Method Article

Can virtual fences reduce wombat road mortalities? Are severely under-powered studies worth the effort?

ABSTRACT

Aims: Evaluation of the effectiveness of methods of roadkill mitigation requires field experiments and thus experimental designs. We evaluated such a published experiment that used a virtual road-fence (VF) system to attempt to reduce roadkill of bare-nosed wombats (*Vombatus ursinus*) and draw general conclusions on the ability of the implemented design, given overall mean rates of roadkill, to detect a substantial reduction of 50%.
Study design: The study site consisted of contiguous fenced and unfenced road sections that were each monitored for road kill for 995 and 322 days pre- and post-installation, respectively. The study design was an unreplicated BACI (Before-After-Control-Impact) design.
Methodology: No statistical analysis of the roadkill data were carried out in the original study. We used a Poisson/log-link Generalized Linear Model fit to their BACI data as a single 2x2 table of counts and define a single interaction parameter as a function of control-adjusted rate reduction due to the VF. This parameter can also be defined as the logarithm of the ratio of odds of a random kill in the fenced section occurring in the post-installation period to the corresponding odds for the unfenced section. A null hypothesis of no effect of the VF corresponds to a log odds-ratio of zero. We use simulation to show that estimates of this parameter have close to a Gaussian distribution and from this derive an estimate of the statistical power of the design to detect a hypothetical effect of a 50% reduction in rate. We also used simulation to estimate the corresponding power to detect this reduction if the BACI had been physically replicated three times.
Results: The estimated log odds-ratio and the corresponding percent reduction inferred for the VF were -0.02 (SE=0.53) and 1.8%, respectively, and the statistical power of the design to detect a hypothetical effect of a 50% reduction in rate was only .35 but gave an improved power of .7 under the hypothetical replication.
Conclusion: Our results showed that there was very low confidence in this design combined with similar overall roadkill rates being able to detect a 50% reduction. This increased to medium-to-high confidence if the experiment had been replicated in three locations. The very small point estimate of reduction could have been so low compared to a true but modest effect purely by chance, thus our conclusion is that there is very low confidence in any estimate of the VF's effect obtained using this study's design in the context of common roadkill rates.

Keywords: roadkill mitigation device; BACI; MBACI; Poisson GLM; statistical power

1. INTRODUCTION

Many devices have been developed to mitigate roadkill and claims made as to their efficacy. However, when subjected to rigorous scientific research and evaluation these devices have been shown to have little or no effect in reducing roadkill [1,2]. There have been five studies using field trials in Australia to evaluate the efficacy of Wildlife Safety Solutions [3] DD430 virtual fencing (VF) (iPTE, Traffic Solutions Ltd., Graz, Austria) in reducing roadkill focusing on native marsupial species. Three have been published in peer-reviewed journals and two are unpublished.

Fox et al. [4] reported an almost 50% reduction in fenced road sections compared with the control sections for Tasmanian pademelons (Thylogale billardierii). The Fox et al. [4] study used an unreplicated Impact (i.e. VF operating) versus Control (i.e. no VF installed) design for the main statistical comparison using a paired t-test with replicates assumed to consist of 38 monthly counts of roadkill for each of (i) aggregation across species, (ii) Tasmanian pademelons, and (iii) Bennett's wallabies (Notamacropus rufogriseus) to compare fenced and unfenced road section roadkill rates (number $month^{-1}$ km^{-1}). This gave statistically significant t-statistics ($P<0.0001$ for i and ii and $P<0.05$ for iii). Note that their Impact vs Control comparison is completely confounded with road section within periods of Feb 2014 to Oct 2015 (3.2 km length fenced section) and Nov 2015 to Mar 2017 (above 3.2 km section plus 1.9 km extension to give 5.2 km length fenced section). The Control (i.e. unfenced) sections were defined by the total 13 km section minus fenced section within periods. They also used a "Before-After" comparison limited to the 1.9 km extension of the VF using a paired t-test calculated using replication consisting of 12 monthly counts. Fox et al. [4] note that their study did not employ spatial replication which means that "treatment" (i.e. Impact vs Control) level replication in the sense of [5] is absent. Coulson and Bender [6] and Englefield et al. [7] note these limitations while [6] notes additional limitations of [4] study with the response to [6] given by Fox and Potts [8].

Englefield et al. [7] estimated a range of 13% to 32% reduction in roadkill rates due to the VF for the three most prevalent species killed of Bennett's wallabies, Tasmanian pademelons and common brush-tail possums (BP) (Trichosurus vulpecula). Englefield et al. [7] obtained spatial replication within each of two "blocks" by turning six 750 m contiguous road sub-sections of a 4.5 km section of installed VF on and off according to a strict design protocol and also obtained two temporal replications by repeating the BACI (Before-After-Control-Impact) design [9] using two before-vs-after replications (i.e. the "blocks"). This gave a 2-replicate MBACI (Multiple BACI, [10]) each with three within spatial replicates. Using the MBACI's 24 roadkill rates (i.e. number per month per km) as a subset of the counts for the 6 sections by 5 periods for their study, which also considered cross-over and simple on-off comparisons of rates and the fit of a Gen-eralised Additive Model (GAM) [11] of counts assumed to be Poisson distributed with smooth spatial and temporal trends incorporated, they found no statistically significant reduction ($P>0.10$) in control-adjusted roadkill rate for each of the above three species they modelled. The above analyses in addition to that for the BACI also showed no statistically significant effect of the VF. Further, a simulation study they carried out confirmed the study design had a .78 power to detect a reduction of the order of 48%. Stannard et al. [12] correctly quote the above point estimates of percentage reduction in rates due to the VF in the discussion of their paper but in their introduction contradict this correct result by claiming "Englefield et al (2019) found that roadkill rates were reduced by 21 to 57% depending on species". This incorrect interpretation was also noted by Coulson and Bender [13]. The values of 21 to 57% do not depend on multiple species and are in fact artificially inflated reductions of actual roadkill rate due to the virtual fence (VF). These artificial reductions were applied in order carry out the above power analysis based

on simulated Poisson distributed kill numbers using an operating model derived from the fit of the above GAM to the Bennett's wallaby roadkill data.

Stannard et al. [12] did not estimate a percentage reduction but presented a monthly rate for bare-nosed wombats (Vombatus ursinus) for the study 1.5 km fenced road section of 0.7 month^{-1} for the 995 days of monitoring before VF installation and 0.6 month-1 for the 322 days of monitoring after installation. Their study design is an unreplicated BACI design. They reported raw roadkill counts and rates per month for each of the cells of the BACI which can by classified by a 2x2 table with columns of pre-installation ("before") and post-installation ("after") and rows of the monitored road section outside the fenced section ("control") and fenced section ("Impact"). However, no statistical analyses were described for these counts or monthly rates to determine if the reduction reported in the fenced section was statistically significant with or without adjustment for the "control" change in rate as given in Englefield et al. [7]. This lack of any statistical analysis of the roadkill counts was also noted by Coulson and Bender [13] who supplied the standard chi square analysis for a 2x2 contingency table for the null hypothesis of no association between the factors defining the table margins (i.e. Control vs Impact and Before vs After factors) (see Appendix). We compare and contrast that approach with the hypothesis testing using generalized linear models (GLM) [14] that we apply.

Appleby and Jones [15] and GDH [16], both unpublished studies, demonstrated that virtual fencing was not effective at reducing roadkill numbers.

2. METHODOLOGY

2.1 Conceptual Framework

BACI designs have the desirable property that (i) nuisance spatial effects or confounders are eliminated by the differencing, on the linear predictor scale in generalized linear models, of "after" minus "before" for the impact site (i.e. road section) mean rates, and (ii) nuisance temporal effects or confounders are eliminated by the second differencing of the "after" minus "before" mean rates for the control from the above difference for the impact site. Note (ii) assumes that the nuisance temporal effect is additive and thus independent of the temporal effect of interest which is the effect of the impact (i.e. operation of the VF). However, a persistent problem with BACI designs is in obtaining replicates of the experimental units at the level at which the intervention is applied (i.e. the "treatment" level), in this case the road section/monitoring period with the VF installed and switched on (i.e. "Impact" section). Subdividing an impact road section into contiguous sub-sections or the monitoring period into contiguous sub-periods to provide replicate counts for statistical analysis is termed pseudo-replication ([5]; see more recently [17]) and does not allow statistically valid uncertainty estimates for the estimate of the impact treatment (see [18] for some exceptions to this rule that are not relevant here). The set of monthly rates analysed using paired t-tests in Fox et al. [4] is an example of pseudo-replication since the treatments (i.e. impact vs control) were not replicated by design (i.e. the extension was a post-design ad hoc manipulation).

The debate on the utility or otherwise of this technology for mitigating wildlife roadkill needs a scientifically robust literature by considering, among other important issues, the necessity for valid statistical inferences [9,19]. To this end we emphasis in this commentary the validity of experimental designs and associated statistical inference that has been applied and provide a statistical inference procedure using the data given in Stannard et al. [12] to compensate for the lack of such in that study.

2.2 Experimental Design, Material and Methods

In order to provide some sense of the statistical significance of the results of monitoring roadkill in the BACI trial of Stannard et al. [12] we used their published data to fit a Poisson generalized GLM to the above 2x2 table of raw counts. Without spatial or temporal replication given that only a single 2x2 table of counts was available for estimation, it was necessary to assume a pure Poisson process generated the counts in order to estimate the uncertainty of the estimated effect of the VF. We use simulation to investigate the statistical distribution of the estimated effect and derive an estimate of the statistical power of the design to detect a hypothetical effect of a 50% reduction in rate. We also used simulation to estimate the corresponding power to detect this reduction if the BACI had been physically replicated three times. The assumption of a Poisson was investigated using the GAM analyses of data from Englefield et al. [7].

We used the data given in the Stannard et al. [12] (see Table 1 below) to fit a Poisson generalized linear model (GLM) with log link function [14] to the 2x2 table of raw counts with the corresponding monitoring periods expressed on a log scale as an "offset" (see Appendix) which effectively converts raw counts to period standardised rates. Taking into account main effects given by 2-level factors defined as "site" (i.e. impact site or "inside VF" vs control site or "outside VF") and "period" (i.e. pre- vs post-installation), a significant interaction with corresponding single parameter (i.e. single degree of freedom) that is negative and significantly different from zero would infer a reduction in roadkill due to the operation of the VF. Given this interaction parameter is say par the percent reduction in rate (R) is given by 100{1-exp(par)} so that par=log(1-R/100). Note also, that par corresponds to a log odds-ratio, where the numerator odds is the odds of a random roadkill occurring in the post-installation monitoring period for the "inside the VF" road section and the denominator being the corresponding odds for the "outside the VF" road section after adjusting for the different lengths of monitoring period (see Appendix). When the odds-ratio is unity, the log odds-ratio is zero corresponding to a null hypothesis that the effect of the VF on roadkill rates, that after adjusting for the control (i.e. nuisance pre- versus post-installation temporal effects), is zero, while a statistically significant negative log odds-ratio indicates there is a statistically detectable reduction in rate due to the VF. Assuming a Gaussian distribution for the estimate of par, the above hypothesis can be tested using a z-statistic which is the estimate of par divided by the estimate of its standard error which is then compared to the tails of a standard Gaussian (i.e normal) distribution. The adequacy of the above assumption of a Gaussian distribution is investigated using simulation as described below.

An identical estimate of par and its standard error is obtained if the 2x2 table is fitted as a binomial GLM with logit link function for two independent binomials, one for the Control section (i.e. binomial probability for the After count as a proportion of the total of Before and After) and similarly one for the Impact section, since constraining the corresponding Control and Impact marginal totals by fitting the corresponding margins as a factor in the Poisson GLM gives equivalent estimates of the non-marginal parameters since the sum of predictions gives exactly the marginal totals [14]. The hypothesis test for a log odds-ratio of zero is also equivalent in the binomial GLM to the constrained Poisson GLM (see Appendix).

The extra information provided by the pre- versus post-installation counts in the control road section the installed section of road is exploited in the above analysis to adjust for any temporal effects not explained by the operation of the VF as mentioned in the introduction (also see Appendix). This is the GLM equivalent to the estimated reduction given in Table 2 of Englefield et al. [7] for their multiple (i.e. replicated) BACI analysis (MBACI) using a linear

model fitted to standardised rates of roadkill (i.e. see their Table 2 footnote “Values within brackets are the Impact values after subtracting the corresponding Control estimates”).

The assumption of a Poisson distribution is a strong assumption that could under-estimate the uncertainty of the estimate of reduction in roadkill rate if the counts are over-dispersed relative to a Poisson. Such over-dispersion cannot be estimated from this 2x2 table of counts but in order to test empirical support for a pure Poisson we used count data from the Englefield et al. [7] study for each of the three species modelled to estimate the degree of over-dispersion and test if the difference from that of a pure Poisson process with theoretical dispersion parameter of unity was statistically significant. Without spatial replication in the Stannard et al. [12] study, the above analyses based on the assumption of a pure Poisson process (or another discrete distribution with known scale parameters) is required.

Finally, we used simulation of the 2x2 table of counts using the observed counts as expected values for a randomly drawn 1500 sets of 4 independent pure Poisson realisations and fitting the GLM to each table to obtain 1500 estimates of par. This was done to investigate the distribution of GLM estimates of par and in particular if it has close to a Gaussian distribution. The adequacy of the Gaussian assumption was investigated using quantile-quantile plot and confidence bounds for the 1-to-1 line based on both the inverse Kolmogorov test statistic and the theoretical normal point-by-point uncertainty using the R-package qqplotr [20]. Given the assumption of Gaussian distribution is reasonable we derived an estimate the statistical power (i.e. 1 minus the probability of a type II error of accepting the null hypothesis when the effect size, par, is non-zero) of the design to detect a hypothetical effect size of an R=50% reduction in rate using the relationship between R and par and the median across the simulation sample of the 2x2 table GLM estimate of the standard error of the estimate of par (see Appendix). We also used three “blocks” (i.e. replicates) of the above simulation to estimate the corresponding power of this reduction if the BACI was replicated three times as separate tables, physically corresponding to the trial being replicated at the same site over three periods or at three similar sites. This was done by simply dividing 1500 sets of 2x2 tables into 500 sets of three replicate blocks without adding between-block variation in the counts. This is because we estimate par for each block using an interaction of “Block” with the par-parameter interaction term and then average the three estimates and then obtain the standard error of this average using only the standard errors of the individual block estimates of par (i.e. ignoring between-block random differences in par) (see Appendix).

The analyses were carried out using the R-software [21] and the code and output is provided in Supplementary Material.

3. RESULTS AND DISCUSSION

Table 1 is the data used to fit the Poisson GLM extracted or calculated (Supplementary Material) from descriptions in Stannard et al. [12].

The estimated log odds-ratio, par, and the corresponding percent reduction, R, were -0.02 (SE=0.53) and 1.8%, respectively, for the observed 2x2 table given in Table 1. Clearly the null hypothesis of a log odds-ratio of zero is not rejected with two-side probability of acceptance of .97. The lack-of-fit chi-square statistic with single degree of freedom (i.e. sum of squares of Pearson residuals) for the Poisson, and equivalently for the binomial GLM, without the interaction term was 0.0011 (P=.97) which is also the same chi-square statistic obtained by Coulson and Bender (2022) (see Appendix).

Table 1. Observed frequency of roadkills of bare-nosed wombats (*Vombatus ursinus*) in the control and impact road sections before and after installation of a virtual fence.

Road section	Length (km)	Pre-installation (Before)		Post-installation (After)	
		Monitoring Period (d)	Roadkill Count (rate per month[1])	Monitoring Period (d)	Roadkill Count (rate per month[1])
Unfenced (Control)	21.2	995	64 (1.93)	322	17 (1.58)
Fenced (VF) (Impact)	1.5	995	23 (0.69)	322	6 (0.56)

[1] Define month as 30 days.

The simulation showed that estimates of this parameter have close to a Gaussian distribution (Supplementary Material) and from this an estimate of the statistical power of the design to detect a hypothetical effect of an R=50% reduction in rate of .35 was obtained. The simulation study to estimate the corresponding power of this reduction if this BACI had been replicated as three blocks of 2x2 tables gave a closer to Gaussian distribution for the average of the three block values of par than that for the simulation of the unreplicated BACI (Supplementary Material) and gave an improved power of .69.

Based on the fit of the generalized additive model (GAM) [11] with Poisson response distribution to the roadkill data in the Englefield et al. [7] study with "offset" [14] of the sum of log of period of monitoring and log of road section length to which the roadkill was attributed (i.e. 6 replicate sections of 750 m each monitored over 5 periods ranging from 14 to 28 d) for each of the species of Bennett's wallabies (BW), Tasmanian pademelons (TP), and common brush-tail possums (BP) the residual deviances with 26 degrees of freedom (df) were 21.6, 23.4, and 36.7, respectively. For the total of all marsupials the residual deviance was 24.43. The quantile of the central chi square distribution with 26 df corresponding to a 0.9 probability is 35.6. Therefore, only for the BP would the response distribution be inferred to be over-dispersed relative to a Poisson and even in this case the over-dispersion is minor with dispersion parameter estimated as only slightly greater than 1 (i.e. approximately 1.41=36.7/26).

Theoretically, the Poisson is a discrete probability distribution for a random variable of the number of events (e.g. roadkills) occurring in a fixed interval of time or space if these events occur with a known constant mean rate and independently of the time since the last event [22]. Due to the reproductive property of the Poisson distribution that the sum of independent Poisson distributions is also Poisson with expected rate the sum of the individual Poisson rates, roadkill counts for small fixed length road sections and time periods, if independently Poisson distributed, can be aggregated across these road length sections and periods without loss of information. It can reasonably be argued that the distribution of roadkill counts should follow a Poisson at least approximately. The empirical analysis of the counts from Englefield et al. [7] largely confirms this assumption, though, lack of independence in counts, which are in effect fatal encounters for wildlife with vehicles where those encounters may be spatially clustered, between contiguous road sections could lead to over- or under-dispersion relative to a pure Poisson process. Nevertheless, in practical terms based on the above empirical results, the assumption of a pure Poisson purpose is considered adequate for focusing the discussion of this mitigation technology on, among other important issues, the necessity for valid statistical inference.

The inferences using the estimation of par we give and that using the chi square test of no association (see Appendix) of Coulson and Bender [13] are in agreement. However, the advantage of our approach is the ability to carry out the statistical power calculations we give.

The overall message in Stannard et al. [12] given in their Abstract that "Virtual fencing implemented in regions that have high wombat roadkill rates may aid in reducing road deaths and species conservation" we believe is unjustified given our results and those of [13] using the data from [12]. Despite the lack of spatial and temporal replication in their study the raw data showed very similar rates pre- and post-installation in the VF-installed road section with rates of 0.7 and 0.6 kills per month, respectively. In the control road section, the monthly rate was also reduced pre- to post-installation with rates of 1.9 and 1.6 kills per month. Note that these last rates when calculated as a rate per month and per kilometre to give 0.091 and 0.075 $month^{-1}km^{-1}$, respectively, are much less that the corresponding rates for the fenced section of 0.46 and 0.37 $month^{-1}km^{-1}$ pre- to post-installation, respectively. This large difference in pre-installation rates between Control and Impact sections does not invalidate the BACI analysis since this spatial effect is removed by the Before to After contrast as mentioned in the Introduction. That there was a reduction in both sections is relevant (see below). These rate results, even though they were bereft of any statistical uncertainty estimates or hypothesis testing, should have been described in the Abstract rather than as the raw counts which without the corresponding time interval of the monitoring period and the length of the section of road monitored are by themselves uninformative. To be comparable across studies roadkill numbers should be standardised for both monitoring period length and length of road section for example as a rate per month per kilometre as given in Englefield et al. [7].

It can be seen from these simple empirical analyses/simulations that this study design is not able to detect as statistically significant substantial and thus practically significant notional effects of the VF. We acknowledge the challenge given limited research resources to implement study designs that have reasonable power to detect practically significant effects of the VF. However, Englefield et al. [7] using daily observations of roadkill on a 4.5 km road section were able to detect effects of close to 48% or greater with acceptable power (=0.78) in a similar study using a modest level of spatial replication. As Englefield et al. [7] show, it is possible to build spatial replication at the "treatment" level into trials of the effectiveness of the VF, with some constraints such as the road sections of VF when switched on being of adequate length so that spill-over effects into adjacent control sections do not substantially inflate or deflate the effect of the VF. We specify potential inflation versus deflation since such spill-over effects could possibly operate via two opposing mechanisms that would only be detectable if the VF was sufficiently effective in reducing roadkill or may in fact be undetectable if they negate each other. First, an effective VF could cause animals approaching the boundary of the Control and Impact road sections to avoid the latter section and inflate the road-kill near the boundary within the adjacent control section, as noted by Coulsen and Bender [6] as the "fence-end" issue, thus over-estimating the reduction in rate due to the VF. Alternatively, any deterrent effect of the VF due to the audio and visual stimulii may spill-over and reduce roadkill in the part of the control road section immediately adjacent to the operating VF section as considered by Englefield et al. [7]. If the Control and Impact road sections are sufficient in length these effects, if they exist in toto, should be minimal when spread over the full length of a given Control road section. The key point is that spatial (i.e. "treatment") replication is possible with the VF as shown by Englefield et al. [7] whereas some mitigation infrastructures, such as crossing structures combined with physical fencing [19], do not logistically allow manipulation to achieve "treatment" replication within study sites. Therefore, it is puzzling why this advantage of VF roadkill mitigation trials was not exploited in Fox et al. [4] or Stannard et al. [12]. Further, the relatively short length of

the fenced section 1.5 km in [12], relative to similar studies [4,7,14,15] that ranged from 3.8 km to 5.2 km and relative to the monitored un-fenced section of 21.2 km, is also puzzling. The limitation this placed on the number of potential road kills in the fenced section combined with the considerable resources spent on monitoring 22.7 km of road for almost four years, suggest that a better balance would have been to have multiple 1.5 km fenced sections interspersed with shorter control sections to give spatial replication within the study road or, alternatively, setup one fenced and one control section per road and replicate this across similar roads in the local area. In terms of monitoring periods, Rytwinski et al. [18] recommend a minimum of either four years monitoring or four replicate sites for Before-After-Control-Impact designs. The Stannard et al. [12] study had a total duration of almost four years but statistical power was still unacceptably low (0.35 to detect a 50% reduction) while the simulation study that replicated the study three times increased the power to a more adequate value of 0.7. Based on these results Rytwinski et al. [19] recommendation could be tightened by specifying a minimum of four years of monitoring combined with three or even four replicate sites each with similar levels of roadkill to the study section. A similar total number of roadkills for a particular comparison (i.e. individual species or combinations of species) is required in the above recommendation since this total affects the precision of mean rates with precision improving with greater kill numbers based on the assumption of Poisson counts and the relationship between the Poisson mean and its variance as noted by Englefield et al. [7].

4. CONCLUSION

Stannard et al. [12] provide in effect one data point, that is the reduction in rate within the fenced section of the road (i.e. without any statistical uncertainty). Using our analysis of the 2x2 table of counts in Table 1, the point estimate of the effect of the VF, adjusted for temporal confounders, expressed as a log odds-ratio combined with its associated standard error could be used in a meta-analysis when combined with corresponding estimates and their standard errors from any future trials [19]. However, Hulbert [18] notes that meta-analyses combining low-power individual studies is no panacea stating that “When for lack of treatment replication, estimates of effect size contain large amounts of ‘noise’ or random error, the output of a meta-analysis will also be ‘noisy’”.

In conclusion, there is very little that can be inferred on the effectiveness of the VF for reducing the roadkill of bare-nosed wombats from this study alone. Therefore, no assessment of the efficacy of this VF on reducing bare-nosed wombat roadkill based on the Stannard et al. [12] study results was warranted even when couched in cautionary language. Their conclusion that this “virtual fencing may provide suitable mitigation strategies to reduce wombat roadkill rates, however further data is needed” would serve future research better if rather than simply recommending “further data” (collection) it was replaced by the statement that “the VF might or equally might not be a useful mitigation measure thus requiring additional studies with acceptable levels of statistical power to draw more definitive conclusions”. For the Stannard et al. [12] study, given its severely under-powered experimental design and given the research resources that were committed, the answer to the question posed in the title that can be inferred from the above is unfortunately ”no”.

Supplementary Materials: The following are available online at …,

S1 Figure S1: Gaussian qqplots using function in R-library qqplotr for 1500 estimates of par from simulations of 2 x 2 Poisson distributed BACI counts, Figure S2: Gaussian qqplots using function in R-library qqplotr for estimates of par from 500 simulations of 3 blocks of 2 x 2 Poisson distributed BACI counts after averaging par across the 3 blocks

S2 R-code to carry out analyses of Table 1 and simulation study of empirical distribution of "par"

S3 R-code output

APPENDIX

Let informative subscripts for the 2x2 table (see Table 1) be given for rows (Control, C; Impact, I) and columns (Before, B; After, A). Let the observed counts of roadkill for the 2x2 table be given by catenating columns $(n_{CB}, n_{IB}, n_{CA}, n_{IA})$, the corresponding monitoring periods (units of 30-day month) be $(t_{CB}, t_{IB}, t_{CA}, t_{IA})$, and corresponding length of road sections (units of km) be (l_C, l_I), respectively, noting that logically $l_{CB} \equiv l_{CA} = l_C$, and $l_{IB} \equiv l_{IA} = l_I$. Note that l_C is assumed as unknown here since its value was not reported in Stannard et al. [12]. Let the expected value of the counts under repeat sampling using the same values of $(t_{CB}, t_{IB}, t_{CA}, t_{IA})$ and (l_C, l_I) be $(t_{CB} l_C \mu_{CB}, t_{IB} l_I \mu_{IB}, t_{CA} l_C \mu_{CA}, t_{IA} l_I \mu_{IA})$ so that $(\mu_{CB}, \mu_{IB}, \mu_{CA}, \mu_{IA})$ are expected roadkill rates (month^{-1} km^{-1}) and let $(\eta_{CB}, \eta_{IB}, \eta_{CA}, \eta_{IA}) = \log_e(\mu_{CB}, \mu_{IB}, \mu_{CA}, \mu_{IA})$ be log of expected rates.

The GLM is expressed as

$$\mathbf{y} = \exp(\mathbf{o} + \mathbf{X}\boldsymbol{\beta}) + \mathbf{e}$$

where

$$\mathbf{y} = (n_{CB}, n_{IB}, n_{CA}, n_{IA})^T$$

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \eta_{CB} + \log_e(l_C) \\ \eta_{IB} - \eta_{CB} + \log_e(l_I / l_C) \\ \eta_{CA} - \eta_{CB} \\ \eta_{IA} - \eta_{IB} - (\eta_{CA} - \eta_{CB}) \end{bmatrix}, \quad \mathbf{o} = \begin{bmatrix} \log_e(t_{CB}) \\ \log_e(t_{IB}) \\ \log_e(t_{CA}) \\ \log_e(t_{IA}) \end{bmatrix},$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4)^T$ is the vector of parameters to be estimated and β_4, using an obvious notation as the 4th element of $\boldsymbol{\beta}$, is the interaction parameter denoted *par* in the main text. The parameter β_4 is denoted *par* in the main text. The response count variable $\mathbf{y}$ is assumed to be Poisson distributed with expected value $\exp(\mathbf{o} + \mathbf{X}\boldsymbol{\beta})$ and $\mathbf{o}$ is the "offset" variable in the definition of the GLM. Note that the fitted GLM is fully saturated with parameters so that the residual deviance is zero and residuals are zero but given the scale parameter [14] is defined as unity for the Poisson then standard errors of parameter estimates can still be obtained.

The interaction parameter β_4 corresponds to the difference in log of rate effects, $\eta = \log_e(\mu)$, (i.e. on the linear predictor scale) of post-installation rate minus pre-installation rate for "inside VF" minus post-installation rate after subtracting pre-installation rate for "outside VF" as given mathematically by $\eta_{IA} - \eta_{IB} - (\eta_{CA} - \eta_{CB})$. This has the effect of reducing the estimated reduction in rate due to the VF for any reduction from pre- to post-installation rate for the "outside the VF" road section. Conversely, the effect would be to increase the estimated reduction in rate due to the VF for any increase in rate from pre- to

post-installation for the "outside the VF" road section. The principle is to adjust the estimated reduction in rate attributed to the VF for any reduction/increase in rate that is unrelated to the operation of the VF which is true *a priori* for the "outside the VF" road section but is also assumed to operate equally within both two road sections and is assumed to be additive to any effect of the VF on the linear predictor scale. Note that the Control road section length l_C is subsumed into the β_1 and β_2 parameters via $\log_e(l_C)$ and $\log_e(l_I / l_C)$, respectively, but these last two values do not appear in the parameter of interest β_4 so that excluding the logarithm of road section lengths from the offset term has no effect on β_4.

Expressing the above principle in model parameters, the adjusted percentage reduction in roadkill rate, *R*, inferred for the Impact treatment is given by

$$R = 100\left\{1-\exp(\beta_4)\right\} = 100\left\{1-(\mu_{IA}/\mu_{IB})(\mu_{CB}/\mu_{CA})\right\}$$
$$= 100\left\{1-(\mu'_{IA}/\mu'_{IB})\right\} = 100\left(\mu'_{IB}-\mu'_{IA}\right)/\mu'_{IB}$$

where $\mu'_{IA} = \mu_{IA}\mu_{CB}$ and $\mu'_{IB} = \mu_{IB}\mu_{CA}$. Conversely, the parameter β_4 can expressed as a function of *R* as $\beta_4 = \log_e(1-R/100)$.

Note, that β_4 corresponds to a log odds-ratio, where the numerator odds is the odds of a random roadkill occurring in the post-installation monitoring period for the "inside the VF" road section and the denominator being the corresponding odds for the "outside the VF" road section.

This property of the above GLM parameterisation can be seen since

$$\beta_4 = \log_e\left\{\left(p_{IA}/[1-p_{IA}]\right)\left(p_{CA}/[1-p_{CA}]\right)^{-1}\right\}$$

where $p_{IA} = \mu_{IA}/\mu''_I$, $\mu''_I = \mu_{IA}+\mu_{IB}$, $p_{CA} = \mu_{CA}/\mu''_C$, and $\mu''_C = \mu_{CA}+\mu_{CB}$.

When these odds-ratio is unity, the log odds-ratio is zero corresponding to the null hypothesis that the control-adjusted effect of the VF on roadkill rates is zero, while a statistically significant negative log odds-ratio indicates that after adjusting for nuisance pre- versus post-installation (i.e. temporal) effects, there is a detectable reduction in rate due to the VF.

Note also that since t_{CB} was identical to t_{IB}, (i.e. t_B) as was the case for t_{CA} and t_{IA}, (i.e. t_A) and which is the usual case for BACI designs, the offset is strictly only necessary in the more general case since the values of t_B and t_A can be subsumed in a similar way to section length in this case using both parameters β_1 and β_3. Also, if the interaction term is dropped from the GLM corresponding to dropping the parameter β_4 from the fit then the residuals are no longer constrained to zero so that the residual deviance is greater than zero since the GLM is no longer saturated with parameters. The Pearson chi square lack-of-fit test statistic, as an alternative to the residual deviance, is the sum of the squared Pearson residuals [14] across the 4 cells of the 2x2 table.

If we fit a binomial/logit GLM to the cell counts n_{CA} and n_{IA} conditional on binomial totals $n_{CB}+n_{CA}$ and $n_{IB}+n_{IA}$, respectively, with terms of a regression constant and a main-effect factor with level 1 as Control and level 2 as Impact, then the single parameter for the contrast of Impact minus Control is the above expected log odds-ratio, since

$$\log_e\left\{\left(p^*_{IA}/\left[1-p^*_{IA}\right]\right)\left(p^*_{CA}/\left[1-p^*_{CA}\right]\right)^{-1}\right\}=\log_e\left\{\mu_{IA}\mu_{CB}/\left(\mu_{IB}\mu_{CA}\right)\right\}$$

$$=\log_e\left\{\left(p_{IA}/\left[1-p_{IA}\right]\right)\left(p_{CA}/\left[1-p_{CA}\right]\right)^{-1}\right\}$$

where

$p^*_{IA}=t_A\mu_{IA}/\left(t_B\mu_{IB}+t_A\mu_{IA}\right)$

$p^*_{CA}=t_A\mu_{CA}/\left(t_B\mu_{CB}+t_A\mu_{CA}\right).$

The null hypothesis of a zero log odds-ratio is therefore tested with the parameter estimate $\hat{\beta}_4$ and its estimated standard error using either Poisson or binomial GLMs (see Supplementary Material for analysis results) since $H_0:\mu_{IA}\mu_{CB}/\left(\mu_{IB}\mu_{CA}\right)\equiv 1\Rightarrow p^*_{IA}\equiv p^*_{CA}=p^*_A\Rightarrow\beta_4\equiv 0$.

The Pearson chi square statistic for the above null hypothesis is a lack-of-fit statistic when β_4 is set to zero rather than a z-statistic for the estimate of β_4 whose distribution was studied in the simulations described in the methods section. The Pearson chi-square statistic under equal binomial probabilities (*ep*) as in H_0 above, is obtained as the usual sum across cells of the squared Pearson residuals, which for the binomial GLM is given by

$$X^2_{ep}=\left[n_{CA}-\hat{E}_{ep}\left(n_{CA}\right)\right]^2/\hat{V}_{ep}\left\{\hat{E}\left(n_{CA}\right)\right\}+\left[n_{IA}-\hat{E}_{ep}\left(n_{IA}\right)\right]^2/\hat{V}_{ep}\left\{\hat{E}\left(n_{IA}\right)\right\}$$

where $\hat{E}_{ep}\left(.\right)$ represents the estimate of the expectation over repeat sampling given fixed values for $\left(l_C,l_I,t_A,t_B\right)$ and each of the Control and Impact margins, and similarly $\hat{V}_{ep}\left(.\right)$ is the estimate of the variance and $\hat{E}_{ep}\left(n_{CA}\right)=\left(n_{CA}+n_{CB}\right)\hat{p}^*_A=\left(n_{CA}+n_{CB}\right)\left(n_{CA}+n_{IA}\right)/N$ and

$\hat{E}_{ep}\left(n_{IA}\right)=\left(n_{IA}+n_{IB}\right)\hat{p}^*_A=\left(n_{IA}+n_{IB}\right)\left(n_{CA}+n_{IA}\right)/N$

where $N=n_{CA}+n_{CB}+n_{IA}+n_{IB}$ and $\hat{p}^*_A$ is the maximum likelihood estimate. We calculate the variances of the After expected cell counts for each of the Control and Impact rows of the 2x2 table as independent binomials substituting the maximum likelihood estimate of p^*_A into the usual variance formula

$$\hat{V}_{ep}\left\{\hat{E}_{ep}\left(n_{CA}\right)\right\}=\left(n_{CA}+n_{CB}\right)\hat{p}^*_A\left(1-\hat{p}^*_A\right)=\hat{E}_{ep}\left(n_{CA}\right)\left(1-\hat{p}^*_A\right)$$

$$\hat{V}_{ep}\left\{\hat{E}\left(n_{IA}\right)\right\}=\left(n_{IA}+n_{IB}\right)\hat{p}^*_A\left(1-\hat{p}^*_A\right)=\hat{E}_{ep}\left(n_{IA}\right)\left(1-\hat{p}^*_A\right).$$

The chi square statistic used by Coulson and Bender [12] is for the test of no association (*na*) between the Control vs Impact and the Before versus After factors, that is, this chi square is a lack-of-fit statistic under the null hypothesis that the expected values of the 2x2 cell counts are determined completely by the product of the marginal probabilities after multiplication by N. This chi square is given by

$$X^2_{na}=\left[n_{CA}-\hat{E}_{na}\left(n_{CA}\right)\right]^2/\hat{E}_{na}\left(n_{CA}\right)+\left[n_{IA}-\hat{E}_{na}\left(n_{IA}\right)\right]^2/\hat{E}_{na}\left(n_{IA}\right)$$

$$+\left[n_{CB}-\hat{E}_{na}\left(n_{CB}\right)\right]^2/\hat{E}_{na}\left(n_{CB}\right)+\left[n_{IB}-\hat{E}_{na}\left(n_{IB}\right)\right]^2/\hat{E}_{na}\left(n_{IB}\right).$$

It can be seen that

$\hat{E}_{ep}\left(n_{CA}\right)=\hat{E}_{na}\left(n_{CA}\right)$ and $\hat{E}_{ep}\left(n_{IA}\right)=\hat{E}_{na}\left(n_{IA}\right)$ since $\hat{E}_{na}\left(n_{CA}\right)=N\left(n_{CA}+n_{CB}\right)\left(n_{CA}+n_{IA}\right)/N^2$ and $\hat{E}_{na}\left(n_{IA}\right)=N\left(n_{IA}+n_{IB}\right)\left(n_{CA}+n_{IA}\right)/N^2$ with corresponding estimates $\hat{E}_{na}\left(n_{CB}\right)$ and $\hat{E}_{na}\left(n_{IB}\right)$ similarly derived. Finally note that

$X^2_{na}=X^2_{ep}$ since $\left[n_{CA}-\hat{E}_{na}\left(n_{CA}\right)\right]^2/\hat{E}_{na}\left(n_{CA}\right)+\left[n_{IA}-\hat{E}_{na}\left(n_{IA}\right)\right]^2/\hat{E}_{na}\left(n_{IA}\right)=$
$\left[n_{CA}-\hat{E}_{ep}\left(n_{CA}\right)\right]^2/\left[\hat{E}_{ep}\left(n_{CA}\right)\left(1-p^*_A\right)\right]$

and

$\left[n_{CB}-\hat{E}_{na}\left(n_{CB}\right)\right]^2/\hat{E}_{na}\left(n_{CB}\right)+\left[n_{IB}-\hat{E}_{na}\left(n_{IB}\right)\right]^2/\hat{E}_{na}\left(n_{IB}\right)=$

$\left[n_{IA}-\hat{E}_{ep}\left(n_{IA}\right)\right]^2/\left[\hat{E}_{ep}\left(n_{IA}\right)\left(1-p^*_A\right)\right]$.

Therefore, the Poisson GLM, binomial GLM, and no-association lack-of-fit chi-square statistics calculated as Pearson chi-squares all give exactly the same statistic for the equivalent null hypothesis, with each expressed in different forms, that the Impact factor level when combined with the After factor level has no effect in addition to that given by the simple main effect model terms which in the context here is a null hypothesis that there is no mitigation benefit of the virtual fence.

Note that the 2x2 table of counts used in this study can be considered as 4 independent Poisson distributed counts so fitting main effects for the Poisson GLM was necessary to estimate all parameters in $\boldsymbol{\beta}$ with the incidental result that all margins were fitted exactly, given this property of the Poisson GLM with canonical link of the natural log function [14]. For 2x2 contingency tables in other types of studies such as those obtained in case-control studies in epidemiology, possibly including J-strata of 2x2 tables (e.g. [23]), more complex distributional assumptions such as the extended hypergeometric [24] are required to account for conditioning on the margins when these have been artificially constrained by the sampling design. Such artificial constraints on sampling at the marginal-factor level do not occur in BACI-type studies. However, where sampling probabilities are unequal since typically as in this particular study $l_C \neq l_I$ and $t_B \neq t_A$ the variable sampling probabilities at the within-table (i.e. cell) level are accounted for in the Poisson GLM either indirectly by being subsumed in the main effects or directly using the offset term. Therefore, the BACI 2x2 contingency table can validly be assumed to consist of 4 independent Poisson distributed counts. These results are demonstrated empirically in the Supplementary Material.

The 2x2 table was replicated using the three independent "Blocks" described in the Section 3 simulations to give a 3-replicate MBACI design, then fitting the Block factor as main effect and interactions of Block with each of the Control vs Impact factor, Before vs After factor, and their interaction, but excluding these last two factors and their interaction as additional main effect terms, then there is an estimate of β_4 and a corresponding estimate of its standard error for each level of Block. These replicated values were averaged with the standard error of this average given by the square root of the sum of the squared standard deviations with the square root then divided by three (i.e. ignoring between-block random differences in *par*). Using "Block" as a fixed effect in this way considers the trials as a "detection-device" for the any reduction effect of the VF rather than an "across-population-units" estimated effect (i.e. blocks as random effects) with associated uncertainty, since the former gives a more precise estimate to determine if there is a detectable effect of given size (e.g. 50% reduction) in the first place. Note that this approach of fitting blocks as fixed effects was used for the two replicate blocks of the BACI design data described in Englefield et al. [7] and importantly there is no loss of information on the VF effect because the comparison or contrast that quantifies the effect of the VF is estimated within-blocks. When there is such

a detectable effect, how it may vary quantitatively across random sites and periods is a question that can be addressed once the former is confirmed. Therefore, for this "detection" objective there was no requirement to add random block-level extra-Poisson variation into the simulated counts despite such random block-level variation being present in a real-world trial since such variation is removed by the above approach of considering blocks as fixed effects. Note that the replicate estimates of β_4 have close to zero covariance as expected for independent replicates. The average can then be assumed to have a Gaussian distribution for the z-test of the null hypothesis of no effect of the VF and for the power analysis for any nominal percent reduction due to the VF when this percentage is converted to a log odds-ratio (see above). The power, *P*, is calculated as

$P = 1 - Prob(Z > z_c)$ where $Z \sim N\left(\beta_4(R=50), SE^2_{\hat{\beta}_4}\right)$, $Prob(Z' < z_c) = 0.05$, and $Z' \sim N\left(0, SE^2_{\hat{\beta}_4}\right)$ where $N(\theta, \sigma^2)$ is a Gaussian distribution with mean θ and variance σ^2.

COMPETING INTERESTS DISCLAIMER:

Authors have declared that they have no known competing financial interests OR non-financial interests OR personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

[1] Bender, H. Deterrence of kangaroos from roadways using ultrasonic frequencies – efficacy of the Shu Roo. Department of Zoology, University of Melbourne, 2001.

[2] D'Angelo, G.J.; D'Angelo, J.G.; Gallagher, G.R.; Osborn, D.A.; Miller, K.V.; Warren, R.J. Evaluation of wildlife warning reflectors for altering white-tailed deer behavior along roadways. Wildlife Society Bulletin 2006; 34: 1175-1183.

[3] Wildlife Safety Solutions. What is Virtual Fencing? https://www.wildlifesafetysolutions.com.au/ (30 Oct 2020).

[4] Fox, S.; Potts, J.M.; Pemberton, D.; Crosswell, D. Roadkill mitigation: trialing virtual fence devices on the west coast of Tasmania. Aust. Mammal. 2019, 41, 205–211. doi:10.1071/AM18012

[5] Hurlbert, S.H. Pseudoreplication and the design of ecological field experiments. Ecol. Monogr. 1984; 54: 187-211.

[6] Coulson, G.; Bender, H. Roadkill mitigation is paved with good intentions: a critique of Fox et al. (2019). Aust. Mammal. 2019; 42,122-130 https://doi.org/10.1071/AM19009

[7] Englefield, B.; Candy, S.G.; Starling, M.; McGreevy, P.D. A Trial of a Solar-Powered, Cooperative Sensor/Actuator, Opto-Acoustical, Virtual Road-Fence to Mitigate Roadkill in Tasmania, Australia. Animals 2019; 9(10), 752; doi:10.3390/ani9100752

[8] Fox, S.; Potts, J. Virtual fence devices – a promising innovation: a response to Coulson and Bender (2019). Aust. Mammal. 2022; 42(1), 131-133. https://doi.org/10.1071/AM19031

[9] Rytwinski, T.; van der Ree, R.; Cunnington, G.M.; Fahrig, L.; Findlay, C.S.; Houlahan, J.; Jaeger, J.A. G.; Soanes, K.; van der Grift, E.A. Experimental study designs to improve the evaluation of road mitigation measures for wildlife. J. Environ. Manage. 2015; 154, 48–64. doi:10.1016/j.jenvman.2015.01.048

[10] Conner, M.M.; Saunders, W.C.; Bouwes, N.; Jordan, C. Evaluating impacts using a BACI design, ratios, and a Bayesian approach with a focus on restoration. Environ. Monit. Assess. 2016; 188, 1–14. [CrossRef]

[11] Wood, S.N. Generalized additive models, 2nd ed.; Chapman and Hall/CRC: London, UK, 2017.

[12] Stannard, H.J.; Wynan, M.B.; Wynan, R.J.; Dixon B.A.; Mayadunnage, S.; Old, J.M. Can virtual fences reduce wombat road mortalities? Ecol. Eng. 2021; 172: 106414.

[13] Coulson, G.; Bender, H. Wombat Roadkill Was Not Reduced by a Virtual Fence. Comment on Stannard et al. Can Virtual Fences Reduce Wombat Road Mortalities? Ecol. Eng. 2021, 172, 106414. Animals, 2022; 12, 1323. https://doi.org/10.3390/ani12101323

[14] McCullagh, P.; Nelder, J.A. Generalized Linear Models, (2nd Ed.) Chapman and Hall, London, UK, 1989.

[15] Appleby, R.; Jones, D. An Experimental Trial of 'Virtual Fence' Devices in an Effort to Reduce Vehicle Collisions with Wallabies on Heinemann Road, Mount Cotton. Environmental Futures Research Institute. Available from Griffith University, Queensland. 2020.

[16] GDH Wildlife Roadkill Monitoring - Huon Highway, available from Department of State Growth, Tasmania. 2020.

[17] Freeberg, T.M.; Lucas, J.R. Pseudoreplication Is (Still) a Problem. J. Comp. Psychol. 2009; 123(4), 450–451.

[18] Hurlbert, S.H. On misinterpretations of pseudoreplication and related matters: a reply to Oksanen. OIKOS 2004; 104:3.

[19] Rytwinski, T.; Soanes, K.; Jaeger, J.A.; Fahrig, L.; Findlay, C.S.; Houlahan, J.; van der Ree, R.; van der Grift, E.A. How effective is road mitigation at reducing road-kill? A meta-analysis. PLoSOne 2016; 11, e0166941. doi:10.1371/journal.pone.0166941

[20] Almeida, A.; Loy, A.; Hofmann, H. ggplot2 Compatible Quantile-Quantile Plots in R. The R Journal 2018, 10(2); 248-261. doi.org/10.32614/RJ-2018-051

[21] R Core Team. R: A Language and Environment for Statistical Computing; R Foundation for Statistical Computing: Vienna, Austria, 2018.

[22] Haight, F. A. Handbook of the Poisson Distribution, John Wiley & Sons, New York, NY, USA, 1967. ISBN 978-0-471-33932-8

[23] Woolson, R.F.; Bean, J.A.; Rojas, P.B. Sample Size for Case-Control Studies Using Cochran's Statistic. Biometrics 1984; 42(4), 927-932.

[24] Harkness, W.L. Properties of the Extended Hypergeometric Distribution. Ann. Math. Statist. 1965; 36 (3), 938-945.