Appendix S1. Thermal aerial control trial methods

Aerial hunting of deer using thermal imaging to detect deer was undertaken in the forested area of the upper Snag Burn over five days in early winter (7–9, 27, & 29 June) 2023. Weather conditions were clear and windless with air temperatures ranging between 0°–7°C (NIWA 2024). A MD 500D helicopter (MD Helicopters, LLC; Mesa, USA) was used with the following crew configuration: (i) Pilot on the forward left; (ii) thermal camera operator on the forward right; and (iii) shooter on the back right. The thermal camera was a FLIR T865® (Teledyne FLIR LLC; Wilsonville, USA) fitted with a 34mm lens with IR resolution of 640 x 480 pixels, refresh rate of 30 Hz and thermal sensitivity of <40 mK @ 30°C. This was paired with a class 4 high powered laser pointer and connected to a 12-inch HD screen in front of the operator and a 4k digital video recorder. The thermal camera was chest mounted and stabilised by a cinematic strong arm. An Android tablet showing treatment area boundaries and equipped for live tracking was used by the pilot and camera operator to monitor coverage in real time, enabling the area to be systematically searched.

The aerial sharpshooter utilised a semi-automatic .223 calibre Schmeisser AR15 M4FL (Krefeld, Germany) platform, fitted with a Pulsar Talion XG35 thermal scope (Ukon Advanced Optics Worldwide; Vilnius, Lithuania) and an auxiliary non-magnifying red dot optical scope. Nontoxic copper ammunition was used to avoid any risk of lead poisoning of the indigenous endangered alpine parrot, kea (*Nestor notabilis*), which is resident in the treatment area.

To aid detection of deer on buffer control flights we used a Pulsar Lexion XP38 thermal imaging monocular (Pulsar, Yukon Advanced Optics Worldwide, Lithuania) with a 38mm lens

with IR resolution of 640 x 480 pixels, 50 Hz refresh rate and thermal sensitivity of <40 mK @ 30°C.

Appendix S2. Estimation of population size using DNA-based mark-recapture - methods and analysis.

The January 2023 faecal pellet survey was conducted along transects spaced about 200 m apart (as per MacDonald (2019); Ebert et al. (2021)). A variable swath width averaging about 2 m was searched along each transect, wider in open areas and narrower in areas with dense fern cover. Slopes > 50 degrees were difficult to search, potentially biasing search effort. To evaluate search effort bias in relation to slope we applied a 10 m buffer to transects in 10° slope increments (0–10°, 11–20°, 21–30° and so on) and calculated the proportional area of slope searched in relation to the proportional area available in the treatment area.

We located 109 groups of pellets each considered to represent a single recent defecation event. Pellets from each group were swabbed with a sterile swab that was then stored in Longmire's buffer solution (Longmire et al. 1997). The swabbed pellets were stored in sterilised vials in case further extraction of fDNA was required. Samples were stored on ice after each search day.

In March 2023, DNA was extracted from the swabs using readymade reagents from the Qiagen® QIAamp 96 DNA QIAcube HT Kit following the manufacturer's instructions. All faecal samples underwent sex testing in duplicate, with male and female deer DNA targeted by specific amplification of a region of the Amelogenin gene using an adapted method from (Yamauchi et al. 2000). Sex test results were visualised through gel electrophoresis on a 4% agarose gel using GelRed™ stain and ultraviolet light via a Gel Doc XR+ System (Bio-Rad)

and compared to male and female known-sex positive controls. Both replicates were required to be positive to detect sample sex. Tentative results were reported where only a single replicate per sample produced a sex test result.

Genetic profiling was performed by amplifying 11 microsatellite markers for deer following an EcoGene® genotyping protocol for deer. Unfortunately, one marker (RT7) did not score consistently and so was excluded, resulting in only 10 microsatellite markers being available for DNA profiling. Microsatellite amplicons were visualised as above using gel electrophoresis and diluted (where required) based upon relative band strength ahead of capillary electrophoresis using an Applied BioSystems 3500xL Genetic Analyzer and the GeneScan™ 500 LIZ™ dye Size Standard. Resulting data files were analysed and alleles scored using GeneMapper v 6.0 (Applied Biosystems by Life Technologies). Final DNA profiles per sample, or consensus genotypes, were recorded where alleles were observed at each marker in at least two or more replicate profiles.

DNA profiles were grouped based on genotype similarity per marker and sex test information to give an estimate of the minimum number of contributing individuals. These groupings were made conservatively, allowing for potential allelic drop out per marker (where one of two possible alleles at a marker may fail to be detected) where an allele was in common between two sample profiles, but the genotype did not match (e.g. homozygotes compared to heterozygotes). Each identical grouping was designated as a hypothetical individual deer through DNA profile exclusions/genotype mismatches to other groupings.
We analysed the fDNA survey using the R (R Core Team 2025) package capwire (Miller et al. 2005; Pennell et al. 2013). This uses a maximum likelihood-based approach specifically designed for genomic data, enabling population size estimation with data from a single field

survey. As an urn model, it uses counts of detections within a single survey as a proxy for the number of times each identified deer was captured within that survey (Miller et al. 2005).

The model can account for heterogeneity of sampling by allowing (if necessary) for two different rates of detection (two-innate rates model, TIRM). This reduces the likelihood of bias in the population size estimate if there are two groups of individuals with different capture probabilities, even if the cause of this is not known. A likelihood ratio (LR) test was used to decide whether to use a TIRM model or, alternatively, an equal catchability model (ECM) where the null hypothesis is that all individuals are equally likely to have been captured. As the LR test did not identify a significant difference in capture probabilities ($p = 0.19$), we assumed that individuals were equally likely to have been detected, and the ECM model was therefore used.