reasons for adopting this definition. First, owing to the significant effort required to conduct capture-recapture monitoring at each 10 m × 10 m grid, only four sampling units were recommended, thus limiting the number of replicates per treatment to two. Second, the sampling units were to be selected from sites of high abundance. These sites are unlikely to provide a generic representation of frog abundance throughout Whareorino Forest because the relative proportion of high-, medium- and low-density sites is unknown (see section 6). Given that this sampling is not completely random, and the small number of sampling units, it is not statistically appropriate to extrapolate trends in abundance beyond these grids.

1.6 PILOT STUDY

We conducted pilot studies between January 2004 and March 2005, trialling two sampling unit (i.e. grid) sizes for capture-recapture monitoring. We used data from these studies to obtain preliminary estimates of capture-recapture parameters, from which we produced power estimates to aid in the design of the long-term study. The recommendations for monitoring design and data analysis are given in section 6.

2. Pilot study field methods

Two pilot study grids (one $5\,m \times 7\,m$, one $10\,m \times 10\,m$) were established in representative frog habitat, which supported high densities of frogs, in Whareorino Forest (for a detailed habitat description see Thurley 1996). The $10\,m \times 10\,m$ grid was created as an extension of the smaller $5\,m \times 7\,m$ grid (established first) to enable comparison of the quality of the data collected from different grid sizes. Hence the $5\,m \times 7\,m$ grid was embedded in the $10\,m \times 10\,m$ grid.

Each grid was divided into lanes 2-m wide and searched at night by torchlight for emerged frogs. Searching commenced 1 hour after sunset and one complete search of the grid(s) was made each night. Grids were searched for up to 3 consecutive nights, depending on the length of each field session. Table 1 details the number of sessions, dates of when those sessions were completed between January 2004 and March 2005, and grid size. Weather conditions (relative humidity, air temperature, cloud cover, rainfall, maximum wind speed, average wind speed, weather over the previous 24 hours) were recorded at the start and finish of searching for the sessions starting November 2004.

Emerged frogs were captured and held individually in re-sealable plastic bags for processing and then returned to their point of capture. A new disposable (nitrile) glove was used for handling each frog to minimise the spread of amphibian disease should any be present at the site. Time of capture, capture location (to the nearest $1\,m^2$), habitat, height above ground, and age class were recorded on capture, and search effort was recorded by all observers.

Captured frogs were photographed with the aid of a photographic mirror stage, that enables four views (lateral left, lateral right, dorsal and frontal) of the frog to be captured in one digital photograph (Wallace 2004; Smale et al. 2005). Frogs were later identified from these photographs for capture-recapture analysis following the methods developed by Bradfield (2004). Individual frog recognition was achieved using their unique patterned skin markings.

TABLE 1. DATE, NUMBER AND LENGTH OF MONITORING SESSIONS COMPLETED PER GRID.

DATE	NO. OF CONSECUTIVE NIGHTS	GRID SIZE
January 2004	2	$5\,m \times 7\,m$
February 2004	4	$5\,m \times 7\,m$
November 2004	3	$5\,m \times 7\,m$*
November 2004	3	$10\,m \times 10\,m$*
February 2005	3	$10\,m \times 10\,m$
March 2005	3	$10\,m \times 10\,m$

* Data collected concurrently for grid size comparison.

3. Capture-recapture methodology

Capture-recapture analysis was conducted on two datasets from pilot studies, using the free software package MARK (White 2005) and programs using the 'R' statistical package (R Development Core Team 2005).

These studies used a robust design (Pollock 1982), each with three primary periods or sessions (the three different months), within each of which there were repeated nights of sampling (the secondary periods). Analysis of the robust design (Kendall et al. 1995) assumes the population is closed to birth, recruitment, death and migration within each secondary period, but allows for these processes to occur between primary periods. Further assumptions are that the animals are captured independently, and all are correctly identified. If there are at least four sampling nights within a session, the robust design analysis is able to account for heterogeneity of capture, in which some animals are more likely to be captured than others. This avoids the underestimation of population size that occurs when heterogeneity is present but not allowed for in the model. *Leiopelma* frogs are likely to have heterogeneity of capture, as well as possible temporary emigration underground (rather than sideways off the grid). At any session, some individuals may be completely unavailable for capture over all the nights of that session. With the robust design, temporary emigration may be estimated for all sessions except the first and last (Kendall & Nichols 1995). Robust design analysis is available on the MARK software package, but it has many parameters, and requires a large dataset for successful model fitting.

For quick monitoring in the short term, two simpler methods are available. To compare abundance between the last two sessions, closed-population models (Otis et al. 1978) may be used. Provided the models are based on statistical likelihood (Pledger 2000), comparisons may be done by model selection (e.g. comparing models with either constant or a declining population) using Akaike's Information Criterion (AIC). The AIC measures lack of fit of the model to the data. The model with the lowest AIC is selected, although differences of less than about 3 or 4 indicate there is no clear choice between two models (Burnham & Anderson 2002). The closed-population models may allow for various influences on capture probability—for example, variation through time, heterogeneity, trap-shyness, weather covariates or search effort, but not temporary emigration. It is important to identify these sources of variations in capture probability as they bias the abundance estimates if they are present but not allowed for in the models. To monitor abundance over three or more recent sessions, a likelihood version (Schwarz & Arnason 1996) of the Jolly-Seber model (Jolly 1965; Seber 1965) is useful. AIC values may be used to compare models with constant abundance with models where abundance is declining between sessions. The Jolly-Seber model cannot allow for heterogeneity or temporary emigration.

4. Pilot study results

Dataset 1 was from the 5 m × 7 m grid over 9 nights of sampling in 2004: in January (2 nights), February (4 nights) and November (3 nights). Fifty-five different frogs were caught. Dataset 2 was from the 10 m × 10 m grid over 9 nights, 3 each in November 2004, and February and March 2005, in which 116 different frogs were captured.

The robust design analysis failed with Dataset 1, because of too few captures and recaptures, but it was successful with Dataset 2 (Table 2). In all models tried, survival probability ϕ depended on time (denoted by ϕ(t)), with capture probability p for each individual being modelled as constant (Model {ϕ(t), p(.)}), varying through time from night to night (Model {ϕ(t), p(t)}), or having a behavioural effect, e.g. trap-shyness (Model {ϕ(t), p(b)}). It was not possible to check for heterogeneity of capture probability among individuals (Model {ϕ(t), p(h)}), as this requires at least 4 nights of sampling within a primary period. Covariates for p were also tried, with models having p dependent on search time, time on the site or various weather covariates (temperature, humidity, cloud cover, rainfall or wind speed at the start or finish of sampling, and previous rainfall).

The result of the robust design analysis was to select Model {ϕ(t), p(t)}, with time variation in p (Table 2). The second best model, with AIC 3.2 higher

TABLE 2. RELATIVE AIC VALUES FOR THE ROBUST MODELS WITH DATASET 2. THE SURVIVAL PROBABILITY ϕ VARIED THROUGH TIME, AND DIFFERENT MODELS WERE TRIED FOR NIGHTLY CAPTURE PROBABILITY p. RELATIVE AIC = (AIC FOR CURRENT MODEL) – (AIC FOR BEST MODEL). THE BEST MODEL HAS AIC = 0.

MODEL TYPE	MODEL	DESCRIPTION	RELATIVE AIC
No covariates for p	M(0)	Constant p	32.0
	M(t)	Nightly variation of p	0.0
	M(b)	Trap-shyness	11.4
Search effort covariates for p	M(site.time)	Person-hours at site	3.2
	M(search.time)	Person-hours searching	12.2
Weather covariates for p	M(avwS)	Average wind speed at start	28.8
	M(avwF)	Average wind speed at finish	34.0
	M(ccS)	Percentage cloud cover at start	17.5
	M(ccF)	Percentage cloud cover at finish	34.0
	M(humS)	Relative humidity at start	15.4
	M(humF)	Relative humidity at finish	33.9
	M(maxwS)	Maximum wind speed at start	30.2
	M(maxwF)	Maximum wind speed at finish	34.0
	M(prev)	Weather in previous 24 hours	27.8
	M(rainS)	Rain at start	32.6
	M(rainF)	Rain at finish	32.6
	M(tempS)	Temperature at start	24.4
	M(tempF)	Temperature at finish	20.4

than the best model, had p dependent on site time (the number of person-hours spent on the site that night, by all searchers). This covariate really only reflects the fact that when more frogs are available for capture, a longer time must be spent on site; we could probably not increase capture probabilities by staying longer. All other models had AIC at least 11.4 higher than the chosen model $\{\phi(t), p(t)\}$. This includes the weather-variable models. The best weather variable for modelling p was humidity at the start of sampling, but it had AIC 15.4 above that of the p(t) model. The weather variables provided some information about capture probabilities, but not a complete explanation. The night-to-night variation in capture probability is consistent with other *Leiopelma* studies (e.g. Pledger 1999; Bell et al. 2004; N. Webster, DOC, pers. comm. 2003). The choice of p(t) supports using the simpler Jolly-Seber model, by pooling information within each session (primary period) to score each frog as either caught (at least once) or not caught in that session. This ignores the fine detail of capture patterns within a session, but is helpful for providing more precise estimates of population size and survival probabilities.

Parameter estimates from closed-population and robust design analyses of the 10 m × 10 m grid are in Table 3. The average nightly capture probability for a frog on the surface and available for capture (i.e. not in temporary emigration) is 0.31, with a standard deviation of 0.11. Note the particularly high capture probability of 0.67 one February night. The closed-population

TABLE 3. PARAMETER ESTIMATES FROM DATASET 2 USING THE ROBUST DESIGN.

PARAMETER TYPE	SESSION*	PARAMETER	ESTIMATE	STANDARD ERROR
Per night				
Nightly capture	1	p_{11}	0.28	0.07
probability p for	1	p_{12}	0.35	0.08
frogs available	1	p_{13}	0.30	0.07
for capture	2	p_{21}	0.16	0.04
(not temporarily	2	p_{22}	0.67	0.07
emigrated	2	p_{23}	0.41	0.06
underground)	3	p_{31}	0.24	0.06
	3	p_{32}	0.29	0.06
	3	p_{33}	0.23	0.06
Per session				
Abundance N	1	N_1	63.5	9.7
(available frogs)	2	N_2	79.6	6.0
	3	N_3	103.7	16.5
Proportion of all	1	γ_1	Not estimable	
frogs in temporary	2	γ_2	0.26	0.15
emigration, γ	3	γ_3	Not estimable	
Between sessions				
Survival probability	1 to 2	ϕ_1	0.96	0.14
	2 to 3	ϕ_2	0.61	0.13

* 1 = November 2004; 2 = February 2005; 3 = March 2005.

abundance estimates of the available frogs are in Table 3, but using the robust design adjusts the N_2 estimate up from 79.6 to 104.7 after temporary emigration has been estimated (26% of all frogs). The robust design cannot estimate temporary emigration at the first and last sessions, so the N_1 and N_3 values in Table 2 are likely to be underestimates.

We need a closed-population session of at least 4 nights to test for heterogeneity. The February session in Dataset 1 had 4 nights, but there were not enough captures for a clear selection of the 'best' model. Combining the February and March sessions from Dataset 2 as if they were from a closed population gave a model choice including both time and heterogeneity as factors for p (Model M(t+h) of Pledger 2000). No behavioural effect (e.g. trap-shyness) was detected. However, there is some doubt about this model selection, as a null hypothesis of a closed population over those 2 months was rejected (closure test, $\chi^2 = 30.835$, df = 2, $P < 0.0001$).

5. Power analysis

The preliminary results from the pilot studies enable us to assess the power to detect a decline in the population under different study designs. In this report, power, the probability of detecting the decline, is labelled 'very high' if above 0.9, 'high' if between 0.8 and 0.9, 'medium' if between 0.7 and 0.8, and 'low' if below 0.7. Calculation of power for the χ^2 tests uses the method given in Lebreton et al. (1992). Each scenario has a certain combination of number of grids, number of nights of sampling per session, size and type of capture probabilities, presence or absence of heterogeneity and temporary emigration, significance level (α) to be used for the statistical test, initial population and amount of decline of population. For each scenario, we found the power to detect the decline, using four different statistical tests. Each test may be a short-term analysis based on only the last two sessions, or a long-term analysis based on several sessions. The test may aim at detecting a decline over all the study grids, or a decline only on either the treatment or non-treatment grids.

5.1 SCENARIOS TESTED

Suppose, for example, there are four grids being monitored, grids 1 and 2 in the non-treatment area, and grids 3 and 4 in the treatment area. The four tests are outlined below.

A *short-term analysis* to detect an *overall population decline*. Data from the two most recent sessions are analysed as a set of closed populations. Suppose the population sizes are N_1, N_2, N_3 and N_4 at the former session, and M_1, M_2, M_3 and M_4 at the latter session. The null hypothesis of no change (each $M_i = N_i$) is tested against an alternative of changed population values (the M_i values differ from N_i, and on average are lower). Appropriate closed-population models are fitted for each site at each session. Likelihood-based models allow us to take a joint likelihood over all sites, giving an overall test with more power to detect population changes than a test of a single site. This is a χ^2 likelihood ratio test on 4 degrees of freedom.

A *short-term analysis* to detect a *population decline in non-treatment sites only*. This has the same null hypothesis as the test in 1, but the alternative is that the M_i values at the non-treatment sites are lower than N_i, while there is no change at the treatment sites. This χ^2 test is on 2 degrees of freedom.

A *long-term analysis* to detect an *overall population decline*. This is based on the last few sessions. The likelihood-based version of the open-population Jolly-Seber model given by Schwarz & Arnason (1996) enables us to test at any site a null hypothesis of a constant population against an alternative of a declining population (Schwarz 2001). Over four sites, we may use joint likelihoods to construct the appropriate χ^2 test of a null hypothesis of no change at any site, versus an alternative hypothesis of a decline at each site.

A *long-term analysis* to detect a *population decline in non-treatment sites only*. This has the same null hypothesis as the test in 3, but the alternative is a decline at only the non-treatment sites.

5.2 TEST RESULTS

Tests 1–4 are used, together with estimates from the pilot studies, to find the power curves and hence decide the numbers of sites and nights of sampling needed to detect changes and trends in the population size. Figures 1–4 give detailed results of tests 1–4, respectively. As expected, they show that higher power is obtained with more grids, higher initial population size, higher capture probabilities, more sampling nights per session, higher pre-set significance levels and larger percentage drops in population size N.

5.2.1 Test one

For *short-term detection* of an *overall decline* (Test 1), Fig. 1 shows power curves including a basic scenario of four sites (two non-treatment, two treatment), an initial population $N_1 = 100$, an average nightly capture probability $p = 0.3$, 4 nights per session and testing the hypothesis at a 20% significance level ($\alpha = 0.2$). The values $N_1 = 100$ and $p = 0.3$ were suggested by the pilot study for a 10 m × 10 m grid. This basic scenario appears in all four plots in Fig. 1. There is an acceptably high power to detect an overall drop in N—for example, a drop of 20% is detected with very high power (probability 0.943). Figure 1A shows the effect of varying the number of nights of sampling. Searching for 4 nights per session gives much better power than 3 nights. Figure 1B shows the effect of varying the initial number of frogs on the grid, which could be achieved by changing the grid size. The 10 m × 10 m grid in good frog habitat has N about 100, with very high power. The 5 m × 7 m grid has N about 60, which still has high power (probability

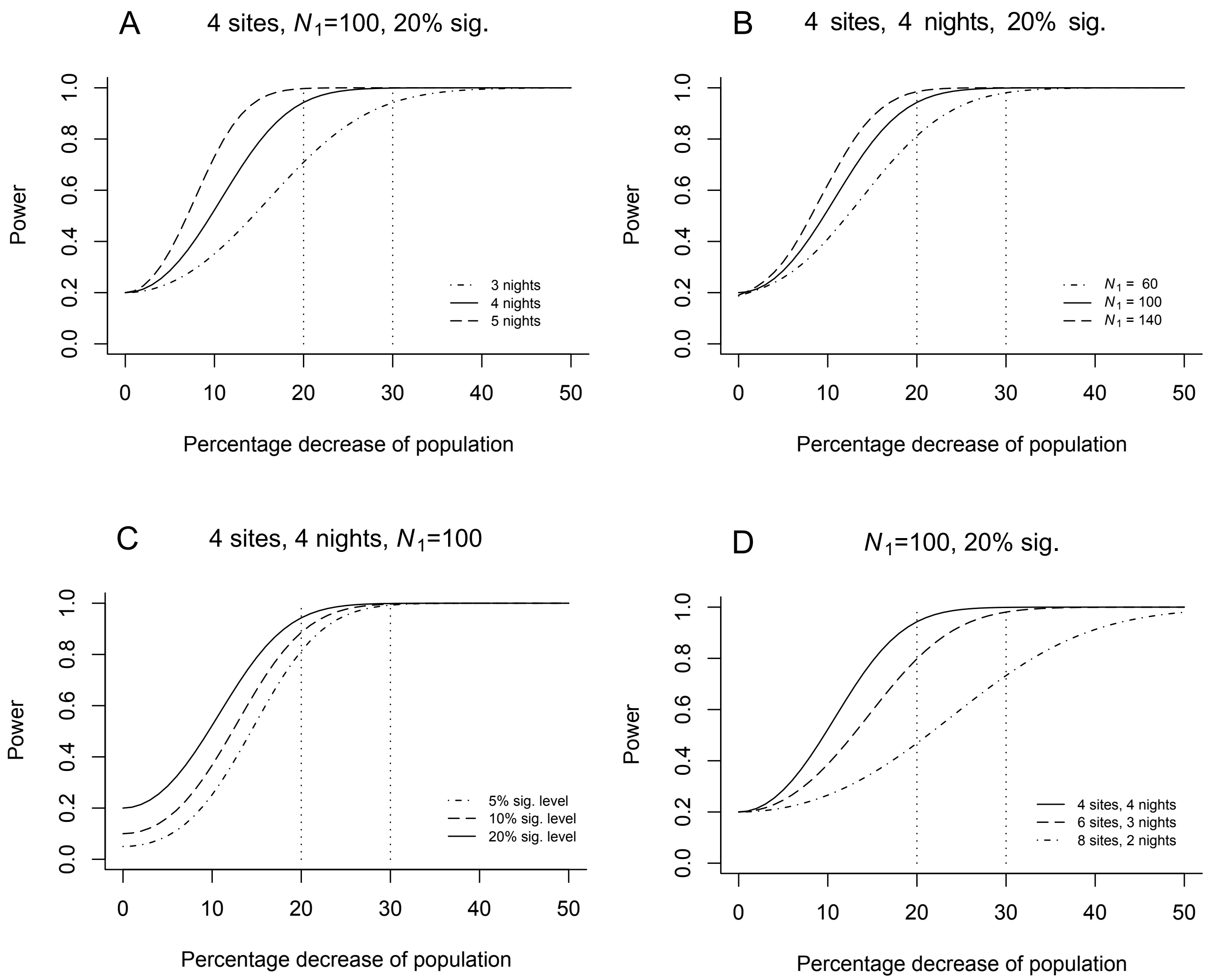

Figure 1. Power to detect a decrease in population size N at all sites, using the last two sessions. The basic scenario has four sites, initial abundance $N_1 = 100$, average nightly capture probability $p = 0.3$, $k = 4$ nights of sampling per session and a 20% significance level for the testing. Plots A, B and C show the effect of varying k, N_1 and significance level, respectively. Plot D shows the effect of increasing the number of sites while compensating by decreasing k to keep on-site search time approximately constant. The basic scenario appears as a solid line in all four plots, and the vertical lines at 20% and 30% drops are for reference.

0.815 for a 20% drop). However, with $N = 60$ there is a risk of having too few captures to be able to fit the capture-recapture models to the data. Figure 1C shows the effect of using different significance levels in the hypothesis test. Where an early warning of population decrease is needed, setting α high is a safer policy, even at the (say) 20% risk of a false positive (Skalski & Robson, 1992). In Fig. 1D, two alternative designs are considered. Bearing in mind the economic constraints on sampling effort, three options with similar on-site effort are explored. The first is four sites with 4 nights per session (the basic scenario), the second is six sites (three non-treatment, three treatment) with 3 nights per session, and the third is eight sites (four non-treatment, four treatment) with 2 nights per session. There was considerable loss of power with the second and third options, caused by sessions having fewer different animals caught, and fewer recaptures. The second and third options would also be more demanding logistically, involving more travel and set-up costs.

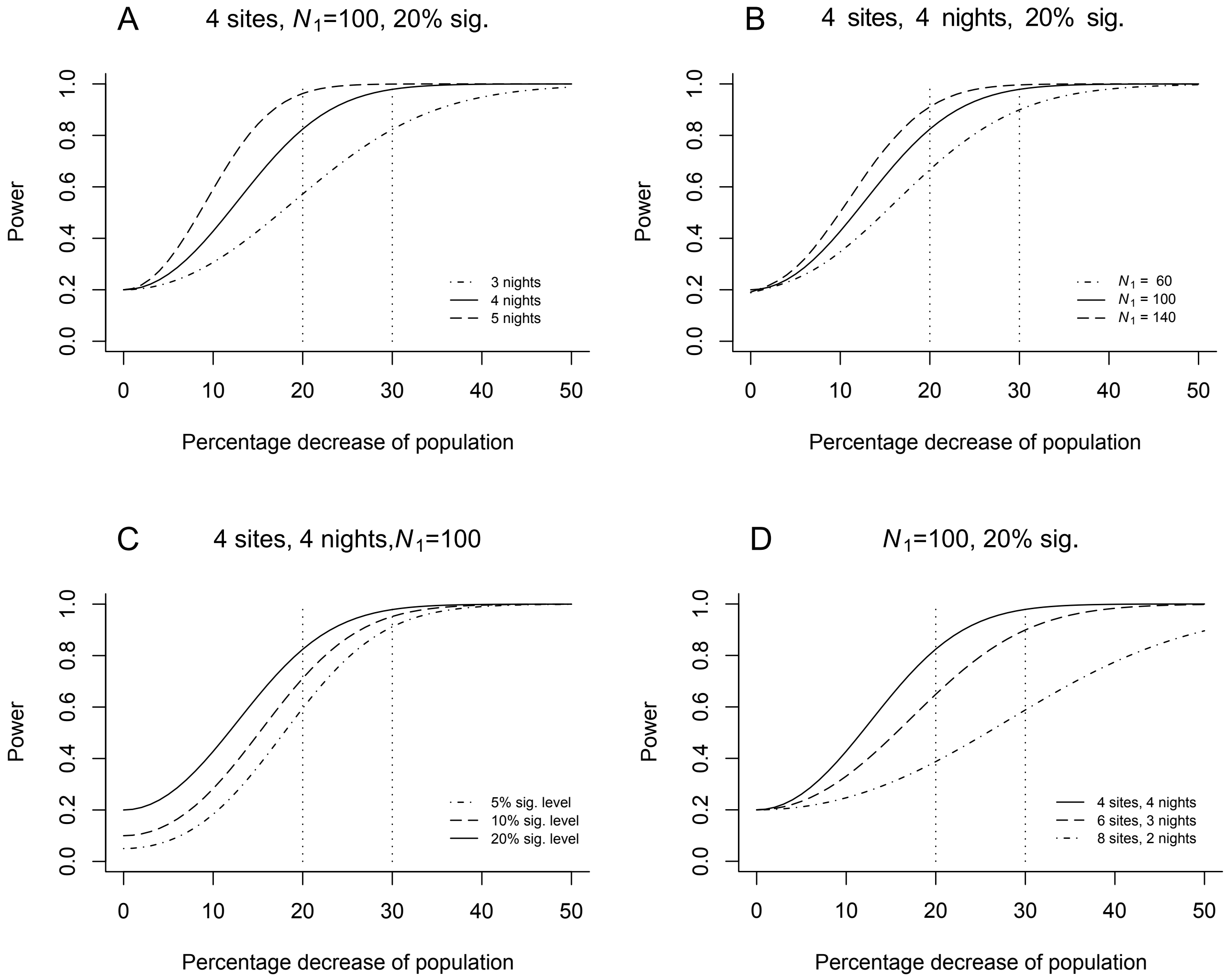

Figure 2. Power to detect a drop in population size N occurring at the non-treatment sites only, using the last two sessions. The basic scenario has four sites, initial abundance $N_1 = 100$, average nightly capture probability $p = 0.3$, $k = 4$ nights of sampling per session and a 20% significance level for the testing. Plots A, B and C show the effect of varying k, N_1 and significance level, respectively. Plot D shows the effect of increasing the number of sites while compensating by decreasing k to keep on-site search time approximately constant. The basic scenario appears as a solid line in all four plots, and the vertical lines at 20% and 30% drops are for reference.

5.2.2 Test two

For *short-term detection* of a *decline in non-treatment grids only* (Test 2), Figure 2 shows power curves for the same basic scenario and variations. The power to detect a 20% drop in population with the basic scenario is now 0.824—lower than when the drop is over all four sites, but still acceptably high. However, if smaller grids are used with $N = 60$, the power to detect a 20% decrease is low (0.668). The power curves in Fig. 2 show the same patterns as those in Fig. 1, but are generally lower.

The power curves did not take account of possible heterogeneity of capture, for which we do not yet have a good estimate. With heterogeneity, N is underestimated. If a similar amount of heterogeneity is present at both sessions, the change in N is detected with slightly reduced power, even if using Model M(t) (not allowing for heterogeneity). This is because we are comparing an initial and final population, and the underestimates tend to

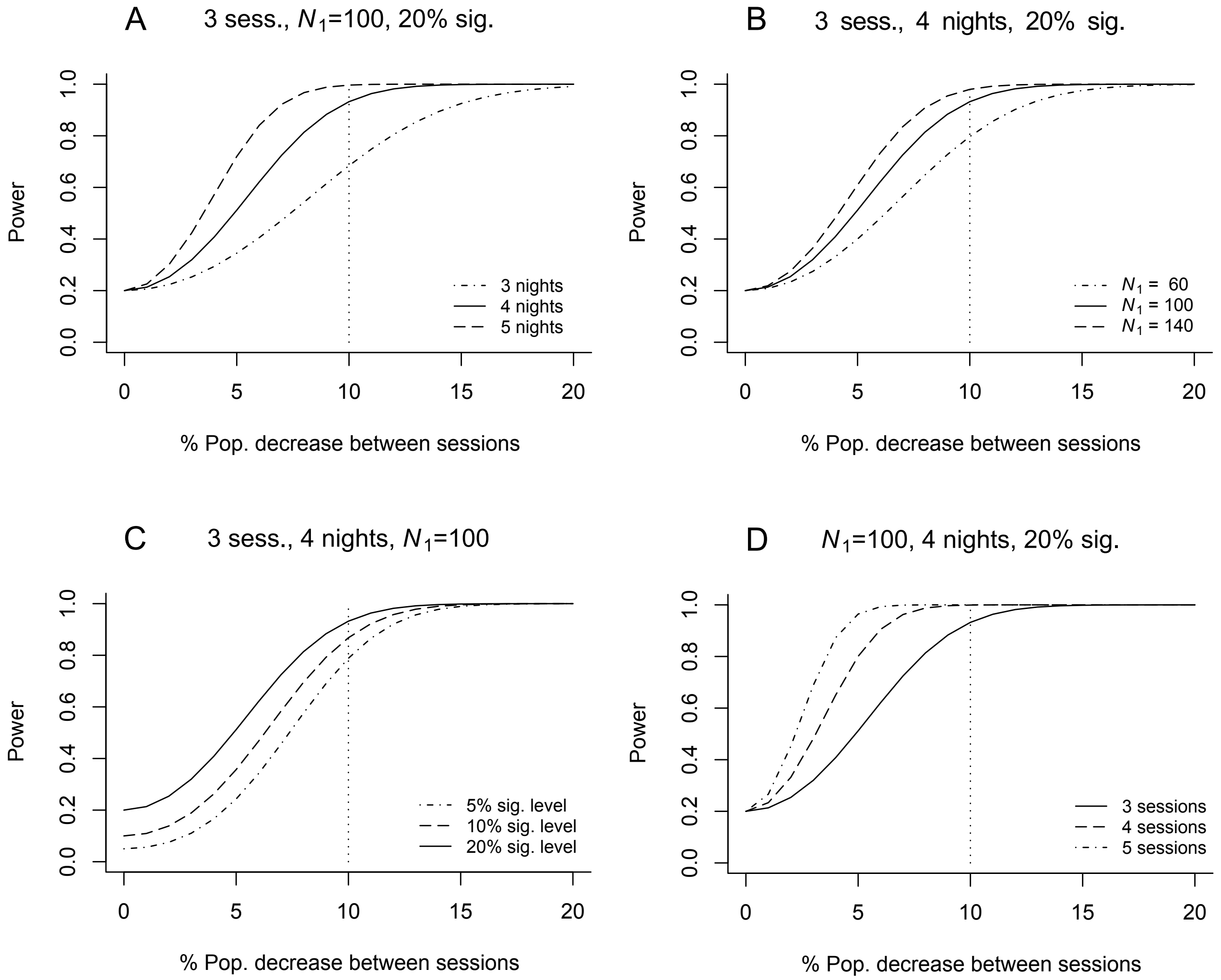

Figure 3. Power to detect a constant decline from session to session in population size N at all sites, using the last K sessions. The basic scenario has four sites, initial abundance $N_1 = 100$, average nightly capture probability $p = 0.3$, K = three sessions, $k = 4$ nights of sampling per session and a 20% significance level for the testing. Plots (A), (B), (C) and (D) show the effect of varying k, N_1, significance level and K respectively. The basic scenario appears as a solid line in all four plots, and the vertical line at the 10% drop is for reference.

cancel each other out. If there is heterogeneity of capture at the second session but not the first, there could be a false positive signal, as the second N is underestimated, so that even a steady population appears to decrease. A major problem could arise if the first session has more heterogeneity than the second. If there is a decrease in abundance, the more substantial underestimate of the first N could compensate for the subsequent drop and make the population look steady (a false negative result, no apparent drop). If at least 4 nights of sampling are completed in each session, we may check for the presence of heterogeneity and allow for it in the modelling.

Similarly, if temporary emigration is comparable at both sessions, detection of a drop is reliable. If there is more temporary emigration at the time of the second session, a false positive could occur, while the problem of a false negative could occur if there is more temporary emigration at the time of the first session. We are unable to test for temporary emigration in a short-term analysis of only two sessions.

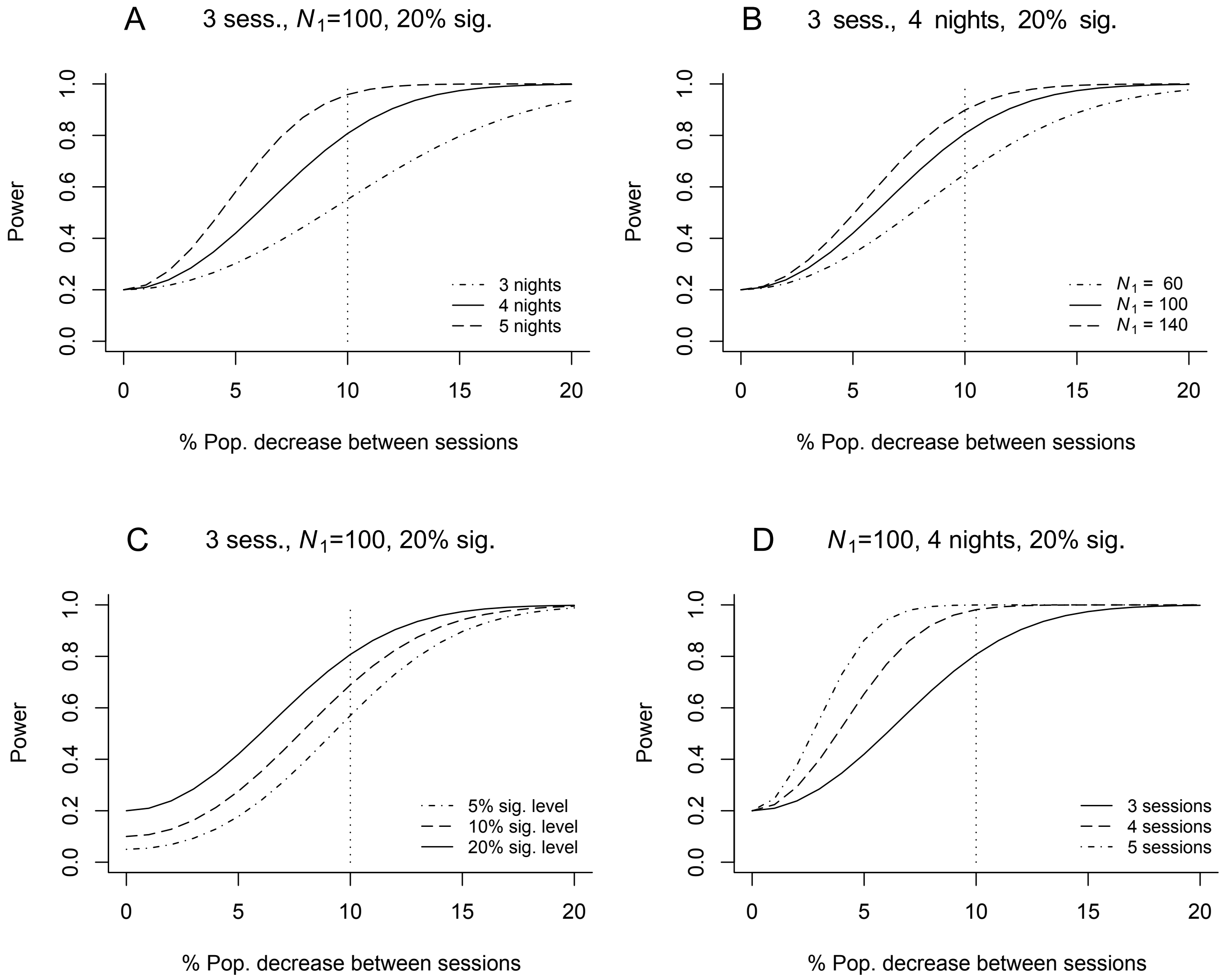

Figure 4. Power to detect a constant decline from session to session in population size N occurring at the non-treatment sites only, using the last K sessions. The basic scenario has four sites, initial abundance $N_1 = 100$, average nightly capture probability $p = 0.3$, K = three sessions, $k = 4$ nights of sampling per session, and a 20% significance level for the testing. Plots (A), (B), (C) and (D) show the effect of varying k, N_1, significance level and K, respectively. The basic scenario appears as a solid line in all four plots, and the vertical line at the 10% drop is for reference.

5.2.3 Tests three and four

Power curves for *long-term detection* of a downward trend in population over several sessions are in Figs 3 (Test 3, overall decline) and 4 (Test 4, decline on non-treatment grids only). The Schwarz & Arnason (1996) likelihood-based version of the Jolly-Seber model allows us to combine the likelihoods and do these tests using combined data over multiple grids. The basic scenario for Figs 3 and 4 is three sessions, 4 nights per session, initial population of 100, and a 20% significance level. Plots A to D show variation in number of nights, initial population, significance level and number of sessions, respectively. Even with only three sessions in the analysis, there is good power to detect an ongoing 10% decrease in population between sessions, provided there are 4 nights per session. If the decline is only on the non-treatment grids (Fig. 4), power is a little lower than for an overall decline (Fig. 3). With the basic scenario, power to detect an overall ongoing decline of 20% between sessions is 0.932, while if the same decline is only on the non-treatment grids the power is 0.807.

5.3 ALTERNATIVE SCENARIOS

Other values of population decrease may be considered. There is low power for immediate detection of a population decrease of 10% with the basic scenario and closed-population models. The power is 0.523 for a decline at all sites (Fig. 1) and 0.428 for a decline at only the non-treatment sites (Fig. 2). Power is improved if the decrease continues and data from the last three sessions are analysed using the Jolly-Seber model (power = 0.932, Figs 3 and 4), but by then the 10% decrease has occurred twice. With the closed-population models, bringing the power up to 0.9 for immediate detection of a 10% decrease would need an increase of initial *N* to 336 per grid, which would require grid size to increase by more than a factor of three. Alternatively, three or four times as many grids of size 10 m × 10 m would achieve this power. With a non-linear system like this, there are diminishing returns of increased power as the area sampled increases. If instead it is necessary to detect a 30% decrease, this power can be achieved with an initial *N* of only 35 per grid (decrease of 30% at all sites) or 59 per grid (decrease of 30% at treatment sites only). However, making the grids smaller for these values of initial *N* would again run the risk of obtaining insufficient data for capture-recapture analysis to be feasible. Reducing the number of grids instead would remove the essential replication at the treatment and non-treatment sites.

Tables of power from a more detailed investigation at all combinations of the factor levels are available as Excel files from the authors or the Waikato Conservancy Office, DOC, Hamilton.

6. Recommendations

6.1 STUDY DESIGN

Sampling design. The best capture-recapture monitoring design is Pollock's (1982) robust design, comprising a series of monitoring sessions, each consisting of a number of consecutive nights' searching. This allows different models to be compared, to evaluate not only population size, survival and recruitment, but also details of capture probabilities, behavioural responses and the presence of heterogeneity and/or temporary emigration, all of which can affect the population size estimates.

Number and size of grids. Power analysis showed that four 10 m × 10 m grids, two in the non-treatment area and two in the treatment area, will give good power for detection of changes in population size. The grids need to be placed in good frog habitat, as the calculations assumed about 100 frogs to be present on the grid.

Frequency of visits. The timing of visits may be varied to suit weather conditions, frog emergence and staff availability. More frequent visits will give faster detection of a decrease in population size. With the resources available, two or three visits per year seem possible. Visits need not be equally spaced, although a long gap will leave more time for a sharp decline (e.g. caused by chytrid fungus or drought) to go undetected.

Number of consecutive nights of sampling. Four or 5 nights of sampling are recommended. With only 3 nights the power to detect a decline is lowered and it is not possible to check for heterogeneity (which biases the *N* estimates). However, if sessions occur when only 2 or 3 sampling nights are possible, the data can still be used in the two basic analyses suggested in the second paragraph of this section (above), provided there are enough captures.

Minimum number of captures per session. Capture-recapture analysis methods become very unstable, or impossible, if the number of different animals caught within a session is less than about 25, or if the total number of within-session recaptures over the sampling nights is less than about 15. Flexibility about staying an extra night if the numbers look low would be helpful here, although recaptures take some time to identify from natural markings and the numbers would not be available in the field.

These recommendations are summarised in Table 4.

TABLE 4. SUMMARY OF PROPOSED MONITORING DESIGN.

ISSUE	RECOMMENDATION
Type and size of sampling unit	10 m × 10 m grid.
Number of replicates	Four grids in total with two in the treatment block and two in the non-treatment block.
Number of replicates monitored per night	One.
Number of consecutive nights per monitoring session/grid	4 nights minimum if heterogeneity is to be appraised. Otherwise, enough nights to give 25 first captures and 15 recaptures within the session.
Location of sampling units	Randomly (within 150 m of field base) in representative frog habitat that will yield high frog densities; approximately 100 frogs in a 10 m × 10 m grid.
Minimum number of captures per session	40 captures, with 25 first captures and 15 recaptures within the session.
Number of sessions per year	Two or three, depending on resources.
Timing of sessions	When frog emergence is high, e.g. November (late spring), January/February (summer), March (autumn).
Stratification	By treatment.
Effort required	1 month fieldwork per session, totalling either 8 or 12 weeks of field work depending on whether two or three sessions are completed.

6.2 NOTES

- For the monitoring programme, four grids should suffice. It is more important to ensure that each session has 4 nights of data and enough captures and recaptures in each session.
- The grid locations should be chosen in the same or similar habitat type, to enable valid comparison between samples. Locations with high frog density are best, for effective monitoring. Logistics (e.g. proximity to field base) will dictate the exact locations.
- Any attempt to extrapolate capture-recapture abundance estimates to the whole 600-ha population would require a much broader sampling scheme. There would need to be grids in all different habitat types, and grids with lower frog densities would need to be larger in order to obtain valid capture-recapture estimates.
- The focus in this report is on choosing a few sites to monitor intensively. They are not randomly chosen, but are selected for their high frog densities. This best meets the objective of detecting a decline in abundance. The populations are defined as the frog populations at these monitoring sites.
- Further stratification, by altitude (e.g. 500 m and 700 m above sea level) has been proposed but would reduce replication and the precision of estimates, and for these reasons it has not been attempted.
- Sampling sessions need not be spaced equally through the year. If the goal is to detect a decline within 2–8 months of it occurring, two or three sessions per year should suffice. Sampling in the seasons when frog emergence is highest (spring, summer, autumn) should reduce the risk of not getting enough captures. There is no way of deciding frequency and spacing of visits statistically—it is a matter of budget and common sense, and should be decided by management. More frequent visits of course give faster detection, so leaving a long gap is undesirable in case a sudden decline is missed altogether. More frequent visits cannot be bought at the expense of the minimum number of sites (four, for replication in the non-treatment and treatment blocks) and the minimum grid size and number of sampling nights per session to achieve the necessary numbers of captures for analysis to be possible (25 different animals per session per site, plus at least 15 recaptures, although occasional values below this may still acceptable).
- While sessions with 4 or more nights of sampling are ideal, as the evaluation of heterogeneity is possible, analysis will still be possible (provided there are enough captures) if field conditions sometimes dictate only 2 or 3 nights of sampling. There would be some loss of power to detect change.
- If the first few sessions show no sign of heterogeneity of capture, it would be safe to reduce the number of nights per session to three, provided enough different frogs and recaptures are still obtained.

- The grids will need to be monitored over 4 successive weeks, and the order of sampling them is important. It would be good if, say, low numbers on the treatment grids and high on the non-treatment grids (or vice versa) could be avoided. However, this is more likely to be driven by the weather than by calendar time. If the first week was on a treatment grid in dry weather and sample numbers were down, a weather forecast of continuing dry weather might suggest doing a non-treatment grid next, while a wet forecast could indicate that doing the other treatment grid next would redress the balance. Flexibility during the month would be an advantage; however, field logistics may limit such flexibility.

6.3 LOGISTICS AND RESOURCES

Up to 12 weeks of field monitoring per year could be budgeted into work plans, with a maximum of 4 weeks of field time budgeted for each monitoring session. On this basis monitoring four grids would require a week (5 days, 4 nights) to complete each monitoring session at each grid. At least three observers, preferably four, are needed to complete the field work.

There is some flexibility in this schedule, provided 4 consecutive nights are completed on each grid per session.

6.4 STATISTICAL ANALYSIS METHODS

A sequence of analyses would be best completed over the short, medium and longer term.

6.4.1 After each session

- Complete a short-term analysis of the last two sessions, using results from all grids. Use the closed-population model M(t), and test for changes in population density, as in Tests 1 and 2 in the power calculations.
- Use all the closed-population models to check if heterogeneity is present. If so, replace model M(t) with model M(t+h). This will only be possible if there are at least 4 nights per session.
- If there is evidence of a declining population, consider monitoring again fairly soon, or taking any other appropriate action.
- These analyses could be completed on MARK, but they are quite complicated to do. Annotated code that will run in either 'R' or S-Plus will be provided by the second author (Pledger).
- Two tests for monitoring abundance in the short term were described in Section 5. Other tests could also be formulated. One possibility is that in the short term, N decreases by a certain percentage in the non-treatment sites and a different percentage at the treatment sites. The model for the alternative hypothesis would have $M_1 = bN_1$, $M_2 = bN_2$, $M_3 = cN_3$ and $M_4 = cN_4$, and the null hypothesis would be that declines are similar over non-treatment and treatment grids, Ho: $b = c$. Code for this test is included.

6.4.2 In the medium term

- Use the Jolly-Seber model to check for a trend over the last few sessions.
- Analyses may be pooled, as in the power calculations, or done separately for each grid.
- Survival and recruitment rates may also be estimated and compared. Since survival is more precisely estimated than population size, any dataset with sufficient power to detect changes in N will also be able to detect changes in probability of survival, ϕ. Recruitment estimates are usually less precise.
- The version of the Jolly-Seber model on MARK does not yet allow for the estimation of trends, so the second author (Pledger) will provide annotated code that will run in either 'R' or S-Plus.
- Note that the Jolly-Seber model cannot deal with heterogeneity of capture.

6.4.3 In the longer term

- Once a longer run of sessions is available, the full robust design analysis is possible (Kendall et al. 1995; Kendall & Nichols 1995). This can take account of either heterogeneity or the influence of weather, or search effort covariates on the capture probabilities. As well as estimates of abundance, survival, recruitment and capture probabilities, it will provide estimates of temporary emigration (the proportion underground and unavailable for capture) for all sessions except the first and last.
- This is a standard analysis, which is now available on MARK, as the 'robust design with heterogeneity'.

7. Acknowledgements

Funding for this investigation was provided by the Department of Conservation Science Advice Fund (Investigation no. 2005/22). Jessica Wallace, university student on work placement to DOC 2003/2004, is specially thanked for collecting data in the first session, contributing to development of photographic techniques and data management from the first three sessions. Michael Crossland provided technical advice on photographic techniques and photo identification methods. Peter Lei assisted with individual frog identification and data management. The following people provided valuable assistance in the field: Michael Crossland, Peter Lei, Leigh Marshall, Kyle Chalmers, Jonathon McDonald, Mike McGirr and Mark Lamas.

8. References

Bell, B.D.; Pledger, S.; Dewhurst, P.L. 2004: The fate of a population of the endemic frog *Leiopelma pakeka* (Anura: Leiopelmatidae) translocated to restored habitat on Maud Island, New Zealand. *New Zealand Journal of Zoology 31*: 123–131.

Bradfield, K.S. 2004: Photographic identification of individual Archey's frogs, *Leiopelma archeyi*, from natural markings. *DOC Science Internal Series 191*. Department of Conservation, Wellington. 36 p.

Burnham, K.P.; Anderson, D.R. 2002: Model selection and multimodel inference: a practical information-theoretic approach. 2nd edition. Springer, New York 2. 353 p.

Eggers, K.E. 1998: Morphology, ecology and development of Leiopelmatid frogs (*Leiopelma* spp.), in Whareorino Forest, New Zealand. Unpublished MSc thesis, Massey University, Palmerston North. 169 p.

Fitzgerald, B.M.; Campbell, D.J. 2003: Examination of eight *Leiopelma* frogs from Whareorino Forest for signs of predation. Unpublished report, Ecological Research Associates of NZ, Wallaceville, Upper Hutt. 3 p.

Hitchmough, R.A.; Bull, L.; Cromarty, P. (Comps) 2007: New Zealand threat classification system lists—2005. Department of Conservation, Wellington. 194 p.

Jolly, G.M. 1965: Explicit estimates from capture-recapture data with both death and immigration—stochastic model. *Biometrika 52*: 225–247.

Kendall, W.L.; Nichols, J.D. 1995: On the use of secondary capture-recapture samples to estimate temporary emigration and breeding proportions. *Journal of Applied Statistics 22*: 751–762.

Kendall, W.L.; Pollock, K.H.; Brownie, C. 1995: A likelihood-based approach to capture-recapture estimation of demographic parameters under the robust design. *Biometrics 51*: 293–308.

Lebreton, J.D.; Burnham, K.P.; Clobert, J.; Anderson, D.R. 1992: Modelling survival and testing biological hypotheses using marked animals: a unified approach with case studies. *Ecological Monographs 62*: 67–118.

Otis, D.L.; Burnham, K.P.; White, G.C.; Anderson, D.R. 1978: Statistical inference for capture data on closed animal populations. *Wildlife Monographs 62*. 135 p.

Pledger, S. 1999: Monitoring protocols for Motuara Island Frogs *Leiopelma pakeka.* School of Mathematical Computer Sciences, Victoria University of Wellington, Wellington. 22 p.

Pledger, S. 2000: Unified maximum likelihood estimates for closed capture-recapture models using mixtures. *Biometrics 56*: 434–442.

Pollock, K.H. 1982: A capture-recapture design robust to unequal probability of capture. *Journal of Wildlife Management 46*: 752–757.

R Development Core Team 2005: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. www.R-project.org (viewed 10 July 2007).

Schwarz, C.J. 2001: The Jolly-Seber model: more than just abundance. *Journal of Agricultural, Biological and Environmental Statistics 6*: 195–205.

Schwarz, C.J.; Arnason, A.N. 1996: A general methodology for the analysis of capture-recapture experiments in open populations. *Biometrics 52*: 860–873.

Seber, G.A.F. 1965: A note on the multiple-recapture census. *Biometrika 52*: 249–259.

Skalski, J.R.; Robson, D.S. 1992: Techniques for Wildlife Investigations: design and analysis of capture data. Academic Press, San Diego.

Smale, A.; Holzapfel, A.; Crossland, M. 2005: Development of a capture-recapture monitoring programme for Archey's frog (*Leiopelma archeyi*) in New Zealand based on photographic identification of individual frogs. Proceedings of the Society for Research on Amphibians and Reptiles in New Zealand. *New Zealand Journal of Zoology 32*: 225.

Sutton, N. 2005: Presence/absence survey of Archey's frog in Whareorino Forest and the Herangi Range, Northern King Country. Unpublished report, Department of Conservation, Te Kuiti. 11 p.

Thorsen, M. 1998: Determination of a standardised methodology for long-term monitoring of mainland *Leiopelma* species. Unpublished report to the Department of Conservation, Hamilton. 34 p.

Thorsen, M. 1999: A resurvey of Archey's frogs Whareorino Forest and Mt. Moehau 18–24 August 1999. Unpublished report to the Department of Conservation, Hamilton. 8 p.

Thurley, T. 1996: A survey of native frogs (*Leiopelma archeyi* and *L. hochstetteri*) in Whareorino Forest, northern King Country. Unpublished MSc thesis, Victoria University of Wellington, Wellington. 97 p.

Thurley, T. 2000: Monitoring Archey's Frog (*Leiopelma archeyi*) at Whareorino Forest, August/September 2000. Unpublished report, Maniapoto Area Office, Department of Conservation, Te Kuiti. 9 p.

Thurley, T. 2003: Emergency predator control for the Whareorino Forest Archey's frog population. Unpublished report, Maniapoto Area Office, Department of Conservation, Te Kuiti. 4 p.

Thurley, T.; Bell, B. 1994: Habitat distribution and predation on a western population of terrestrial *Leiopelma* (Anura: Leiopelmatidae) in the northern King Country, New Zealand. *New Zealand Journal of Zoology 21*: 431–436.

Wakeland, M.; Smuts-Kennedy, C.; Thurley, T.; Webster, N. 2003: Artificial cover objects for Leiopelmatid frogs. *DOC Science Internal Series 120*. Department of Conservation, Wellington. 17 p.

Wallace, J, 2004: Identification of *Leiopelma archeyi* using natural markings: trialling two methods for taking digital photographs. Unpublished report, Department of Conservation, Hamilton. 66 p.

White, G.C. 2005: The MARK package. www.warnercnr.colostate.edu/~gwhite/mark/mark.htm (viewed 10 July 2007).

DOC Research & Development Series

DOC Research & Development Series *is a published record of scientific research carried out, or advice given, by Department of Conservation staff or external contractors funded by DOC. It comprises reports and short communications that are peer-reviewed.*

Individual contributions to the series are first released on the departmental website in pdf form. Hardcopy is printed, bound, and distributed at regular intervals. Titles are also listed in the DOC Science Publishing catalogue on the website, refer www.doc.govt.nz under Publications*, then* Science & technical*.*

New Zealand Government